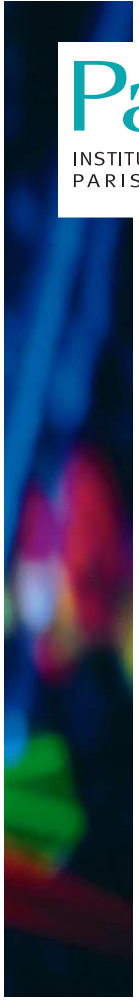
Low dose X-ray image processing

Paolo Irrera



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUT OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0031

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité “Signal et Images”

Présentée et soutenue publiquement par

Paolo IRRERA

le 17 juin 2015

Traitement d’images de radiographie à faible dose

Débruitage et rehaussement de contraste conjoints, et détection automatique de points de repère anatomiques pour l’estimation de la qualité des images

Jury :

Laure	BLANC-FÉRAUD	CNRS - Sophia Antipolis	Rapporteur
Michel	DESVIGNES	Gipsa-Lab - Grenoble INP	Rapporteur
Jalal	FADILI	ENSICAEN - Université de Caen	Examinateur
Antoine	FEYDY	AP-HP-Cochin	Examinateur
Isabelle	BLOCH	Télécom ParisTech	Directeur
Maurice	DELPLANQUE	EOS imaging	Co-directeur

TELECOM ParisTech
école de l’Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

THÈSE

EDITE - ED 130

P H D T H E S I S

to obtain the title of

PhD of Science

of

TELECOM ParisTech

Specialty "Signal and Image Processing"

defended by

Paolo IRRERA

on June 17, 2015

Low dose X-ray image processing

Joint denoising and contrast enhancement, and automatic detection of anatomical landmarks for image quality estimation

Jury :

Laure	BLANC-FÉRAUD	CNRS - Sophia Antipolis	Reviewer
Michel	DESVIGNES	Gipsa-Lab - Grenoble INP	Reviewer
Jalal	FADILI	ENSICAEN - Université de Caen	Examiner
Antoine	FEYDY	AP-HP-Cochin	Examiner
Isabelle	BLOCH	Télécom ParisTech	Advisor
Maurice	DELPLANQUE	EOS imaging	Co-advisor

Remerciements

Je tiens à remercier en tout premier lieu Isabelle et Maurice pour l'encadrement exemplaire de cette thèse et leur soutien tout au long de ce chemin ardu et gratifiant à la fois. Merci Isabelle pour tes conseils toujours pertinents et pour m'avoir transmis la passion pour la recherche. Merci Maurice pour avoir toujours mis en valeurs mes travaux en me faisant confiance et pour m'avoir guidé dans le métier de la recherche en entreprise. Soyez certains que tout ce que j'ai appris ne sera pas oublié et que je vous en serai toujours reconnaissant. Encore une fois, un grand merci !

Je remercie Laure Blanc-Féraud et Michel Desvignes pour avoir accepté d'être rapporteurs de ce manuscrit, ainsi que à Jalal Fadili et Antoine Feydy pour leur évaluation. Je tiens également à remercier Andrés Almansa et Charles Deledalle pour leur avis objectif sur l'avancement des travaux de thèse au moment de l'évaluation mi-parcours.

La thèse a été réalisée dans le cadre d'une convention CIFRE et elle est fruit d'une collaboration entre Télécom ParisTech et EOS imaging. Merci à l'association nationale de la recherche et de la technologie (ANRT) pour avoir contribué à financer cette thèse.

Merci à Marie Meynadier et à tous les collègues d'EOS imaging pour l'implication au travail qu'ils montrent au quotidien et pour la bonne ambiance. En particulier je voudrais remercier Christophe Gomes, Dominique Decarsin et le radiologue Nor-Eddine Regnard pour leur aide dans la validation clinique.

Une grand merci à l'ensemble des professeurs et chercheurs du groupe de traitement et interprétation des images du département TSI pour leur ouverture au dialogue. Cela a été d'une grande aide pendant ma période d'étudiant et m'a motivé à enrichir mes compétences dans le domaine de l'imagerie avec ce doctorat.

Merci à mes parents et à ma sœur qui m'ont constamment supporté en m'écoutent et en dispensant de précieux conseils.

Enfin, cette thèse aurait été surement beaucoup plus difficile sans mes plus chers amis qui, au cours du temps, sont devenus une véritable famille. Ainsi merci Michele, Fernando, Michella, Sara, Caitlin et Guillaume pour votre soutien mais surtout pour tous les moments de rigolade.

Résumé

Nos travaux portent sur la réduction de la dose de rayonnement lors d'examens réalisés avec le système de radiologie EOS. Deux approches complémentaires sont étudiées.

Dans un premier temps, nous proposons une méthode de débruitage et de rehaussement de contraste conjoints pour optimiser le compromis entre la qualité des images et la dose de rayons X. Nous étendons le filtre à moyennes non locales pour restaurer les images EOS. Nous étudions ensuite comment combiner ce filtre à une méthode de rehaussement de contraste multi-échelles. La qualité des images cliniques est optimisée grâce à des fonctions limitant l'augmentation du bruit selon la quantité d'information locale redondante captée par le filtre.

Dans un deuxième temps, nous estimons des indices d'exposition (EI) sur les images EOS afin de donner aux utilisateurs un retour immédiat sur la qualité de l'image acquise. Nous proposons ainsi une méthode reposant sur la détection de points de repère qui, grâce à l'exploitation de la redondance de mesures locales, est plus robuste à la présence de données aberrantes que les méthodes existantes.

En conclusion, la méthode de débruitage et de rehaussement de contraste conjoints donne des meilleurs résultats que ceux obtenus par un algorithme exploité en routine clinique. La qualité des images EOS peut être quantifiée de manière robuste par des indices calculés automatiquement. Étant donnée la cohérence des mesures sur des images de pré-affichage, ces indices pourraient être utilisés en entrée d'un système de gestion automatique des expositions.

Synthèse des travaux de thèse

Cette thèse a été réalisée dans le cadre d'une convention CIFRE et financée par l'ANRT. La thèse s'est déroulée à Télécom ParisTech au laboratoire CNRS LTCI (dans l'équipe de Traitement et Interprétation des images) et dans les locaux d'EOS imaging. Le doctorat a été dirigé par Isabelle Bloch (Télécom ParisTech) et Maurice Delplanque (EOS imaging). Le sujet de cette thèse a porté sur le traitement d'images de radiographie à faible dose. Le but des travaux de thèse était d'exploiter des méthodes de traitement d'images pour baisser la dose de rayonnement lors d'examens réalisés avec le système de radiologie EOS.

Contexte

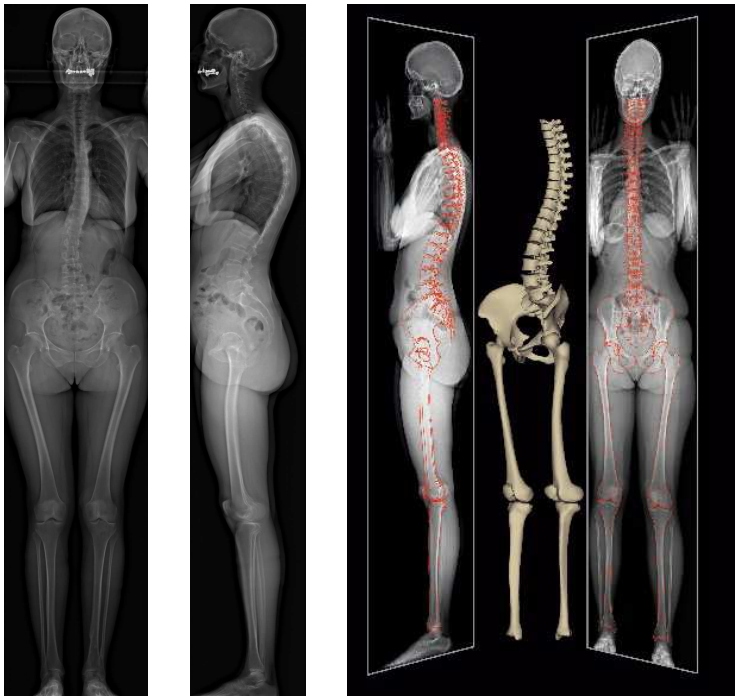
Le système EOS

EOS est un système d'imagerie médicale qui permet de faire de la stéréoradiographie à faible dose. EOS a la particularité d'offrir l'acquisition simultanée d'images en vues de face et de profil d'un patient debout (figures 1a et 1b). Il est aussi possible d'obtenir le modèle tridimensionnel du squelette du patient à l'aide du logiciel propriétaire *sterEOS* (figure 1c).

Les images sont acquises en balayant verticalement un bras en forme de C sur lequel sont montés deux couples de tubes à rayons-X et de détecteurs. Le bras en forme de C est ainsi associé à deux chaînes de traitement d'images, l'une pour l'image en vue de face et l'autre pour celle en vue de profil. Le system EOS utilise un détecteur gazeux multifils issu des travaux de Georges Charpak (Charpak et al., 1968), prix Nobel de physique. Grâce à un effet d'avalanche, les électrons qui traversent le corps du patient génèrent des charges électriques elles-même amplifiées dans le gaz. Cet effet permet d'obtenir des images de qualité diagnostique tout en exposant les patients à une faible quantité de radiations. Les paramètres d'acquisition sont : le pic de tension du tube à rayons X en kilo volts (kV), le courant du tube en milli-ampères (mA) et le facteur de vitesse de balayage verticale du bras en forme de C.

Le principe ALARA en radiologie

L'exposition aux radiations est reconnu comme un risque potentiel pour le corps humain et les dispositifs d'imagerie médicale en sont la source principale. Une étude sur l'usage de systèmes diagnostiques dans six hôpitaux (Smith-Blindman et al., 2012) a mis en évidence l'augmentation du nombre d'examens de tomographie entre les années 1996 et 2010. Par conséquent, la dose effective reçue par patient a également été crue. Pour cette raison, le principe ALARA (*As Low As Reasonably*



(a) (b) (c)

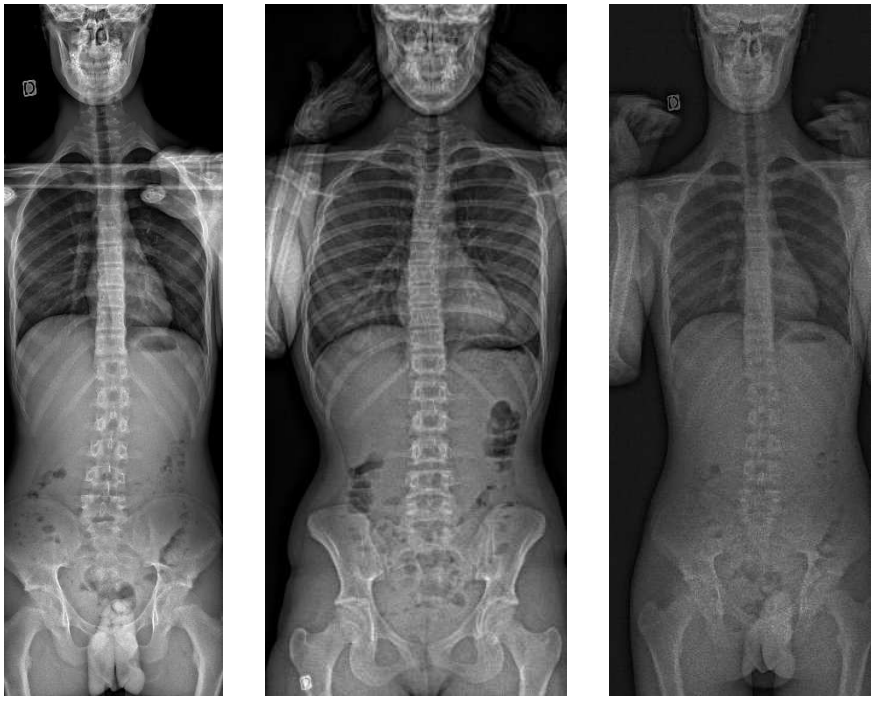
FIGURE 1 : *Images EOS : (a) vue de face ; (b) vue de profil ; (c) reconstruction du squelette à partir d'acquisitions bi-planaires.*

Achievable[1]) a été de plus en plus mis en valeur par les professionnels du monde de la santé. Cette doctrine repose sur l'idée que les paramètres d'acquisition doivent être fixés en sorte de satisfaire le but de l'examen tout en minimisant la quantité de dose reçue par le patient. Autrement dit, il est possible que la qualité des images acquises augmente avec la dose, mais cela ne répond pas forcement à un besoin clinique.

Le système EOS est ALARA car il permet de remplacer, pour certaines applications, des dispositifs classiques de radiologie qui sont plus irradiants. Tout cela ne comporte pas pour autant une réduction de la qualité des images. Par exemple, Deschênes et al. (2010) montrent ce point en comparant des images de rachis acquises avec EOS et avec un système CR (*Computed Radiography*). En outre, EOS prévoit différentes fonctionnalités qui peuvent répondre à des besoins cliniques variés à des niveaux de dose optimaux. Les images *EOS Low Dose* (figure 2a) sont utilisées à des fins diagnostiques, c'est-à-dire pour évaluer structures et/ou lésions. L'exposition aux radiations peut être réduite jusqu'à un facteur 7[2] avec les images *EOS Micro Dose* (figure 2b) qui sont employées pour des examens de suivi. Finalement, les images de pré-affichage (figure 2c) peuvent servir pour contrôler la position du patient avant d'acquérir les images diagnostiques ou de suivi.

[1] Aussi basse que possible

[2] Les valeurs de dose fournies sont en KERMA dans l'air, mesurés en grays.



(a) (b) (c)

FIGURE 2 : *Fonctionnalités EOS : (a) EOS Low Dose, c'est-à-dire diagnostique (*$115,62\ \mu Gy$*) ; (b) EOS Micro Dose, c'est-à-dire examen de suivi (*$14,64\ \mu Gy$*) ; (c) pré-affichage (*$1,44\ \mu Gy$*).*

Approches choisies et défis

Deux approches sont étudiées dans cette thèse pour réduire l'exposition aux rayons X. Dans un premier temps, nous proposons d'optimiser la méthode de post-traitement des images acquises pour combler la perte de rapport signal sur bruit (SNR) due à la baisse de dose de rayons X. Dans un deuxième temps, nous présentons comment estimer la qualité des images pendant l'acquisition dans le but d'exploiter cette information pour guider une boucle d'asservissement en charge de définir les paramètres d'exposition optimaux au sens ALARA.

Optimisation de la qualité des images

Les images acquises à faible dose sont significativement affectées par le bruit. En outre, l'évaluation clinique est compliquée en raison de la faible visibilité des structures anatomiques fines. Par conséquent, le contraste doit être rehaussé tout en faisant attention à ne pas augmenter le bruit.

De plus, comme les images EOS peuvent couvrir la totalité du corps, les propriétés du signal changent en fonction des régions anatomiques. Par exemple, dans les genoux, où les rayons X sont peu absorbés, l'image est riche en texture osseuse. En revanche, dans des régions à forte absorption rayons X associées aux tissus mous, les variations dans l'image sont principalement dues au bruit. Il est ainsi nécessaire de proposer une méthode capable de s'adapter à ces altérations dans les images à trai-

ter. Cela est une caractéristique fondamentale aussi pour une autre raison. En effet, la qualité des images dépend aussi des paramètres d'acquisition et du morphotype du patient car ces facteurs influencent la quantité du signal qui arrive au détecteur.

Ces questions sont traitées en proposant une approche qui permet de rehausser le contraste tout en minimisant l'impact du bruit qui affecte les images EOS d'entrée. Nous présentons d'abord notre extension du filtre par moyennes non locales qui permet de restaurer les images EOS qui sont corrompues par un bruit qui dépend du niveau de signal. Ce filtre est combiné avec le module de rehaussement de contraste de différentes manières, qui peuvent répondre à des besoins cliniques différents. Finalement, nous minimisons le nombre des paramètres de la méthode pour faire en sorte que l'approche soit robuste aux propriétés hétérogènes des images EOS.

Quantification automatique de la qualité des images à l'acquisition

Il est très utile de pouvoir vérifier que les opérateurs utilisent bien le système en suivant le principe ALARA. Il serait donc intéressant de quantifier cette information. De plus, cette mesure obtenue à partir d'images à très faible dose, comme par exemple les images de pré-affichage, pourrait être utilisée comme entrée d'une boucle d'asservissement en charge de définir les paramètres d'acquisition optimaux pour les images diagnostiques suivantes.

La mesure de qualité des images doit être associée à une région anatomique significative pour le besoin clinique de l'examen en cours. Pour autant, une seule évaluation n'est pas représentative pour des images du corps entier parce que plusieurs structures anatomiques sont présentes. Par conséquent, il est nécessaire d'associer une valeur à chaque structure anatomique d'intérêt. Cependant, détecter ces régions anatomiques n'est pas une tâche simple car il faut tenir compte des différents morphotypes et âges des patients ainsi que des multiples modalités d'acquisition. De plus, des cas particuliers comme la présence d'objets métalliques ou de structures déformées par des pathologies ne doivent pas induire la méthode d'estimation en erreur.

Nous proposons une approche reposant sur des points de repère, répondant au problème de la présence de plusieurs structures dans l'image. Une estimation locale est ensuite associée à chaque point de repère. Cela rend la mesure associée à une région anatomique moins sensible à la présence de valeurs aberrantes. Nous proposons ainsi une technique de détection automatique de points de repère.

Débruitage et rehaussement de contraste conjoint

État de l'art sur le traitement d'images de radiographie

Il existe essentiellement trois catégories de méthodes pour améliorer la qualité des images par rayons X.

Le premier type d'approche consiste à rehausser le contraste d'une image d'entrée en exploitant la décomposition multi-échelles. Cette dernière consiste à représentant une image en N_c canaux qui sont obtenus par filtrage itératif et qui codent l'information en fonction des niveaux de fréquence. L'image d'entrée est donc représentée par N_c canaux passe-bande $\{\mathbf{d_0}, \mathbf{d_1}, \ldots, \mathbf{d_{N_c-1}}\}$ et un résidu de basse fréquences $\mathbf{l_{N_c}}$. Le pyramides laplaciennes (Burt and Adeldon, 1983) sont un exemple de technique

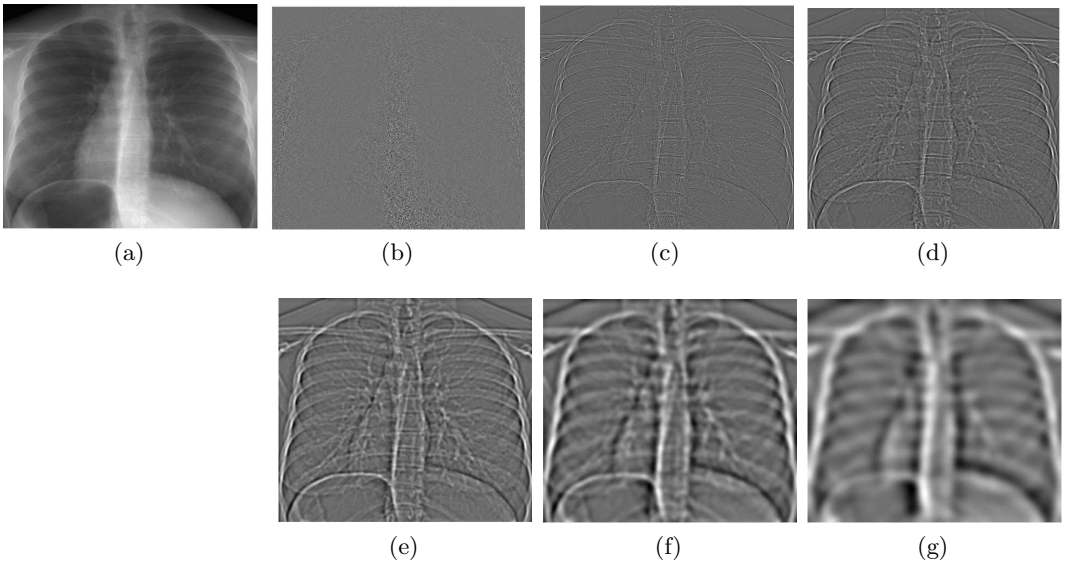


FIGURE 3 : *Décomposition multi-échelles d'une image EOS : (a) image d'entrée ; (b)* $\mathbf{d}_0$ *: texture osseuse et bruit ; (c)* $\mathbf{d}_1$ *: structures fines des poumons (par exemple) ; (d)* $\mathbf{d}_2$ *: structures des poumons plus grossières (par exemple), (e)* $\mathbf{d}_3$ *: bords des vertèbres (par exemple), (f)* $\mathbf{d}_4$ *: cage thoracique (par exemple), (g)* $\mathbf{d}_5$ *: enveloppe de la colonne vertébrale (par exemple).*

de décomposition multi-échelles linéaire. Dans ce cas, les images passe-bande et les résidus de basse fréquence à chaque niveau sont obtenus de la manière suivante :

$$\begin{cases} \mathbf{l_t} = \downarrow (g(\mathbf{l_{t-1}})) \\ \\ \mathbf{d_t} = \mathbf{l_{t-1}} - \uparrow (\mathbf{l_t}) \end{cases} \quad t \in [1, N_c] \tag{1}$$

où $\mathbf{l}_0$ coïncide avec l'image d'entrée $\mathbf{u}$, $\downarrow$ et $\uparrow$ sont les opérateurs de sous-échantillonnage et sur-échantillonnage, respectivement, et $g(.)$ est un filtre gaussien. Dippel et al. (2002) ont montré que cette technique est préférable aux ondelettes car elle permet de préserver le rendu visuel radiologique après que les images passe-bande $\mathbf{d}_t$ aient été rehaussées en contraste. Cependant, Li et al. (2005) ont montré qu'il est préférable d'opter pour une décomposition non-décimée qui s'affranchit des artéfacts d'aliasing dus au sous-échantillonnage. Par conséquent, nous utilisons une version des pyramides laplaciennes non-décimées qui sont aussi appelées ondelettes non-décimées isotropiques (IUWT, Starck et al. (2007)). La figure 3 montre un exemple de décomposition d'une image EOS avec la méthode IUWT. La synthèse est l'opération qui inverse la décomposition multi-échelles. Dans le cas IUWT il s'agit simplement d'ajouter toutes les images passe-bande au résidu de basse fréquence. Ensuite, l'image $\mathbf{e}$ peut être obtenue en appliquant les fonctions ξ_t aux niveaux de détail $\mathbf{d}_t$, c'est-à-dire de la manière suivante :

$$\mathbf{e} = \mathbf{l}_{N_c} + \sum_{t=0}^{N_c-1} \xi_t(\mathbf{d}_t). \tag{2}$$

Les fonctions ξ_t sont toutes fondées sur un même modèle mais leurs paramètres changent selon le niveau de détail, donc du type d'information à rehausser. Le modèle en question correspond à une fonction non-linéaire qui augmente le contraste dans des régions à faible activité et le préserve là où l'énergie est haute, par exemple en présence d'un contour. Dans ces travaux de thèse nous utilisons les fonctions proposées par Dippel et al. (2002).

Le problème de la première catégorie d'approches est que la présence du bruit est négligée, ce qui n'est pas acceptable dans le cadre du traitement d'images à faible dose. Une solution consiste à appliquer un filtre de débruitage sur l'image $\mathbf{u}$ pour obtenir une image $\mathbf{y}$ où le bruit a été fortement réduit. C'est ensuite l'image $\mathbf{y}$ qui est utilisée comme entrée du rehaussement de contraste. Cette approche a été utilisée, par exemple, par Sakata and Ogawa (2009) qui proposent deux décompositions en ondelettes en cascade, la première pour débruiter et la deuxième pour rehausser le contraste. Cependant, cette approche peut amener à une perte d'information. En effet, comme nous pouvons le déduire de l'analyse des résultats de filtres utilisés pour des applications en imagerie médicale (Coupé et al., 2008; Cerciello et al., 2012; Xu et al., 2012), il est très difficile de préserver la totalité de l'information diagnostique. Par exemple, en filtrant des images à rayons X, la texture osseuse peut être significativement détruite.

Le dernier type d'approche tient compte du fait qu'il n'est pas concrètement possible de séparer l'information des oscillations dues au bruit sur des images à rayons X à de faible contraste. Par conséquent, aucune information n'est retirée aux images d'entrée, mais les zones de l'image qui sont censées être corrompues par le bruit ne sont pas rehaussées. En pratique, cela revient à utiliser la somme pondérée coefficient par coefficient suivante :

$$\hat{\mathbf{d}}_t = (1 - \mathbf{n}_t)\xi_t(\mathbf{d}_t) + \mathbf{n}_t\mathbf{d}_t \tag{3}$$

où l'image $\mathbf{n}_t$ est une carte de bruit dont les valeurs sont comprises entre 0 et 1. Les niveaux de $\mathbf{n}_t$ sont d'autant plus élevées que la présence de bruit est significative. Les cartes de bruit dépendent du niveau de fréquence pour prendre en compte le fait que l'impact du bruit est moins significatif à des échelles plus grossières. Le souci principal des méthodes qu'on retrouve dans la littérature est lié au fait que les cartes de bruit dépendent de paramètres empiriques. Par exemple, Stahl et al. (1999) utilisent des valeurs d'écart type local pour distinguer le bruit de l'information. Les cartes de bruit sont sous-optimales vis-à-vis du nombre élevé de cas possibles qui se présentent dans les images à traiter. Par conséquent nous étudions d'autres solutions.

Filtre par moyennes non locales pour des images à rayons X

Le filtre par moyennes non locales (NL-m) a été proposé par Buades et al. (2005a) et il repose sur une représentation par patchs. Cette dernière rend la méthode de filtrage robuste à la présence du bruit en s'appuyant sur la présence d'informations redondantes dans l'image. Par conséquent le filtre NL-m permet de préserver différentes structures dans l'image grâce au fait que la représentation dans l'espace des patchs décrit bien le contexte spatial de l'image. De plus, en présence d'une quantité suffisante d'information redondante, la solution converge vers celle qui est optimale au sens bayésien (Buades et al., 2005b). En considérant que les images à rayons X

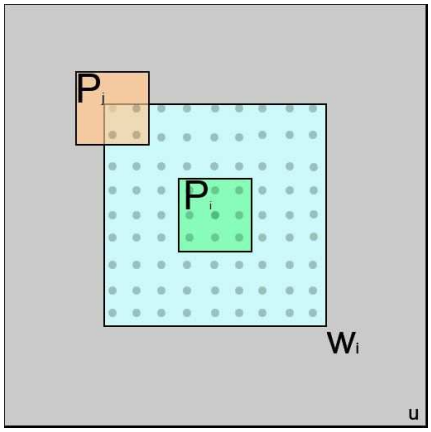


FIGURE 4 : *Recherche de patchs similaires. Le patch vert P_{x_i} centré au pixel x_i est comparé aux patchs oranges P_{x_j} centrés aux pixels $x_j \in W_i$. Dans cet exemple, $|P| = 3 \times 3$ et $|W| = 9 \times 9$.*

du corps entier sont très redondantes, la représentation par patchs est bien adaptée à la restauration des images EOS.

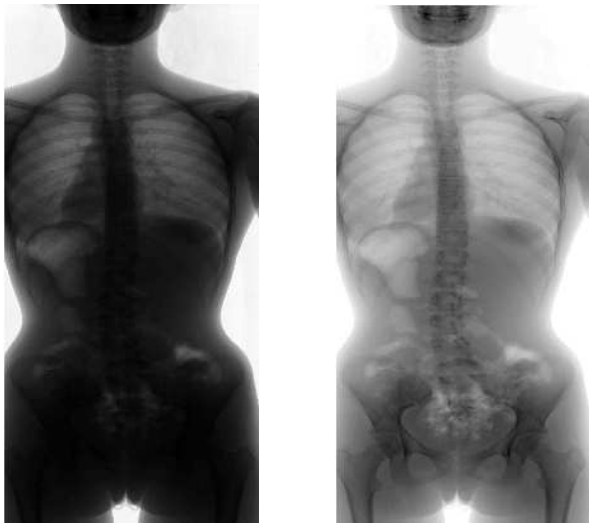
La valeur d'un pixel x_i de l'image filtrée est obtenue par une moyenne pondérée des valeurs d'entrée dans un voisinage de taille $|W|$ (voir figure 4). En pratique, les poids sont définis en utilisant la formule suivante :

$$\omega(i,j) = \varphi\left(\frac{d(P_{x_i}, P_{x_j})}{2h^2|P|}\right) \tag{4}$$

où P_{x_i} est la patch centrée au pixel x_i, φ est une fonction monotone décroissante (par exemple $\exp(-t)$), $d(.)$ quantifie la distance entre les patchs, h est un paramètre de lissage qui dépend du niveau de bruit et $|P|$ est la taille des patchs. En particulier le modèle de distance $d(.)$ doit être choisi en fonction du type de bruit. Par exemple, dans le cas de bruit additif gaussien blanc, il s'agit d'une distance euclidienne. Dans ce dernier cas, la méthode peut être implémentée en utilisant les images intégrales comme Darbon et al. (2008) le proposent.

L'adaptation du filtre aux images EOS passe par la modélisation du bruit qui affecte les images. Si nous considérons les images acquises $\mathbf{u}_r$ après la correction en offset et gain (figure 5a), le bruit suit une distribution poissonienne. Cela est obtenu en mesurant l'écart type (figure 6a) et le rapport signal sur bruit (figure 6c) sur des régions homogènes dans des images de fantôme. Cependant, les images acquises ne sont pas celles qui sont utilisées en entrée du module de rehaussement de contraste. En effet, pour compenser l'atténuation exponentielle reçue par les rayons X qui traversent le corps du patient, une fonction logarithmique est appliquée à l'image $\mathbf{u}_r$ pour obtenir l'image $\mathbf{u}_l$ (figure 5b). La figure 6b montre que cette opération inverse la courbe de bruit, c'est-à-dire que le bruit devient plus fort lorsque le signal devient plus faible. Le bruit peut être localement approximé par une distribution gaussienne dont l'écart type est fonction du signal (Hensel et al., 2006). Il est donc nécessaire d'estimer ces niveaux de bruit.

La méthode du percentile (Ponomarenko et al., 2007) permet d'estimer l'écart type du bruit additif gaussien blanc à partir d'une image. Une image d'entrée $\mathbf{u}$

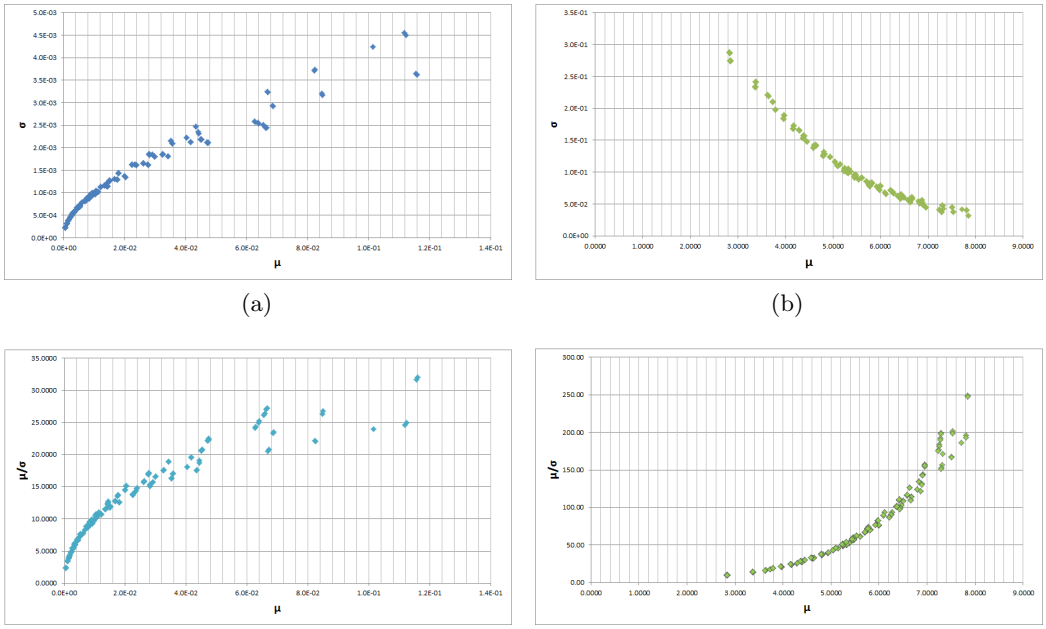


(a) (b)

FIGURE 5 : *Les images acquises sont d'abord corrigées en gain et offset. Les images EOS peuvent être représentées dans les domaines suivants : (a) linéaire ($\mathbf{u}_r$) ; (b) logarithmique ($\mathbf{u}_l$).*

(figure 7a) est d'abord filtrée en passe haut pour ne pas tenir compte de la composante de basse fréquence de l'image. À partir de l'image obtenue $\mathbf{u}_h$ (figure 7b), des mesures d'énergie (par exemple la variance de la distribution des niveaux de gris) sont associées à l'ensemble des fenêtres superposées de taille $|W|$ qui recouvrent la totalité de l'espace des pixels. Une valeur de référence est ensuite retenue en prenant un percentile bas des mesures d'énergie triées par ordre croissant. Le choix du percentile bas est dû au fait que les composantes de signal sont éparses alors que ce n'est pas le cas du bruit. Ainsi, si l'énergie associée à un bloc est élevée, cela signifie que de l'information y est présente. Colom and Buades (2013) ont ensuite développé cette approche en faisant tomber l'hypothèse de bruit additif gaussien blanc. Cette méthode permet d'estimer des niveaux de bruit qui dépendent du signal. En pratique, l'image d'entrée est divisée en un nombre N_r des régions non superposées en fonction du niveau de signal μ_i. Dans chaque région i le bruit est considéré additif Gaussien blanc et la méthode du percentile est utilisée pour estimer le niveau de bruit σ_i. Finalement, l'ensemble d'échantillons $\{(\mu_1, \sigma_1), \ldots, (\mu_{N_r}, \sigma_{N_r})\}$ sont interpolés linéairement pour estimer une courbe de bruit. Nous exploitons cette méthode pour modéliser la relation entre signal et bruit sur des images cliniques initiales avant rehaussement de contraste.

En considérant le modèle de bruit des images après la conversion logarithmique, il est possible d'adapter le filtre NL-m pour débruiter les images EOS. Effectivement, nous supposons que le bruit dans une petite fenêtre est approximativement constant car la densité des tissus ne change pas brusquement. Ensuite, en considérant que le paramètre de lissage h (voir équation 4) est proportionnel à σ^2 (Buades et al., 2005b), la distance entre les patchs est ajustée en fonction du niveau de bruit estimé



(a) (b)

(c) (d)

FIGURE 6 : *Analyse du modèle de bruit : (a) $\sigma(\mathbf{u}_r)$; (b) $\sigma(\mathbf{u}_l)$; (c) $SNR(\mathbf{u}_r)$; (d) $SNR(\mathbf{u}_l)$. Les courbes (a) and (c) montrent que la distribution du bruit dans les images $\mathbf{u}_r$ est poissonnienne, mais cette propriété n'est pas préservée après la conversion au logarithme.*

avec la méthode des percentiles de Colom and Buades (2013) :

$$d(P_i, P_j) = \frac{\|\mathbf{u}(P_i) - \mathbf{u}(P_j)\|_2}{2{\sigma_i}^2|P|}. \tag{5}$$

Les poids sont calculés de la manière suivante :

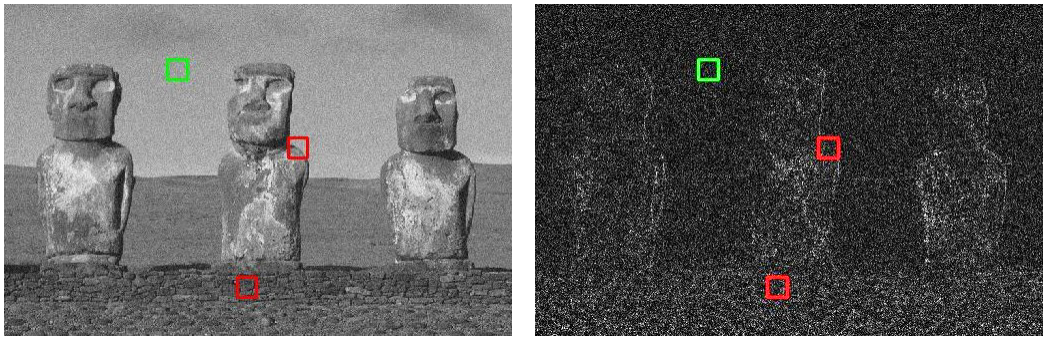
$$\varsigma_X(i,j) = \exp\left(-d_X(P_i, P_j)\right). \tag{6}$$

Par la suite le filtre proposé est noté avec l'acronyme XNL-m. Ce filtre permet de prendre en considération la variabilité intra- et inter-patients de manière complètement automatique comme il est montré dans la figure 8. En détail, par rapport au filtre NL-m classique, le filtre XNL-m permet de préserver la quasi-totalité de l'information dans des zones à faible absorption et, en même temps, de ne pas laisser de résidus de bruit là où ce dernier est plus haut.

Les images dans les figures 8d et 8h sont obtenues en appliquant le filtre XNL-m et, ensuite, sont rehaussées en contraste. Les résultats sont bons, mais il est néanmoins possible de remarquer une perte de la finesse de la texture osseuse. Une solution reposant sur le calcul d'une carte de bruit est donc proposée.

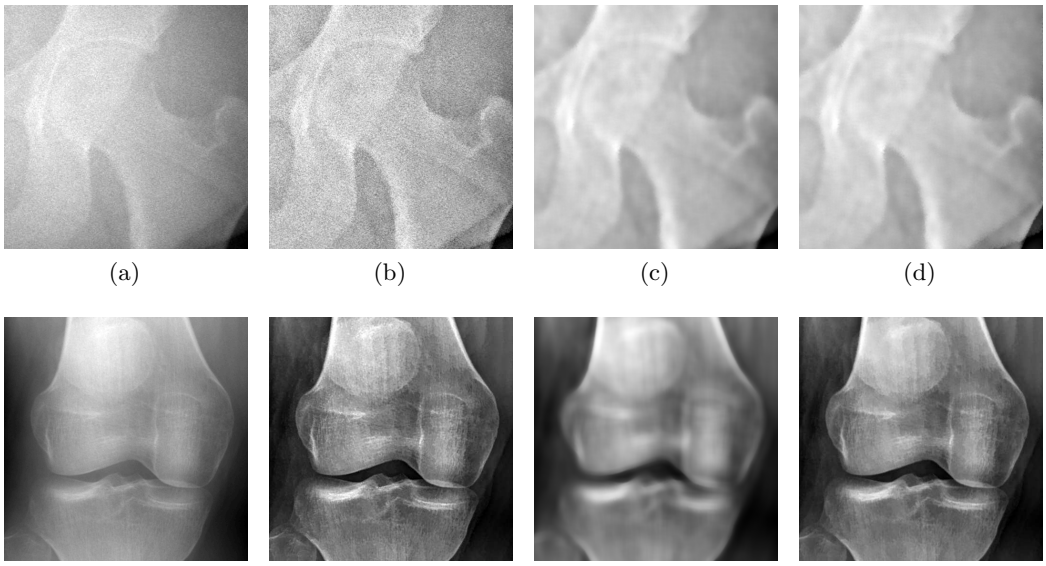
Carte de bruit

Les images passe-bande $\mathbf{d}_l^u$ obtenues en décomposant l'image $\mathbf{u}$ en plusieurs niveaux multi-échelles présentent à la fois des composantes de signal et de bruit. En



(a) (b)

FIGURE 7 : *Estimation du bruit par la méthode des percentiles : (a) image d'entrée ; (b) image filtrée passe haut. Les fenêtres ou il n'y a pas de composantes de signal (verte) sont retenues en considérant un bas percentile des mesures d'énergie. En revanche, celles en correspondance de bords et texture (rouge) sont exclues.*

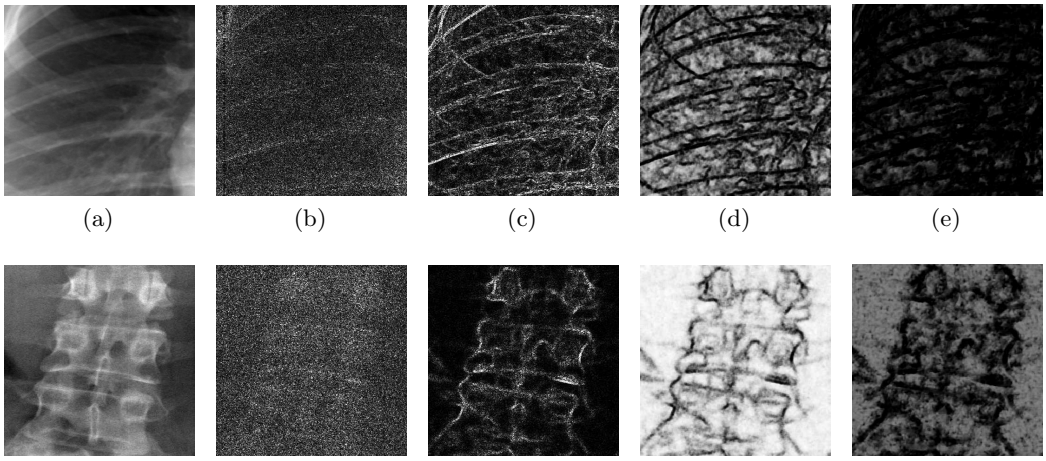


(a) (b) (c) (d)

(e) (f) (g) (h)

FIGURE 8 : *XNL-m versus NL-m avec paramètres définis à la main Le paramètre de lissage h est réglé : (b,f) pour bien restaurer un genou ; (c,g) pour bien restaurer la tête fémorale dans la même image ; (e,h) automatiquement avec le filtre XNL-m.*

revanche, les images passe-bande $\mathbf{d}_t^{\hat{y}}$ obtenues à partir de l'image résultante du filtre XNL-m $\hat{\mathbf{y}}$ présentent uniquement des composantes de signal, quoiqu'une partie de l'information pourrait manquer. On considère à titre d'exemple une région dans les poumons (figure 9a) et une autre dans l'abdomen (figure 9f). En observant les niveaux de détail $\mathbf{d}_0^u$ (figures 9b et 9g), nous remarquons que les structures anatomiques sont noyées dans le bruit et, donc, à peine visibles. Au contraire, l'information est bien distinguable sur les niveaux de détail $\mathbf{d}_0^{\hat{y}}$ (figures 9c et 9h). Les cartes de



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

FIGURE 9 : *Cartes de bruit. Première ligne (a-e) : région de poumons ; Deuxième ligne (f-j) : région lombaire. Colonnes de gauche à droite : (a,f) images d'entrée ; (b,g)* $\mathbf{d}_0^u$ *; (c,h)* $\mathbf{d}_0^{\hat{y}}$ *; (d,i)* $\mathbf{n}_0$ *; (e,j)* $\mathbf{n}_1$.

bruit sont calculées de la manière suivante :

$$\mathbf{n}_t = \delta_t \left(1 - \frac{2\mathbf{c}_t^u \mathbf{c}_t^{\hat{y}} + \epsilon}{(\mathbf{c}_t^u)^2 + \left(\mathbf{c}_t^{\hat{y}}\right)^2 + \epsilon} \right) \tag{7}$$

où $\mathbf{c}_t^u$ et $\mathbf{c}_t^{\hat{y}}$ sont des mesures de contraste[3] associées aux niveaux $\mathbf{d}_t^u$ et $\mathbf{d}_t^{\hat{y}}$, respectivement, ϵ est une valeur constante égale à 1 utilisée pour éviter les singularités, et δ_t est un facteur d'énergie calculé par :

$$\delta_t = 1 - \frac{\Xi(\mathbf{d}_t^{\hat{y}})}{\Xi(\mathbf{d}_t^u)}, \qquad \Xi(\mathbf{d}_t) = 2 \left(\frac{1}{M} \sum_{i=1}^{N} |\mathbf{d}_t(z_i)| \right)^2 \tag{8}$$

où z_i correspond aux coordonnées d'un coefficient i. En analysant les résultats, nous arrivons à la conclusion que les cartes de bruit estimées s'adaptent bien au contenu des régions anatomiques et aux différents niveaux de bruit (figure 9d et 9i). De plus, ces cartes tiennent compte du fait que le bruit est de moins en moins influant à des échelles plus grossières. Les résultats sont obtenus sans devoir fixer empiriquement des paramètres. En effet, les cartes de bruit sont déduites uniquement à partir de la sortie du filtre XNL-m proposé.

Validation

Quatre méthodes sont comparées pour évaluer celle qui s'adapte le mieux au contexte clinique. La première approche rehausse le contraste de l'image en entrée en exploitant une décomposition multi-échelles. La présence de bruit est négligée dans ce cas et on utilise la notation *NE* pour se référer à cette approche. La deuxième

[3]Moyenne de l'amplitude des coefficients dans des fenêtres de taille 5×5.

méthode consiste à estimer l'image non-affectée par le bruit $\mathbf{y}$ à partir de l'image d'entrée $\mathbf{u}$ en utilisant le filtre XNL-m. L'image obtenue à la sortie du filtre est ensuite rehaussée en contraste avec la même technique que celle exploitée dans *NE*. Dans le texte qui suit on notera cette méthode avec *DE*. La troisième méthode exploite les cartes proposées pour contenir le bruit dans le module de rehaussement de contraste multi-échelles. On notera cette méthode *LNCE*. Finalement, on considère aussi un algorithme utilisé en routine clinique pour post-traiter des images EOS, qui repose sur des cartes de bruit dont l'estimation dépend de valeurs prédéfinies sur le niveau de bruit. Cette méthode est notée *EOSE*.

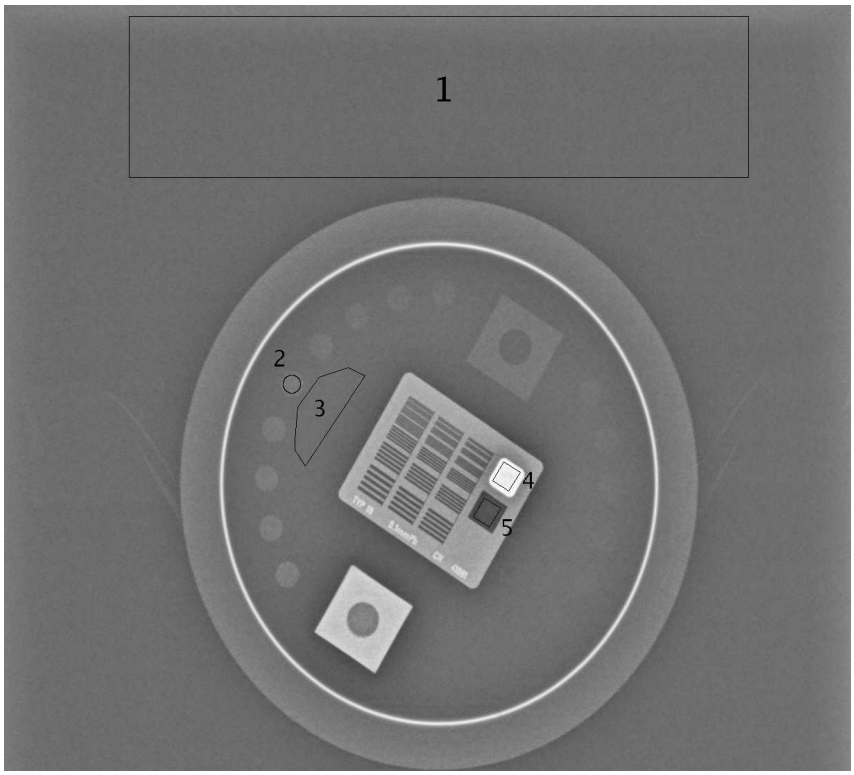


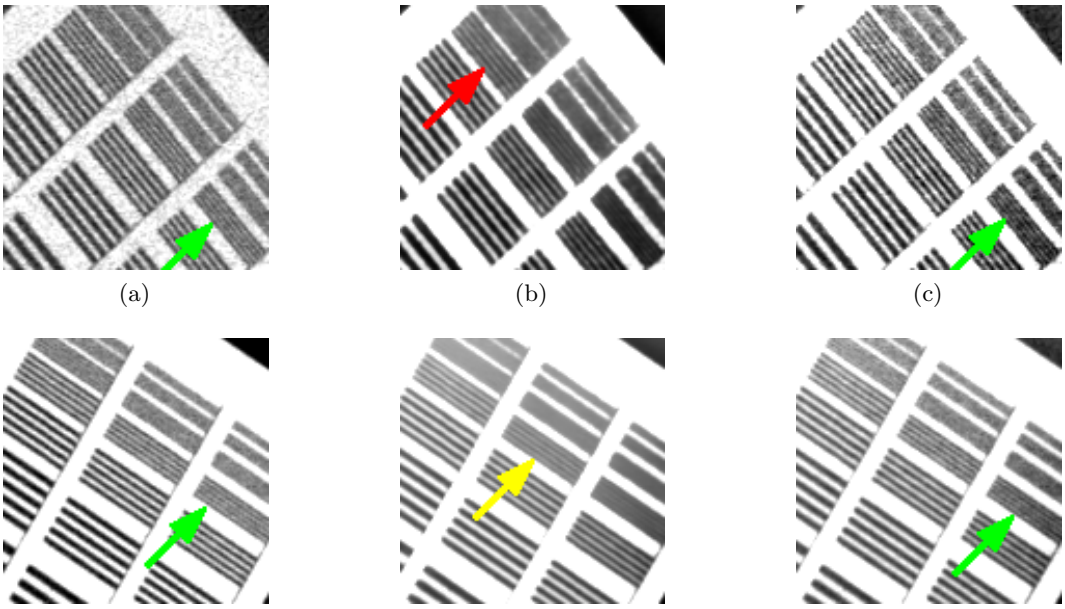
FIGURE 10 : *Fantôme standard PHD5000 fluoro avec les régions d'intérêt annotées à la main pour calculer les mesures suivantes : SNR à partir de la région 1 ; CNR à partir des régions 2 et 3 ; DYN à partir des régions 4 et 5. La grille de haute résolution au centre du fantôme est utilisée pour quantifier la résolution spatiale.*

On considère d'abord des images de fantôme (figure 10) qui nous donnent la possibilité de mesurer le rapport signal sur bruit (SNR), le contraste sur bruit (CNR), la dynamique des niveaux de gris (DYN) et la résolution spatiale. Nous utilisons des blocs de PMMA d'épaisseurs comprises entre 10 *cm* et 30 *cm* qui sont interposés entre le faisceau de rayons X et l'objet pour simuler les changements de morphotypes. De plus, la puissance du signal est également modifiée en faisant varier la dose d'entrée du fantôme entre 10 μGy et 566 μGy.

La table 1 montre de combien augmentent les mesures de SNR, CNR et DYN en moyenne entre l'image de faible contraste et celles post-traitées avec les quatre techniques étudiées. En rehaussant le contraste, le SNR à la sortie est inférieur à celui d'origine dans tous les cas. Cependant, nous observons que *DE* minimise

TABLE 1 : *Variation de SNR, CNR et DYN sur les images post-traitées avec NE, DE, LNCE et EOSE par rapport à l'image en entrée.*

Mesures moyennes	*NE*	*DE*	*LNCE*	*EOSE*
$(SNR_e - SNR_u)/SNR_u$	$-81,0\%$	$\mathbf{-28,9}\%$	$-69,5\%$	$-72,9\%$
$(CNR_e - CNR_u)/CNR_u$	$-20,9\%$	$\mathbf{281,9}\%$	$43,7\%$	$-21,5\%$
$(DYN_e - DYN_u)/DYN_u$	$\mathbf{53,8}\%$	$\mathbf{53,0}\%$	$\mathbf{53,5}\%$	$32,8\%$



(a) (b) (c)

(d) (e) (f)

FIGURE 11 : *Grille de résolution derrière un bloc de* 10 *cm de PMMA. Première ligne (a-c)* 10 μGy *; deuxième ligne (d-f)* 71 μGy*. Première colonne (a, d, g) image d'entrée ; deuxième colonne (b, e, h) DE ; troisième colonne (c, f, i) LNCE. Les flèches pointent les niveaux de résolution les plus fins : vert = résolution préservée ; jeune = perte de résolution égale à* $0,2$ *lp/mm ; rouge = perte de résolution égale à* $0,4$ *lp/mm.*

l'augmentation de bruit. Cette dernière méthode est aussi la meilleure en termes de CNR. Nous constatons également que *LNCE* est la seule approche, outre *DE*, qui rehausse d'avantage l'information par rapport au bruit, ce qui est indiqué par les valeurs positives de la mesure de CNR dans la table 1. En ce qui concerne la dynamique des niveaux de gris, *NE*, *DE* et *LNCE* donnent des résultats similaires parce qu'ils exploitent la même technique pour le réglage du contraste global ce qui n'est pas le cas d'*EOSE* qui s'avère moins performante. Les approches *DE* et *LNCE* sont ainsi les meilleures, avec un avantage pour la première en termes de SNR et CNR. Cependant, en évaluant la résolution spatiale sur des images traitées avec ces deux méthodes on s'aperçoit que *LNCE* est une solution meilleure. En effet, la figure 11 montre que *LNCE* préserve la résolution spatiale des images d'entrée alors que ce n'est pas le cas avec *DE*. En détail, dans le cas où le signal qui arrive au

TABLE 2 : *Pour un niveau de dose diagnostique donné (d_d) on compare de combien augmentent le SNR, CNR et DYN par rapport à une image acquise à un niveau de dose utilisé dans des examens de suivi (d_f) post-traitée avec EOSE and LNCE. L'évaluation est conduite pour une épaisseur de* 10 *cm et* 20 *cm.*

PMMA	d_d/d_f	**EOSE**(d_d)/**EOSE**(d_f)			**EOSE**(d_d)/**LNCE**(d_f)		
		SNR	CNR	DYN	SNR	CNR	DYN
10 *cm*	7, 10	×1, 56	×1, 86	+1%	**×1, 17**	**×0, 95**	**−13%**
20 *cm*	6, 62	×1, 34	×1, 40	+3%	×1, 32	**×1, 01**	**−12%**

TABLE 3 : *Valeurs moyennes de AC et UCII obtenues à partir de tests sur 150 images cliniques.*

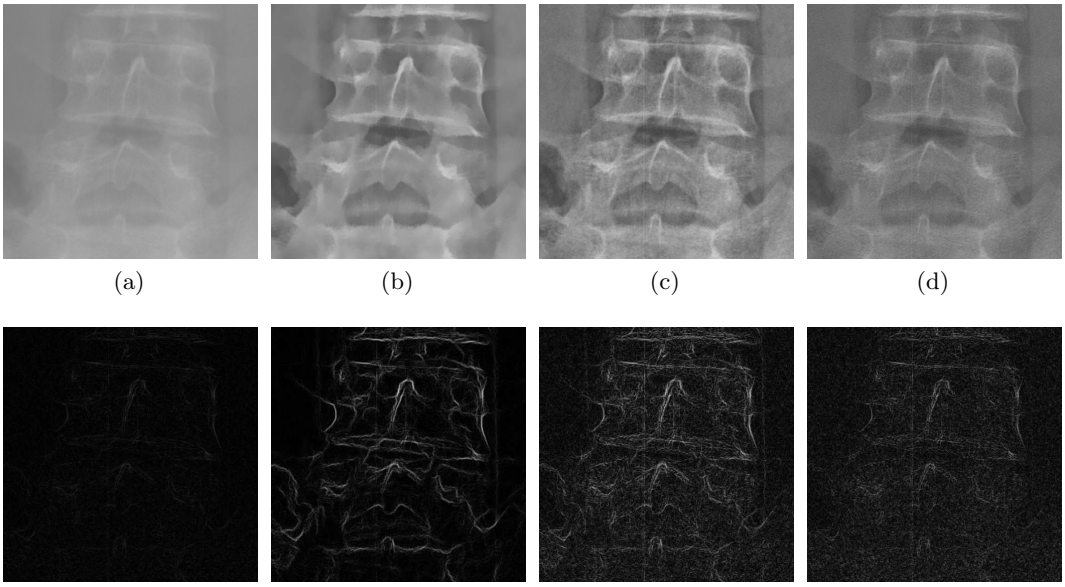
	IN	**NE**	**DE**	**LNCE**	**EOSE**
AC	2, 01	**25, 14**	3, 79	**11, 45**	5, 69
UCII	0	−3, 09	**1, 80**	**1, 37**	−1, 16

détecteur est particulièrement faible (figure 11b) la perte peut s'élever à 0, 4 paires de lignes par millimètre (lp/mm).

Les images de fantôme peuvent aussi être utilisées pour quantifier la réduction de la qualité d'image vis-à-vis d'une baisse de la dose rayons X. La table 2 montre que, même si l'exposition aux rayons X est réduite d'un facteur égal à 7, 10, l'approche *LNCE* permet de préserver presque le même niveau de performance d'images acquises à dose diagnostique et post-traitée avec *EOSE* pour une épaisseur de PMMA à traverser de 10 *cm*. Les résultats sont également encourageants quand l'absorption des rayons X est plus forte, mais il y a néanmoins une différence non négligeable en termes de SNR.

La validation sur des images cliniques est fondée sur des mesures de contraste telles que la moyenne des variances locales (ALV). Ces valeurs sont considérées soit dans l'absolu, soit relativement aux mesures d'entrée. Dans ce dernier cas on obtient des indices d'augmentation de contraste (CII) par rapport à l'image d'entrée. Dans la littérature, ces mesures sont calculées dans des régions définies de manière empirique. En revanche, nous proposons d'associer aux mesures une signification sémantique en traçant à la main des régions d'intérêt clinique. Nous calculons des mesures ALV anatomiques à partir desquelles nous pouvons en inférer deux autres. Nous appelons la première mesure contraste anatomique (AC) et elle indique la valeur absolue de contraste dans des régions anatomiques d'intérêt. La deuxième mesure est appelée indice d'augmentation de contraste non-biaisée (UCII) et elle quantifie de combien le contraste augmente dans des régions d'intérêt par rapport aux zones dont les variations de haute fréquence sont exclusivement dues au bruit.

La table 3 montre que, dans l'absolu, le contraste le meilleur est obtenu avec la méthode NE, mais cela au prix d'une augmentation excessive de bruit comme la valeur négative de UCII l'indique. La méthode DE donne des résultats diamétralement opposés à ces derniers. En effet, uniquement l'information pertinente est rehaussée (UCII = 1, 80), mais le filtre induit aussi une perte de détails comme la valeur faible d'AC le laisse entendre. La méthode proposée de carte de bruit est celle qui offre le meilleur compromis entre AC et UCII. De plus, l'amélioration par rapport à l'algorithme EOSE est significative. Cette évaluation peut être confirmée



(a) (b) (c) (d)

(e) (f) (g) (h)

FIGURE 12 : *Comparaison qualitative du rendu visuel pour L4 et L5 : première ligne = images et deuxième ligne = gradient de Sobel ; colonnes de gauche à droite = IN, DE, LNCE and EOSE.*

en analysant qualitativement les images en figure 12. En particulier, en post-traitant l'image d'entrée avec l'algorithme DE, les bords des vertèbres sont bien préservés alors que la texture osseuse est lissée. Le rendu visuel attendu en radiologie est ainsi perdu. La méthode LNCE préserve la totalité de l'information diagnostique tout en donnant une visibilité des détails supérieure par rapport à EOSE. Ces deux dernières méthodes ont été comparées par un radiologue.

TABLE 4 : *Notes sur la visibilité de structures anatomiques données par un clinicien sur une base de données de* 20 *images post-traitées avec EOSE et LNCE.*

	$\mathfrak{C}^{\mathfrak{f}}$	$\mathfrak{C}^{\mathfrak{l}}$	$\mathfrak{D}^{\mathfrak{f}}$,	$\mathfrak{D}^{\mathfrak{l}}$	$\mathfrak{L}^{\mathfrak{f}}$	$\mathfrak{L}^{\mathfrak{l}}$	$\mathfrak{P}$	$\mathfrak{K}$
EOSE	56%	78%	54%	57%	62%	67%	80%	95%
LNCE	67%	85%	69%	67%	71%	79%	90%	97%
Gain	**+11**	**+7**	**+15**	**+10**	**+9**	**+12**	**+10**	**+2**

Le clinicien a d'abord identifié des ensembles de structures anatomiques d'intérêt dans les régions suivantes :

- Rachis cervical en vue de face ($\mathfrak{C}^{\mathfrak{f}}$)
- Rachis cervical en vue latérale ($\mathfrak{C}^{\mathfrak{l}}$)
- Rachis thoracique en vue de face ($\mathfrak{D}^{\mathfrak{f}}$)
- Rachis thoracique en vue latérale ($\mathfrak{D}^{\mathfrak{l}}$)
- Rachis lombaire en vue de face ($\mathfrak{L}^{\mathfrak{f}}$)

- Rachis lombaire en vue latérale ($\mathfrak{L}^{\mathfrak{l}}$)
- Pelvis en vue de face ($\mathfrak{P}^{\mathfrak{f}}$)
- Genoux en vue de face ($\mathfrak{K}^{\mathfrak{f}}$)

Une note entre 0 et 5 quantifiant le degré de visibilité a été associée à chaque structure. Les résultats dans la table 4 montrent que la méthode proposée apporte une amélioration sensible sur la qualité des images dans toutes les régions anatomiques. Les résultats sont similaires uniquement dans les genoux car les rayons X ne sont guère absorbés. Néanmoins, nous remarquons que dans certaines régions, comme par exemple le rachis thoracique, certaines structures restent difficiles à voir au-delà de l'approche choisie pour post-traiter les images.

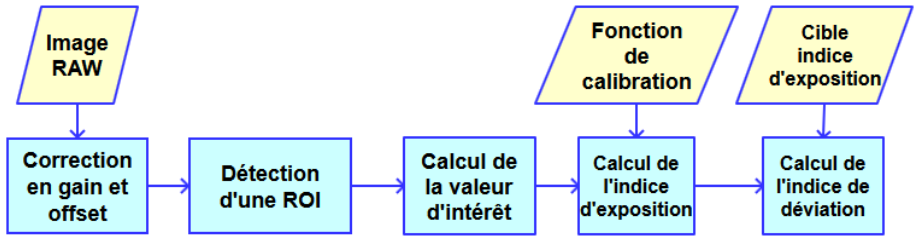
Conclusion

La méthode proposée par carte de bruit (LNCE) est donc la plus appropriée pour traiter des images diagnostiques. En effet, LNCE permet d'améliorer la qualité des images à l'affichage par rapport à une méthode qui aujourd'hui est utilisée en routine clinique. De plus, aucune perte d'information n'a été constatée sur les images post-traitées avec LNCE, alors que ce n'est pas le cas si l'image d'entrée est débruitée avant de la rehausser en contraste (méthode DE). Certains tests sur des images de fantôme montrent que l'approche est également prometteuse en termes de réduction de la dose. Cependant, comme la qualité des images dépend également des paramètres d'acquisition, la visibilité de certaines structures reste de faible niveau indépendamment de l'algorithme choisi. Il est donc important de quantifier la qualité des images dès l'acquisition, ce qui est traité dans les sections suivantes.

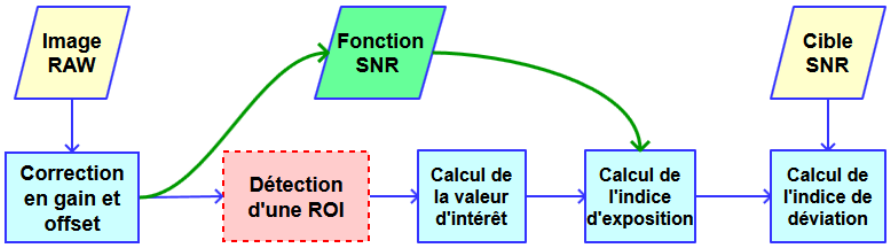
Points de repère pour l'estimation de l'indice d'exposition

Il est utile pour un utilisateur d'être capable de contrôler que la quantité de rayons X émis par un système de radiologie digital soit convenable au sens ALARA. En effet, comme le post-traitement permet d'équilibrer les niveaux de gris dans l'histogramme, les effets de saturation sont cachés. Néanmoins, la force du bruit sur les images change en fonction de la dose de rayonnement. Ainsi une mesure a été standardisée dans le but de fournir un retour aux utilisateurs de systèmes de radiographie digitale. Il s'agit de l'indice d'exposition (EI) qui mesure la quantité de signal qu'arrive au détecteur (IEC 62494-1, 2008).

La figure 13a montre la séquence d'étapes qui permet d'obtenir l'EI comme expliqué dans la norme IEC 62494-1 (2008). Les images acquises passent d'abord par une étape de calibration qui consiste à corriger l'image en gain et offset. Il faut ensuite déterminer une région d'intérêt (ROI) qui soit pertinente selon la finalité clinique de l'examen. À partir de cette ROI, une valeur d'intérêt (VOI) est calculée, par exemple comme la valeur médiane de la distribution des niveaux de gris à l'intérieur de la ROI. Cette VOI est ensuite associée à une valeur d'EI en utilisant une fonction de calibration qui dépend du système de rayons X. La valeur d'EI n'est pas très significative dans l'absolu, mais plutôt si on la compare à une valeur de référence



(a)



(b)

FIGURE 13 : *Algorithmes pour quantifier la qualité des images acquises : (a) indice d'exposition comme proposé dans IEC 62494-1 (2008) ; (b) méthode proposée qui s'appuie sur des valeurs de SNR anatomiques.*

qui est censée représenter la qualité des images ALARA (EI_t). Par conséquent, une valeur d'indice de déviation (DI) est calculée de la manière suivante :

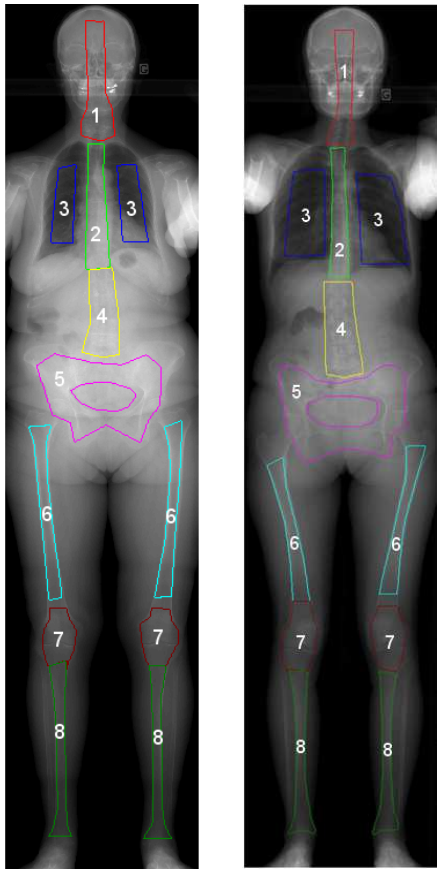
$$DI = 10log_{10}\left(\frac{EI}{EI_t}\right) \tag{9}$$

Nous proposons deux modifications à cet algorithme. En première lieu, nous exploitons l'algorithme d'estimation du percentile pour estimer la puissance du bruit et, implicitement, le SNR[4]. Cela permet d'éviter la modélisation de la fonction de calibration et, donc, d'enlever une entrée de l'algorithme. En deuxième lieu, nous analysons comment définir la ROI. Comme les images EOS sont pour une grande partie des corps entier il n'est pas évident de répondre à cette question. En effet, la valeur de l'EI dans la région des poumons est typiquement 5 fois plus élevée que celle dans le pelvis. Il n'est pas alors suffisant d'utiliser une seule valeur pour des images du corps entier, mais il faut associer un indice à chaque région anatomique. Dans les sections suivantes nous étudions quel type d'approche il est préférable de choisir et pour quelles raisons.

Comparaison de méthodes de définition de ROI et détection

Pour une image du corps entier donnée, nous définissons les huit ROIs suivantes (voir figure 14) :

[4]Par la suite, on parlera de EI mais on utilisera la mesure reposant sur SNR.



(a) (b)

FIGURE 14 : *Vérité terrain sur les valeurs d'EI. Deux exemples de segmentation manuelle sur des images du corps entier. Pour chaque région la VOI est calculée comme la médiane des niveaux de gris dans les régions définies à la main.*

- Tête et rachis cervical ($\mathcal{A}_1$, rouge)
- Rachis thoracique ($\mathcal{A}_2$, vert)
- Poumons ($\mathcal{A}_3$, bleu)
- Rachis lombaire ($\mathcal{A}_4$, jaune)
- Pelvis ($\mathcal{A}_5$, magenta)
- Fémurs ($\mathcal{A}_6$, cyan)
- Genoux ($\mathcal{A}_7$, rouge foncé)
- Tibias ($\mathcal{A}_8$, vert foncé)

Les valeurs d'EI associées aux différentes régions sont ainsi calculées. Par conséquent une image du corps entier sera décrite par un ensemble de huit valeurs d'EI au lieu d'une seule. Nous faisons l'hypothèse que les mesures issues de ROIs segmentées à la main donnent la vérité terrain.

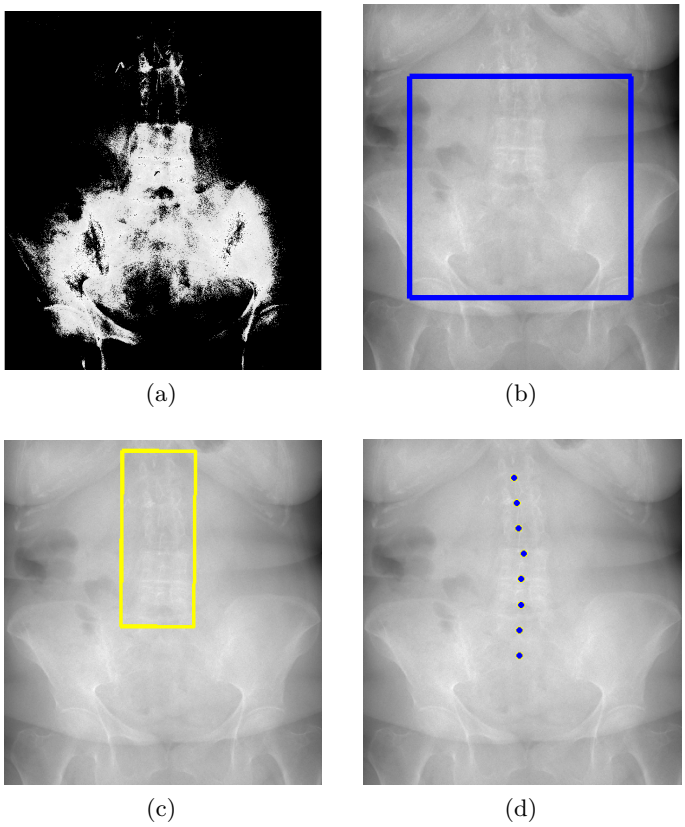
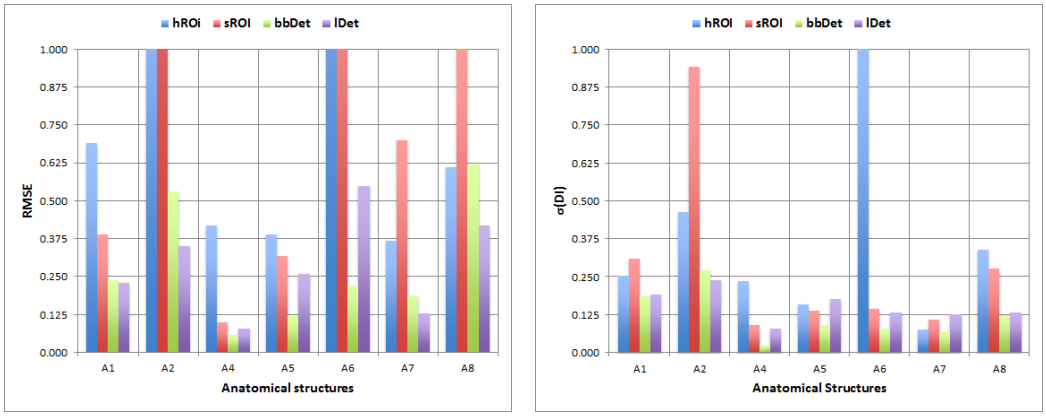


FIGURE 15 : *Méthodes pour définir automatiquement des ROIs pour l'estimation de l'EI : (a) mesures à partir de l'histogramme (hROI) ; (b) centre de l'image (sROI) ; (c) détection de boîtes englobantes (bbDet) ; (d) détection de points de repère (lDet).*

Nous comparons quatre méthodes pour définir des ROIs à partir desquelles estimer des valeurs d'EI. La meilleure approche est celle qui fournit des mesures les plus proches possibles de la vérité terrain. Une des méthodes proposées par Shepard et al. (2009) consiste à utiliser des statistiques calculées à partir de l'histogramme (*hROI*). En particulier, il s'agit de définir deux valeurs de percentile de l'histogramme ($[1\%, 25\%]$ par exemple) et de considérer l'ensemble des pixels entre ces deux valeurs. La figure 15a montre un exemple sur la région abdominale. Une autre approche proposée par Shepard et al. (2009) s'appuie sur l'hypothèse que l'information la plus pertinente est au centre de l'image (voir figure 15b). Dans nos tests nous avons utilisé un carré qui recouvre les 50% du champ de vue restreint à la région anatomique d'intérêt. En alternative aux méthodes qui existent dans l'état de l'art nous proposons de détecter les structures anatomiques auxquelles on veut associer une valeur d'EI. On peut ainsi positionner des boîtes englobantes (voir figure 15c, *bbDet*) ou bien des points de repère (voir figure 15d, *lDet*).

La méthode par points de repère proposée s'appuie sur une analyse locale utilisant une représentation par patchs. Nous considérons des patchs circulaires P_j centrés sur les points de repère l_j. Une mesure d'EI est associée à chaque patch et cela donne un ensemble de mesures locales $e(l_j)$. Ensuite la valeur d'EI d'une région



(a) (b)

FIGURE 16 : *Performances des méthodes d'estimation automatique des valeurs d'EI : (a) racine de l'erreur quadratique moyenne ; (b) écart type de l'erreur. L'erreur est calculée en utilisant l'équation 12.*

anatomique $\mathcal{A}_r$ est obtenue par moyenne pondérée des estimations locales :

$$\mathrm{EI}_r = \frac{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j) e(l_j)}{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j)} \tag{10}$$

où $\mathcal{L}_r$ est l'ensemble des points de repère attribués à la région anatomique $\mathcal{A}_r$ et $\omega(l_j)$ sont les poids qui pondèrent la contribution des mesures locales $e(l_j)$ qui sont définis de la manière suivante :

$$\omega(l_j) = \exp\left(-\frac{H(l_j)}{\alpha_H}\right) \tag{11}$$

où $H(l_j)$ est l'entropie de la distribution des niveaux de gris à l'intérieur d'une patch P_i et α_H est un paramètre dont la valeur est fixée dans tous les tests à 2. Nous attribuons plus d'importance aux distributions à faible entropie parce que cela veut dire qu'une seule contribution de signal est retenue par la mesure locale.

Ces quatre méthodes sont ainsi comparées pour établir celle qui s'approche le plus des valeurs obtenues à partir de la segmentation manuelle. Pour quantifier cette information nous utilisons une mesure d'erreur calculée de la manière suivante :

$$\mathrm{DI} = 10 \log_{10}\left(\frac{\mathrm{EI}}{\mathrm{EI}_{gt}}\right) \tag{12}$$

Nous utilisons donc la formule de l'indice de déviation mais en considérant la vérité terrain comme valeur de référence. À partir d'un ensemble d'images représentatif du type d'images à traiter, nous calculons la racine de l'erreur quadratique moyenne (RMSE, figure 16a) et l'écart type (σ(DI), figure 16b) pour toutes les régions $\mathcal{A}_r$ prises en compte. Les résultats montrent que les approches utilisant sur détection automatique permettent de diminuer les erreurs par rapport aux méthodes de l'état de l'art. De plus, les mesures DI(σ) les estimations sont cohérentes sur différents examens.

TABLE 5 : *Étude sur la robustesse aux erreurs de positionnement pour les méthodes utilisant la détection. Ce tableau montre jusqu'à quel distance en mm l'erreur sur l'estimation de l'EI (équation 12) peut être considéré négligeable, c'est-à-dire DI* $< 0,25$.

	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$
bbDet	10	$<$ **5**	< 5	30	15	5	5	30
lDet-100%	30	**10**	10	30	30	15	15	10
lDet-50%	30	20	**15**	30	30	30	30	20
lDet-25%	30	30	**20**	30	30	30	30	30

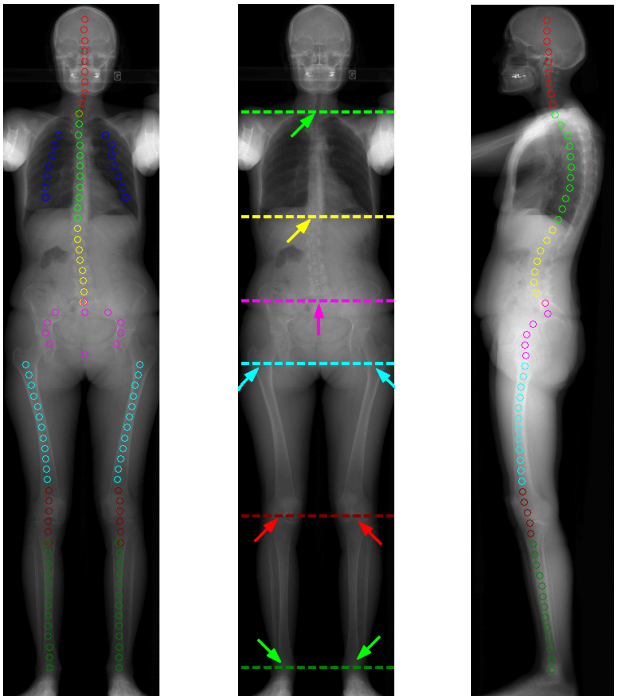
Les méthodes *bbDet* et *lDet* ont des performances similaires si les positions idéales des boîtes englobantes et des points de repère sont utilisées. Il faut donc se poser la question de ce qu'il se passe quand des erreurs sur les positions des objets sont introduits. Pour répondre à cette question nous simulons des erreurs entre 5 et 30 *mm* sur la position des boîtes englobantes (*bbDet*) ou bien des points de repère (*lDet*-100%) pour ensuite mesurer l'indice de déviation, donc l'erreur, par rapport à la valeur d'EI obtenue en considérant les positions correctes. De plus, nous considérons le cas où seulement la moitié (*lDet*-50%) ou un quart (*lDet*-25%) des points de repère sont mal placés. La table 5 montre que la méthode utilisant les points de repère est beaucoup plus robuste à d'éventuelles erreurs de détection, et les indices de déviation ne sont hauts qu'à des distances significatives lorsqu'une partie des estimations locales est correcte. Cela est lié au fait que l'approche par points de repère est redondante et donc,les estimations locales qui ne sont pas correctes seront implicitement écartées.

Ces considérations nous amènent à la conclusion que la définition d'un ensemble de valeurs d'indice d'exposition pour des images EOS du corps entier doit passer par la détection d'ensembles de points de repère. Nous abordons cette question dans la suite.

Détection de points de repère

Plusieurs aspects rendent la tache de détection de points de repère sur une image EOS difficile. En premier lieu, le nombre de cas différents à traiter est très vaste. Par exemple, il faut prendre en compte aussi bien des images diagnostiques que des images de pré-affichage. En deuxième lieu, la détection de structures anatomiques sur des images à rayons X est compliquée parce que la distribution des niveaux de gris pour un objet donné n'est pas homogène en raison de la superposition entre tissus anatomiques de densités différentes. En dernier lieu, l'espace des solutions possibles est très grand, notamment si l'image à traiter couvre le corps entier.

Nous identifions un ensemble de points, que nous appelons points *contrôle*, $c_j \in \mathcal{C}$, qui permettent de diviser l'image en plusieurs tranches sur l'axe vertical. Autrement dit, nous voulons estimer les coordonnées y_j des points de contrôle $c_j \in \mathcal{C}$. Pour atteindre cet objectif, nous définissons des facteurs de proportionnalité à partir des points annotés à la main sur une base de données composée d'une dizaine de patients qui représentent bien la variabilité de la totalité des images à disposition pour la validation (82 patients). En pratique, nous faisons l'hypothèse que les valeurs de y_1



(a) (b) (c)

FIGURE 17 : *Points de contrôle identifiés sur une image de face (a) et associés aux structures suivantes : T1 (c_1 vert) ; T12 (c_2 jaune) ; L5 (c_3 magenta) ; fémur proximal (cyan) ; genoux (rouge) ; chevilles (vert). Cette initialisation est également valide sur l'image de profil correspondante (c).*

et y_6 sont connues, et le facteur de proportionnalité attribué à y_4 sera alors égal à :

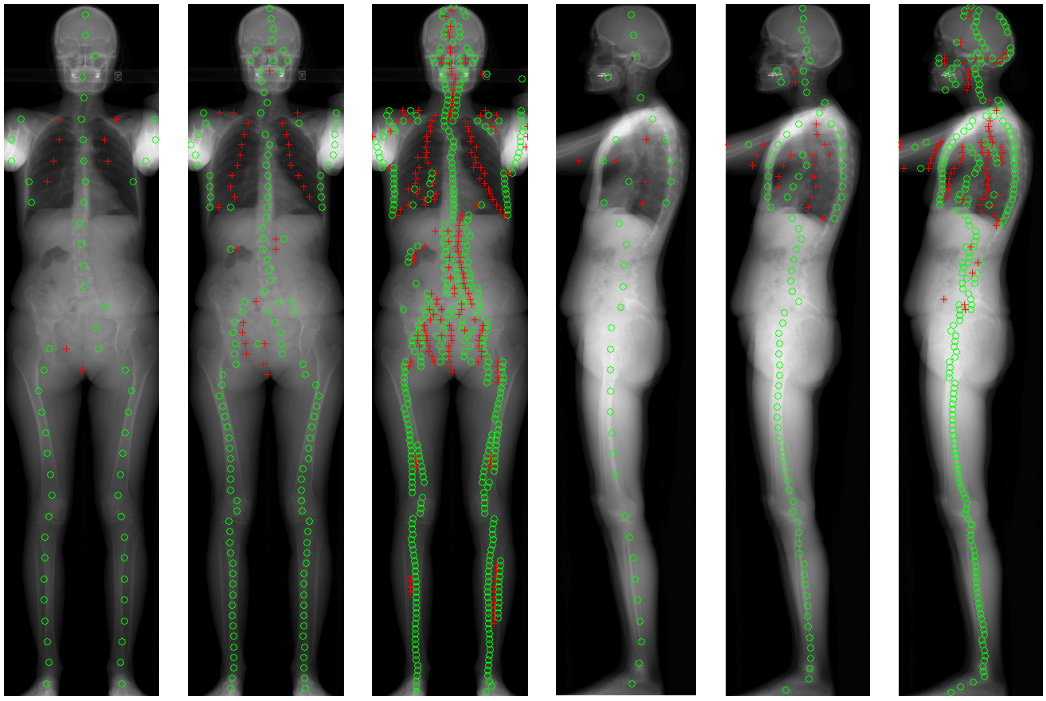
$$\varrho_4 = \frac{y_4 - y_1}{y_6 - y_1}. \tag{13}$$

Cela est fait pour toutes les images de la base de données pour ensuite prendre la valeur minimale et maximale des facteurs de proportionnalité et calculer les extrêmes α_j et β_j des facteurs de proportionnalité :

$$\alpha_j = \min(\varrho_j) - \sigma(\varrho_j) \qquad \beta_j = \max(\varrho_j) + \sigma(\varrho_j). \tag{14}$$

Par conséquent, il est possible de réduire de manière séquentielle l'espace de recherche des points de contrôle ainsi que des points de repère. La détection des points de contrôle initiaux et des positions précises des autres, à savoir ceux associés à T1 et aux chevilles, est conduite en régularisant des points saillants extraits sur l'image à traiter. Les techniques proposées pour la régularisation font partie des contributions de ces travaux de thèse et le lecteur pourra trouver la description détaillée dans la version intégrale du manuscrit (chapitre 5). Dans cette synthèse nous nous limitons à décrire comment les points saillants sont obtenus.

L'attention d'un utilisateur qui regarde une image à rayons X est essentiellement capturée par des variations de signal qui correspondent à des changements de densité des tissus traversés par les rayons X. Les points saillants sont positionnés au centre



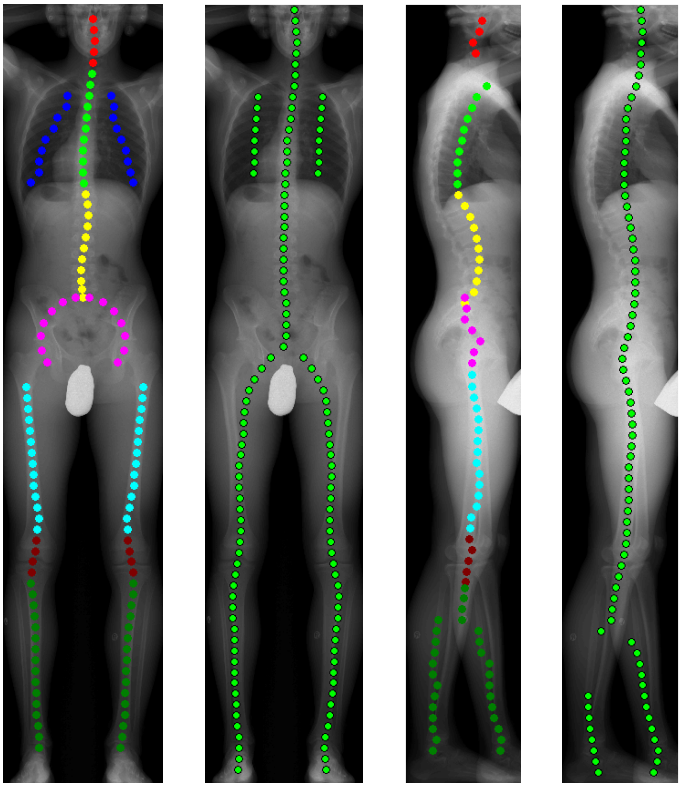
(a) (b) (c) (d) (e) (f)

FIGURE 18 : *Ensembles de points saillants positifs $\mathcal{P}_{(l,s)}$ (rouge) et négatifs $\mathcal{P}_{(h,s)}$ (vert) sur des images en vue de face aux échelles (a) $S = 256$; (b) $S = 128$; (c) $S = 64$ and sur les vues latérales associées (d) $S = 256$; (e) $S = 128$; (f) $S = 64$.*

de fenêtres carrées de côté égal à S dont les niveaux moyens de signal correspondent à des pics positifs (signal au détecteur) ou négatifs (absorption). La figure 18 montre les points obtenus pour un patient, sur des images en vues de face et de profil, pour différentes valeurs de S. La figure 18 montre que la représentation est de plus en plus grossière dès que S augmente. Le paramètre S peut donc être considéré comme un facteur d'échelle. Pour des basses valeurs de S la représentation est plus complète mais il y a plus de valeurs aberrantes, c'est-à-dire des points qui ne sont pas associés à des structures anatomiques d'intérêt. Cependant, si ce paramètre est bien fixé, les points saillants correspondent bien aux structures osseuses pour l'ensemble des pics d'absorption $\mathcal{P}_{(h,s)}$ et aux poumons pour celui relatif aux pics de signal $\mathcal{P}_{(l,s)}$.

Une fois le problème initialisé grâce aux points de contrôles et saillants, les ensembles de points de repère anatomiques sont détectés en utilisant une approche séquentielle. Nous proposons d'exploiter des méthodes adaptées à chaque région anatomique à traiter pour obtenir les points de contrôle et de repère à partir des points saillants. Il faut remarquer qu'une fois détectés les points de repère dans une région anatomique, l'espace de recherche dans les autres régions est significativement réduit. Autrement dit, la méthode séquentielle permet de simplifier progressivement le problème au fur et mesure que les ensembles de points de repère sont localisés. Les détails sur ces aspects ainsi qu'une validation qualitative sur la précision de détection sont largement traités dans la version intégrale du manuscrit (chapitre 5). Nous résumons dans les sections suivantes les résultats sur les valeurs d'EI estimées.

Validation des valeurs d'indice d'exposition automatiques



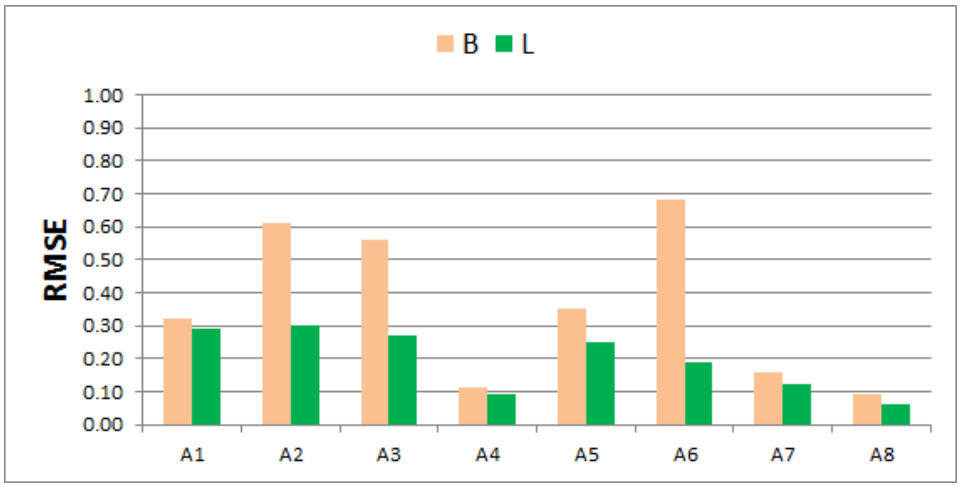
(a) (b) (c) (d)

FIGURE 19 : *Exemple de points de repère détectés automatiquement : (a) Points $\mathcal{L}$ sur une image en vue de face ; (b) Points $\mathcal{B}$ sur une image en vue de face ; (c) Points $\mathcal{L}$ sur une image en vue de profil ; (d) Points $\mathcal{B}$ sur une image en vue de profil.*

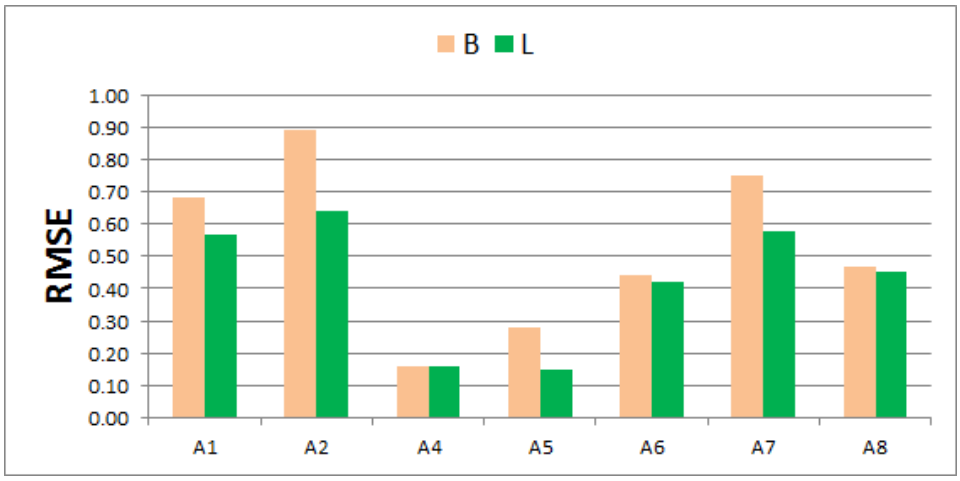
Nous comparons les valeurs d'EI obtenues à partir des points $\mathcal{L}$, qui sont positionnés en utilisant la méthode proposée, à celles extraites à partir des ensembles $\mathcal{B}$ qui sont placés par rapport à l'axe principal de l'image. La figure 19 montre un exemple des points de repère $\mathcal{L}$ et $\mathcal{B}$ résultants pour les images frontale et latérale du même patient. Ces deux techniques de détection sont comparées sur une base de données constituée de 82 examens du corps entier. Nous pouvons aussi y distinguer les trois sous-ensembles suivants :

- $\mathcal{D}_a$: examens diagnostiques de patients adultes d'âges compris entre 54 et 82 ans.
- $\mathcal{D}_b$: examens diagnostiques de patients adolescents ou enfants d'âges compris entre 8 et 16 ans.
- $\mathcal{D}_c$: examens de pré-affichage pour les patients dans le sous-ensemble $\mathcal{D}_b$.

La totalité de cas considérés comporte différents morphotypes, des malformations osseuses peuvent être présentes (scoliose par exemple), les patients ne sont pas forcément centrés par rapport à l'axe vertical de l'image et certaines acquisition sont



(a)



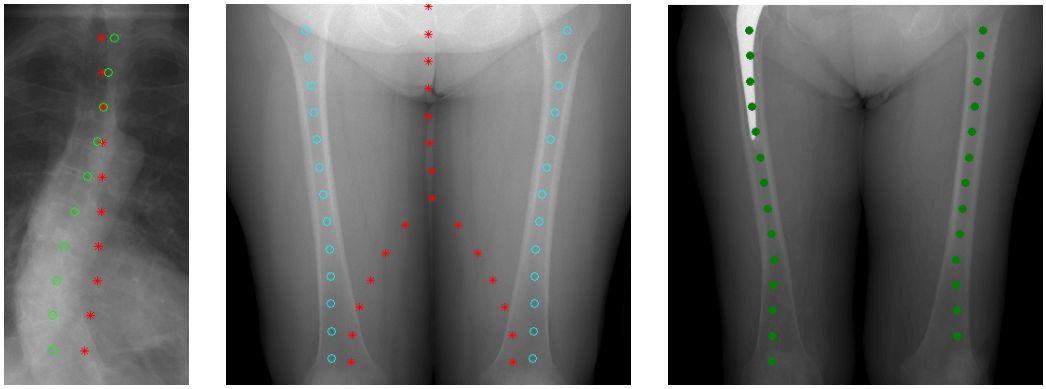
(b)

FIGURE 20 : *Erreurs sur l'estimation des valeurs d'EI en utilisant les points $\mathcal{B}_r$ (rose) et $\mathcal{L}_r$ (vert) pour les régions anatomiques sur des images en vue : (a) frontale ; (b) latérale.*

post-opératoires, ce qui implique la présence d'objets métalliques dans le champ de vue. La validation est menée en calculant le RMSE dont l'erreur est calculée suivant l'équation 12. Nous considérons d'abord la totalité des données et les sous-ensembles indiqués ci-dessus, pour ensuite se focaliser sur des cas particuliers et essayer de comprendre si de l'information significative peut être déduite à partir d'images de pré-affichage.

Les figures 20a et 20b reportent les mesures de RMSE obtenues en utilisant les points $\mathcal{B}_r$ et $\mathcal{L}_r$ pour des images en vues de face et de profil, respectivement. Les régions anatomiques sont les suivantes : tête et rachis cervical ($\mathcal{A}_1$), rachis thoracique ($\mathcal{A}_2$), poumons ($\mathcal{A}_3$), rachis lombaire ($\mathcal{A}_4$), pelvis ($\mathcal{A}_5$), fémurs ($\mathcal{A}_6$), genoux ($\mathcal{A}_7$) et tibias ($\mathcal{A}_8$). Sur des images acquises en vue de face (figure 20a) l'approche proposée, c'est-à-dire $\mathcal{L}_r$, permet de baisser significativement les erreurs des estimations par rapport à l'autre méthode qui prend uniquement les bords de l'enveloppe du patient

comme référence, c'est-à-dire $\mathcal{B}_r$. L'amélioration est particulièrement significative dans les régions $\mathcal{A}_2$, $\mathcal{A}_3$ et $\mathcal{A}_6$. De plus, en observant les résultats par type de données, nous en déduisons que la précision de l'estimation ne change pas pour différents morphotypes ou bien si les images d'entrée sont de pré-affichage. En ce qui concerne les images en vue latérale, la figure 20b indique que l'approche proposée est toujours significativement meilleure. Cependant, il est possible d'obtenir le même niveau de précision que sur les images en vue de face uniquement dans les régions $\mathcal{A}_4$ et $\mathcal{A}_5$. En effet, dans les trois régions des membres inférieures il n'est pas envisageable d'obtenir des valeurs d'EI cohérentes car les jambes sont superposées dans les images. Ensuite, dans certains cas, les images ne sont pas acquises à compter du sommet de la tête et donc le nombre de points de repère est trop faible pour donner une réponse précise dans la région $\mathcal{A}_1$. Finalement, les valeurs d'EI associées à la région $\mathcal{A}_2$ peuvent être influencées par la superposition des épaules avec la partie supérieure du rachis thoracique, ce qui peut causer une sous-estimation des valeurs d'EI. Dans ce cas il faudrait donner a priori plus d'importance aux estimations locales attribuées aux vertèbres dans la partie inférieure du rachis thoracique.



(a) (b) (c)

FIGURE 21 : *Cas particuliers : (a) scoliose ($\mathcal{L}_2$ points verts et $\mathcal{B}_2$ points rouges) ; (b) patient obèse ($\mathcal{L}_6$ points cyans et $\mathcal{B}_6$ points rouges) ; (c) présence d'objets métalliques.*

La figure 21 montre des exemples qui permettent d'évaluer comment la méthode proposée se comporte hors de cas conventionnels. La figure 21a montre que les points de repère $\mathcal{B}_2$ ne sont pas représentatifs lorsque la structure osseuse est déformée par une scoliose, alors que les points $\mathcal{L}_2$ suivent bien le rachis thoracique. L'avantage en terme de valeur d'EI est significative comme il est prouvé par les valeurs de DI égales à $1,36$ et $-0,16$ obtenues quand les points de repère $\mathcal{B}_2$ et $\mathcal{L}_2$ sont utilisés, respectivement. La figure 21b montre la région fémorale d'un patient dont l'indice de masse corporelle est très élevé. Dans ce cas les jambes du patient se touchent et donc la valeur d'EI est sous-estimée, comme la valeur de DI égale à $1,69$ indique. En revanche, la méthode de détection proposée ne dépend pas du morphotype du patient et les points de repère sont bien positionnés en correspondance des fémurs. Par conséquent, la valeur d'EI calculée à partir de $\mathcal{L}_6$ donne une estimation parfaite, à savoir une valeur de DI égale à $0,01$. Les objets métalliques posent un grand problème dans l'estimation de l'EI car ils ne sont pas associés à de l'information diagnostique et en même temps absorbent considérablement les rayons X. La formulation par points

de repère que nous proposons est bien adaptée à cette question. En effet, comme la figure 21c le montre, seulement une partie très réduite des estimations locales sera concernée par la présence d'objets métalliques. En outre, une contribution sera incorrecte seulement dans le cas où les points sont centrés dans l'objet, ce qui est rarement le cas. Pour ces raisons, dans l'exemple donné, les valeurs d'EI calculées pour le fémur à gauche et celui à droite sont pratiquement identiques (41, 24 et 40, 27).

TABLE 6 : *Mesures moyennes et maximales de la divergence de Kullback–Leibler (Δ) calculées entre des distributions de valeurs d'EI sur des images diagnostiques ($\mathcal{D}_b$) et les images de pré-affichage correspondantes ($\mathcal{D}_c$). Les valeurs d'EI sont calculées à partir de points de repère annotés à la main ($\mathcal{M}$) ou bien automatiquement en fonction de l'enveloppe du patient ($\mathcal{B}$) ou avec l'approche de détection proposée ($\mathcal{L}$).*

	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{M}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{B}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{L}(\mathcal{D}_c)$
average$(\Delta \cdot 10^3)$	$2,11$	$8,39$	$4,09$
max$(\Delta \cdot 10^3)$	$4,78$	$58,78$	$27,28$

Le dernier point sur lequel nous nous interrogeons est si les valeurs extraites des images de pré-affichage peuvent être utiles pour estimer des paramètres d'acquisition optimaux ou pas. Pour répondre à cette question nous considérons les distributions de valeurs d'EI, c'est-à-dire l'ensemble des valeurs normalisées, sur une image diagnostique et sur l'image de pré-affichage correspondante, et en calculons la distance en utilisant la divergence de Kullback-Leibler qui est définie de la manière suivante :

$$\Delta(X \mid Y) = \sum_r X(r) \log\left(\frac{X(r)}{Y(r)}\right). \tag{15}$$

La divergence de Kullback-Leibler est une mesure non symétrique qui quantifie la quantité d'information manquante lorsque Y est utilisé au lieu de X, où Y est une approximation de X. Par conséquent, nous voulons quantifier combien d'information est perdue lorsqu'on prend les valeurs d'EI calculées sur les images de pré-affichage à partir de points de repère annotés à la main ($\mathcal{M}$), détectés automatiquement avec l'approche basique ($\mathcal{B}$) ou bien celle proposée ($\mathcal{L}$), par rapport aux valeurs d'EI calculées sur les images diagnostiques à partir de points de repère annotés à la main ($\mathcal{M}$) La table 6 montre les résultats de ce test. Nous en concluons que les images de pré-affichage contiennent effectivement une information suffisante pour prévoir les paramètres d'acquisition optimaux au sens ALARA et que la méthode de détection proposée n'amène pas à une perte d'information significative par rapport aux points annotés à la main. Cette dernière observation montre empiriquement qu'il est possible d'exploiter les valeurs d'EI calculées automatiquement à partir d'images de pré-affichage en entrée d'une boucle d'asservissement en charge de définir le niveau de dose optimal pour l'acquisition suivante utilisée pour faire du diagnostique.

Conclusion et perspectives

Dans la première partie de la thèse nous avons étudié comment optimiser la qualité des images acquises par le système de radiologie EOS en gérant l'équilibre entre augmentation de la visibilité de détails anatomiques et réduction de bruit. Nous

avons ainsi proposé des solutions différentes dans le but de trouver celle qui s'adapte le mieux à une analyse qualitative de type diagnostique de la part de l'utilisateur. Dans cette recherche nous avons apporté plusieurs contributions originales telles que l'extension d'un filtre par moyennes non locales pour le débruitage d'images à rayons X, l'estimation de niveaux de bruit à partir d'images à rayons X et la formulation de cartes non paramétriques qui sont exploitées pour limiter l'augmentation de bruit lors des rehaussement contraste. La solution reposant sur ces cartes de bruit est celle qui est la plus adaptée dans le cadre clinique tout en étant prometteuse pour la réduction de la dose.

Dans la deuxième partie de la thèse nous avons affronté le problème de l'estimation de mesures de la qualité des images pour pouvoir à la fois établir si une image a été acquise en suivant le principe ALARA et, dans le cas où ces mesures sont extraites à partir d'images de pré-affichage, pour obtenir de l'information capable de guider le paramétrage de l'acquisition. Nous avons proposé une nouvelle formulation de l'algorithme d'indice d'exposition à partir de la détection de points de repère. Ce choix a été justifié par une analyse détaillée qui a permis de montrer que cette approche est une solution décisivement plus robuste que d'autres méthodes suggérées dans la littérature. Nous avons ensuite proposé une méthode originale pour détecter ces points de repère sur des images EOS du corps entier. Les résultats montrent que l'approche proposée permit d'estimer de manière complètement automatique si un patient a reçu une quantité de radiations correcte ou pas. De plus, des études sur les images de pré-affichage indiquent que cette méthode pourrait être utilisée pour obtenir les entrées d'une boucle d'asservissement en charge de régler automatiquement le niveau de dose.

En perspective sur les aspects de post-traitement des images, il serait intéressant de prendre en compte le fait que le bruit dépend des échelles alors que les approches par patchs enlèvent uniquement la composante de haute fréquence du bruit. Il serait donc intéressant de tester des approches par patchs qui débruitent à plusieurs échelles. Cela pourrait permettre d'estimer des cartes de bruit plus robustes notamment dans le cas d'acquisitions à très faible dose. Ensuite, nous pourrions tester des décompositions multi-échelles qui représentent mieux l'information affectée par du bruit telles que, par exemple, les ondelettes complexes. Finalement, la validation clinique a été amplement traitée dans nos travaux mais il faudrait avoir plus de retours de la part de cliniciens dans la validation.

En ce qui concerne la méthode de détection des points de repère, il serait intéressant de considérer d'autres cas que celui des images du corps entier. Dans ce cas le nombre de points de contrôle à détecter changerait et, pour éviter de s'appuyer sur une hypothèse a priori sur le contenu de l'image, il faudrait utiliser des approches de localisation supervisées pour trouver les bons points de contrôle. Néanmoins, cela n'est pas trivial parce que la définition de descripteurs robustes est difficile en raison du grand nombre de cas à traiter. Il serait également intéressant de tester des méthodes de recherche séquentielle adaptative, donc pas dans un ordre préfixé, et qui prennent en compte un contrôle rétroactif pour éviter la propagation d'erreurs. Finalement, la méthode nécessite d'être validée en routine clinique.

Abstract

We aim at reducing the ALARA (As Low As Reasonably Achievable) dose limits for images acquired with EOS full-body system by means of image processing techniques. Two complementary approaches are studied.

First, we define a post-processing method that optimizes the trade-off between acquired image quality and X-ray dose. The Non-Local means filter is extended to restore EOS images. We then study how to combine it with a multi-scale contrast enhancement technique. The image quality for the diagnosis is optimized by defining non-parametric noise containment maps that limit the increase of noise depending on the amount of local redundant information captured by the filter.

Secondly, we estimate exposure index (EI) values on EOS images which give an immediate feedback on image quality to help radiographers to verify the correct exposure level of the X-ray examination. We propose a landmark detection based approach that is more robust to potential outliers than existing methods as it exploits the redundancy of local estimates.

Finally, the proposed joint denoising and contrast enhancement technique significantly increases the image quality with respect to an algorithm used in clinical routine. Robust image quality indicators can be automatically associated with clinical EOS images. Given the consistency of the measures assessed on preview images, these indices could be used to drive an exposure management system in charge of defining the optimal radiation exposure.

Contents

Chapter 1

Introduction

This chapter presents the context of the thesis, the motivation of our research and the related challenges.

1.1 EOS: a biplanar low-dose X-ray unit

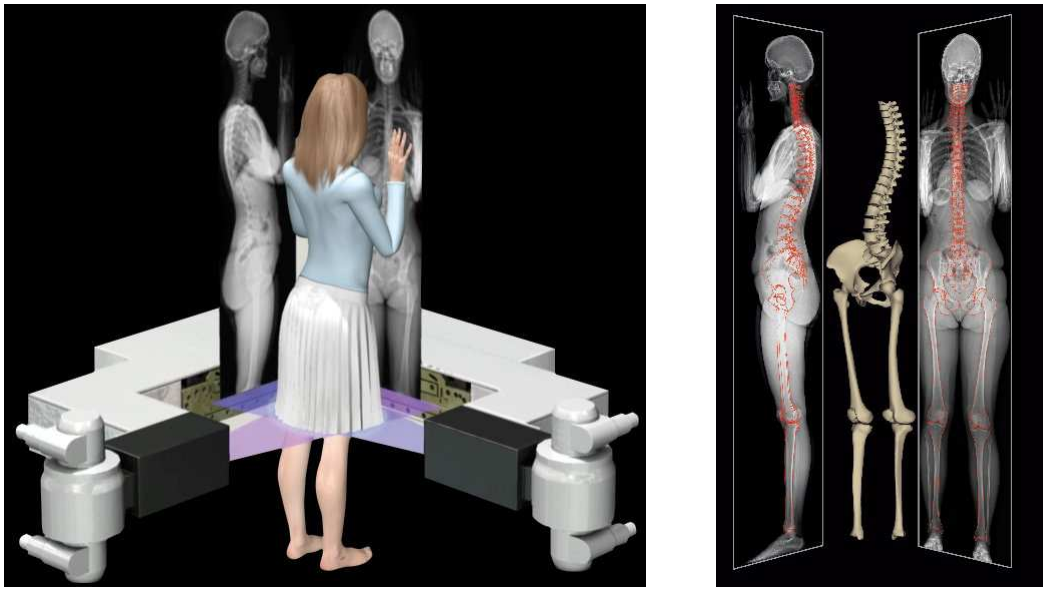
EOS is a Stereo-Radiographic imaging system based on slot scanning principles that is produced by the French manufacturer *EOS imaging*. This is the only musculoskeletal system that allows ultra-low dose, simultaneous acquisition of full body frontal and lateral images in weight-bearing position.

1.1.1 Global architecture and properties

EOS acquisition principle is based on the multiwire gaseous detector that has been conceived from Georges Charpak's[1] researches (Charpak et al., 1968). This detector is capable of extracting a significant amount of information from a single photon while being practically unaffected by scattering, which makes it adapted for X-ray image acquisitions. The detector amplifies the X-rays that pass through the body of a scanned patient by means of electronic avalanche in the gas. In medical applications, the amplification of the detector is crucial as it allows obtaining a high quality X-ray image while keeping the dose low. The detector is composed by 1764 strips focused on the X-ray focal spot, i.e. parallel to the incident X-ray photon trajectories. These channels are placed at 254 μm wide intervals, i.e. the distance that corresponds to the spatial resolution within the horizontal plane. Please refer to Damet et al. (2014) for a quantitative evaluation of the detector performances.

EOS system occupies a 2 m^2 sized area and measures 2.70 m in height. The patient is located in this area in a weight-bearing standing position, or, less often, in a seated position. The possibility of examining the patient in its natural standing position gives a better insight into possible diseases or deformations of the musculoskeletal system than computed tomography (CT) imaging (Dubousset et al., 2005). Figure 1.1a shows how the EOS images are acquired. A c-arm scans the entire body of a patient undergoing an examination by covering, at most, a 45 cm wide and 180 cm high field of view in 23 s on average. The X-ray tube emits a thin collimated X-ray fan-beam that is detected line by line. Two couples of co-linked X-ray tubes

[1] 1992 physics Nobel Price winning.



(a) (b)

Figure 1.1: *EOS: (a) acquisition principle; (b) 3D model from the frontal and lateral view acquisitions.*

and detectors are placed orthogonally to each other on the c-arm and, thus, two images are simultaneously acquired: the frontal (AP) and lateral (LAT) view acquisitions. The two images are then sent to two separated image processing chains. Given this acquisition procedure, EOS is also known as bi-planar X-ray full-body system.

Figure 1.1b shows that the frontal and lateral view acquisitions can be used to extract a personalized 3D model of the skeleton, which is used to obtain some relevant 2D/3D clinical parameters. This task is achieved by using an *EOS imaging* proprietary software called *sterEOS*. It requires the user to identify a limited number of anatomical landmarks. The reconstruction is then completed by exploiting a database of anatomical 3D shapes.

The image quality and radiation exposure depend on how the following parameters are set:

- Peak kilo voltage applied to the X-ray tube (kV). It determines the highest energy of the photons.
- X-ray tube current (mA). The higher mA the bigger the flux of photons is, while the spectrum is not modified.
- The scanning speed (C). It defines the time required to acquire an image. There are 8 speed grades, $C \in \{1, 2, \ldots, 8\}$, and by using the default value ($C = 4$) the scanning speed is 7.6 cm/s. The lowest and fastest scanning speeds are 30 cm/s ($C = 1$) and 3.4 cm/s ($C = 8$), respectively. The lower the scanning speed the lower the noise is, but the dose also increases. Therefore in some cases, e.g. pediatric examinations, it is preferable to reduce the values of C.

Default parameter settings are provided in order to optimally achieve a given clinical task depending on the anatomical areas in the image and the morphotype of the patient. Finally, an acquired image is given as input to the image processing chain that improves its visual rendering to facilitate the diagnosis, and is transmitted as a DICOM image to sterEOS software and to Picture Archival and Communication systems (PACS).

1.1.2 Examples of applications

EOS is mainly a diagnostic X-ray system, i.e. it allows for a morphological analysis of bone structures and lesions. Furthermore, sterEOS provides clinical parameters that help in the quantitative analysis of disorders concerning the skeleton. The image quality has to be good enough to allow the user placing anatomical landmarks in a precise and consistent way, i.e. by maximizing the inter-user consensus. We summarize some applications that exploit these clinical parameters in studies of the spine, the lower limbs and the full body. Please refer to the provided references for detailed descriptions of the presented medical studies.

1.1.2.1 Spine

Somoskeöy et al. (2012) have evaluated the coronal and sagittal spine curvatures by means of Cobb angle measurements (Vrtovec et al., 2009). The purpose of the study was to compare EOS measures with conventional manual annotations. According to the assessments by three experienced users on a database composed by 201 patients, this study shows that sterEOS 3D spinal curvature measurements are accurate and reproducible, without being hindered by drawbacks and errors of conventional 2D measurement methods. This work also highlights that the repeatability of spinal curvature measurements is less limited by scoliotic curve magnitude and sagittal plane deformities which is a key value of measuring in a 3D space. Furthermore, as shown in other studies (Ilharreborde et al., 2013; Illés et al., 2011), the 3D-based curvature parameters are also very useful for evaluating the clinical outcome of a surgery. Finally, Wybier and Bossard (2013) have highlighted that EOS provides higher image quality in the pelvic region acquired from the lateral view with respect to classical radiography, which allows better assessing the pelvic-spinal balance.

1.1.2.2 Lower Limbs

On the lower limbs, EOS can be used to perform goniometry, i.e. knee measurements, coxometry, i.e. hip measurements, hip replacement planning, and so on. Buck et al. (2012) have studied the interchangeability of femoral and tibial torsion measurements computed over EOS and CT scans on patients with osteoarthritis of the knee. The tests conducted on a database composed of 35 patients have confirmed this hypothesis, which has enabled a significant reduction of the exposure for this application. Polkowski et al. (2012) have shown that, by acquiring images in standing positions, EOS allows accounting for the inclination of the acetabulum (cotyloid cavity), which is not possible with CT scans. This information is relevant for preventing potential complications following a total hip arthroplasty. Than et al.

(2012) have studied how hip or knee arthritis influence orthopedic lower limbs deformities by relying on sterEOS hip and knee reconstructions. The patients not affected by arthritis have not manifested deformities, whereas varus and valgus alignments have been observed on the other ones.

1.1.2.3 Full body

Full body EOS images are acquired not only if both the spine and the lower limbs need to be diagnosed, but also to understand how disorders in the spine or in the lower limbs influence the posture of the patient (Lazennec et al., 2011). Following this principle, Le Huec et al. (2011) have proposed a formula that relies on the positions of a set of control points placed all over the body, in order to help defining the surgical planning for the correction of spine imbalance.

1.2 ALARA principle

In the previous section we have presented EOS as a low dose X-ray device. The term dose is clarified and the ALARA principle is introduced here as being fundamental to understand the context of this thesis.

1.2.1 Dosimetric quantities and units

The exposure is a dosimetric measure that refers to ionizing electromagnetic radiations. The amount of energy delivered to a tissue that interacts with ionizing radiations is so high that electrons are liberated from the atoms. Therefore, the tissues can be considerably damaged. The exposure is measured in coulombs per kilogram (Ckg^{-1}) or in roentgens (R).

The exposure is proportional to the absorbed dose that is defined as the energy absorbed per unit mass and is, hence, measured in joules per kilogram (Jkg^{-1}), which are better known as grays (Gy). In medical practice, the absorbed dose depends on the acquisition parameters (see Section 1.1.1), the material used to filter the X-ray beam and the field size. The proportionality factor f between the exposure and the dose depends on the material and on the kV. For example, by considering bone tissues, $f(10\ kV) = 3.5$ and $f(100\ kV) = 1.5$. The exposure is often linked to the dose absorbed by the air, which is also called air kerma (kinetic energy released in a material) as in diagnostic radiology the kerma and the absorbed dose coincide. In this manuscript we will use the kerma in the air as measure of dose.

The equivalent dose is another important quantity. It is proportional to the absorbed dose as a function of the biological effectiveness of a given radiation and is measured in sieverts (Sv). However, in radiology the proportionality factor among the two quantities is equal to one, i.e. they are identical. The equivalent dose is modified by a factor called detriment, that is the stochastic combination of the probability and extent of harming associated with a given tissue. Therefore, the equivalent dose is more relevant from a clinical point of view.

Please refer to Gofman (1981); Kase et al. (2012) for further details on the presented dosimetric quantities and for a complete survey on this subject.

1.2.2 Dose related risks and ALARA principle

Smith-Blindman et al. (2012) have recently presented a statistical study that covers a 15 years period of imaging use in health care facilities of North America. This study analyzes in detail the trends of exposure to artificial radiations. The results indicate that the number of standard radiography examinations remained relatively stable (1.2% annual growth), whereas the number of advanced diagnostic acquisitions significantly increased. In particular, the CT examinations tripled between 1996 and 2010, which increased the amount of dose absorbed by the patients undergoing examinations. Quantitatively, the per capita effective dose by year reached 2.5 mSv. The proportion of patients exposed to high ($> 20 - 50$ mSv) and to very high (> 50 mSv) radiations increased from 1.2% and 0.6% in 1996 to 2.5% and 1.4% in 2010, respectively. Moreover, it was observed that the amount of radiation increased with age: 20% of the patients beyond 45 years old annually received high or very high radiations. The reduction of exposure to X-rays is an important matter regardless the age of the patients because, as shown by Shuryak et al. (2010), the risk of developing radio-induced cancer does not decrease with the age. Specific precautions must, however, be taken in pediatric examinations because the patients statistically have longer lifetime to develop radio-induced cancers. In particular, a study by Ronckers et al. (2008) on women that underwent repeated X-ray examinations during their adolescence indicates that scoliotic women have a probability of developing a breast cancer that is two times higher than those not suffering from this disorder.

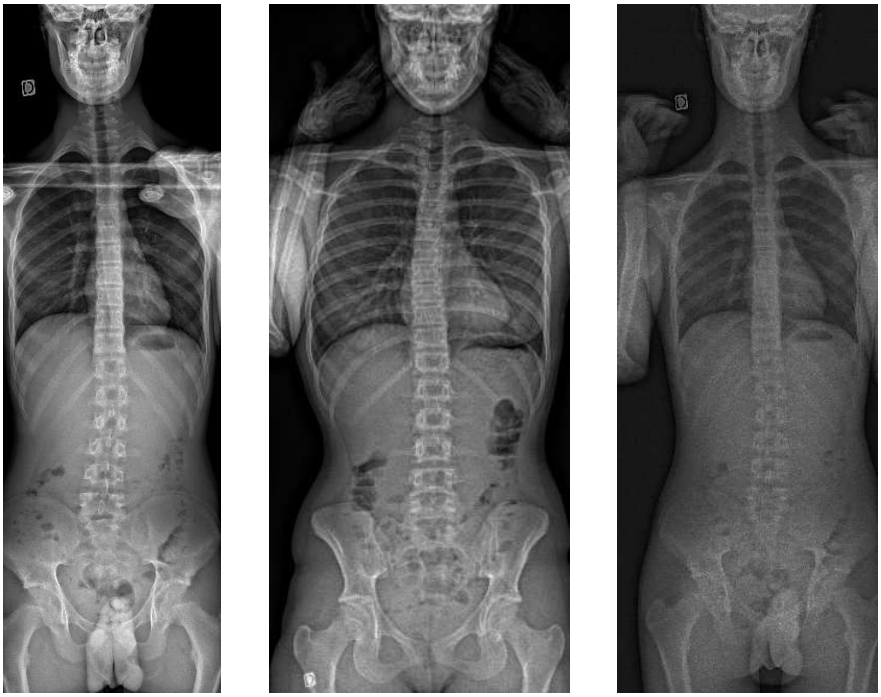
The ALARA (As Low As Reasonably Achievable) principle (Goske et al., 2008) has gained more and more importance in the medical community given the aforementioned dangerous increasing trend of exposure to artificial radiations. The main idea of this guideline is that the radiation exposure should be ALARA, i.e. an image should be acquired with a parameter setting that limits as much as possible the amount of dose absorbed by the patient, while achieving a given clinical need. In fact, a high quality image is not necessarily a good choice if it comes at the price of high radiation exposure. Professionals have been particularly encouraged to follow ALARA guidelines in clinical routine as shown by the *Image Gently* campaign[2].

In this thesis, we apply image processing techniques to reduce the dose by optimizing the post-processing of acquired images (Section 1.3.1). Moreover, in order to help the users to work according to the ALARA guidelines, we study how to automatically provide a feedback on the image quality (Section 1.3.2). In the next section we clarify why EOS is an ALARA oriented device.

1.2.3 EOS and dose

EOS is ALARA because, for certain clinical needs, it allows replacing traditional X-ray devices that are more demanding in terms of dose. Deschênes et al. (2010) have presented a study that compares EOS with a CR system on 50 patients with spinal deformations. According to the results, the average skin dose was reduced from 6 to 9 times when using EOS instead of CR. Besides, the visibility of all the considered structures in both frontal and lateral view images was increased, except

[2] http://www.imagegently.org.



(a) (b) (c)

Figure 1.2: *Types of EOS acquisitions: (a) Diagnostic (*115.62 *µGy); (b) Follow-up (*14.64 *µGy); (c) Preview (*1.44 *µGy).*

the lumbar spinous process. Other examples of the advantages of EOS in terms of dose can be found among the studies presented in Section 1.1.2.

Figures 1.2a, 1.2b and 1.2c show three types of images that can be acquired with EOS: diagnostic, follow-up and preview images, respectively. The term diagnostic is used to denote images that can be used for any application within the scope of EOS device, such as evaluation of structures and lesions. Diagnostic EOS images are acquired with the functionality called *EOS Low Dose*. This name highlights the substantial reduction of dose in standard EOS examinations with respect to other modalities.

Follow-up images are used only in pediatric radiology for balance control assessment of static conditions for the spine and the lower limbs. The radiation exposure is reduced by approximately a factor 7 when using follow-up parameter settings rather than diagnostic ones, which is extremely relevant considering the risks associated with repeated examinations on young patients (Ronckers et al., 2008). Follow-up images are acquired with the functionality called *EOS Micro Dose* that is not intended to be used for the aforementioned diagnostic applications. Nevertheless, the degree of visibility of the anatomical structures is sufficient for the follow-up, which is a good example of enactment of the ALARA principle.

Preview images are optionally acquired to check the positions of the patients and eventually refine the horizontal collimation limits. The air kerma for a preview image is of the order of 1.5 μGy, i.e. approximately 80 times inferior to the air kerma associated with the subsequent diagnostic acquisition. The effective dose is

comparable to half a day of natural radiation exposure.

1.3 Objectives of the thesis and related challenges

According to the presented context, we present the challenges and motivations of this thesis by using the ALARA principle as the basic concept. This analysis justifies the structure chosen for the manuscript which is presented at the end of this chapter.

1.3.1 Optimization of the image quality

The reduction of radiation exposure entails an increase of the noise level that reduces the visibility of anatomical structures. We address this issue by optimizing the image processing chain that provides the display image by using the data acquired by the detector as inputs. Our objective is to allow for a reduction of dose by counterbalancing the drop in signal to noise ratio, due to the inferior radiation exposure, with an optimal algorithm design.

The challenges related to this task are however multiple. First, low dose images are not only affected by noise, but fine anatomical details are not very easy to distinguish either, which is why the contrast has to be enhanced. Secondly, the high dynamic of the detector allows covering the full body of the patient and, thus, the images are rich in different types of features, e.g. bone texture in low absorption regions and smooth areas corresponding to soft tissues. Therefore, the method should be flexible enough to well process all the anatomical regions in the field of view. Furthermore, the noise is signal-dependent, which implies that the noise level changes according to tissue densities. It is worth noting that the concepts of low contrast and high dynamic are not incompatible. Indeed, in this thesis, the contrast denotes the visibility of image features at medium and high frequencies, whereas the dynamic indicates the range of density tissues that can be covered in the field of view of the image. For example, the fact that an EOS lower limb image can properly represent both the pelvis and the ankle does not imply that the texture of bone tissues in the knees is visible enough for the diagnosis. Finally, the image quality depends on the acquisition parameter setting, the patient morphotype and the imaged area. The proposed approach should then adapt to these heterogeneous conditions.

These challenges are addressed by proposing a method that jointly denoises and enhances the contrast of X-ray images. We extend the Non Local-means filter to restore EOS images that are affected by non-homoscedastic noise. We then combine it with the contrast enhancement in different ways that are evaluated according to potential clinical needs. Moreover, particular attention is dedicated to limit the number of parameters involved in the method to make it robust to the aforementioned heterogeneous conditions.

1.3.2 Automatic quantification of the image quality

The ALARA principle is conceptually simple, but it is not trivial for a user to certify if he/she is following it. A classical solution in digital radiography consists in

providing users a feedback that is correlated to the image quality just after the acquisition. This feature would significantly help radiographers and clinicians in their daily practice. Moreover, we could imagine to extract the same kind of information from preview images and use it as input to an automatic exposure control system in charge of tuning the acquisition parameters according to ALARA guidelines. In this thesis we aim at defining a method for the automatic quantification of the image quality regardless the amount of radiation exposure and the patient morphotype.

The main challenge is in this case presented by the richness of information in the images. In practice, the measures have to be associated with anatomical structures of interest because just one measure over the full body is not very meaningful due to the heterogeneity of tissues. On the other hand, the detection of structures of interest and the extraction of robust descriptors are difficult tasks for two reasons. First, on images that have not yet been post-processed, the contrast is low. Secondly, different tissues are superposed because of their projection on the 2D plane of the image. Furthermore, the process has to be efficient despite the wide search area. Finally, the provided measure should not be influenced by the presence of metallic objects, the variations in the positions of anatomical structures due to a pathology, and the changes in patient size and posture.

We propose a landmark-based method that provides consistent intra-exam image quality measures, takes into account the presence of multiple structures in the field of view, and exploits the redundancy of local estimates computed at the landmarks to minimize the impact of eventual outliers. This approach is then combined with a method of detection and recognition of anatomical landmarks on full-body EOS images, which completely automatizes the process of image quality estimation.

1.3.3 Organization of the manuscript

Chapter 2 reviews the state of the art methods that address the problems of noise and low contrast by contextualizing the discussion to medical applications and properties of EOS images. Chapter 3 presents our new method for joint denoising and contrast enhancement of X-ray images. The second aspect studied in this thesis is introduced in Chapter 4 where we both clarify the concept of exposure index and propose a new landmark-based approach that allows quantifying the image quality from local estimates. Chapter 5 presents our approach for the automatic detection of anatomical landmarks on frontal and lateral view EOS acquisitions. Conclusion and perspectives are discussed in Chapter 6.

Chapter 2

Low dose X-ray images: a matter of high noise and low contrast

Low dose X-ray images are affected by a significant amount of noise that reduces the visibility of anatomical structures and, hence, makes the diagnosis more difficult. Therefore, the reduction of noise in these images is a relevant subject. The literature on denoising is particularly rich and the differences among the methods depend on the underlying assumptions about the nature of the signal (e.g. smooth, piece-wise constant and so on) or the noise properties. In general the research on denoising aims at finding flexible methods that perform well in multiple conditions and can adapt to different signal nature and noise properties. In this chapter, we review the state of the art on denoising. This allows identifying the most significant elements to take into account for processing EOS images according to their properties.

In addition to strong noise, the low contrast makes the interpretation of low dose X-ray images difficult, and, therefore, they need to be enhanced. This means that the contrast in low activity regions has to be increased (e.g. fine texture), while keeping the signal stable in strong activity ones (e.g. strong edges at the border of the patient envelop). Moreover, the presence of noise can significantly affect the outputs of contrast enhancement methods because it deteriorates the image representation and makes the quantification of the activity levels ambiguous. Finally, in this chapter we formulate the problem of contrast enhancement of noisy images and give an overview of methods from the literature.

Organization of the chapter. Section 2.1 presents the literature on denoising techniques by over-viewing some classical algorithms and, in particular, patch-based methods that represent the state of the art. Moreover, the extension of the filters to deal with non-homoscedastic noise is discussed and the application of the patch-based filters to medical images is over-viewed. Section 2.2 introduces the problem of contrast enhancement by focusing on multiscale representation and by studying the approaches that consider the presence of noise. Section 2.3 presents the proposed validation protocols that allow comparing different post-processing algorithms for both phantom and clinical X-ray images. Finally, Section 2.4 concludes the chapter by summarizing the most relevant considerations for processing EOS images.

2.1 Image denoising

2.1.1 Introduction to the denoising problem

An image is mathematically represented by N observations placed on a regular grid $\Omega \in \mathbb{Z}^d$ of dimension d (equal to 2 because, in this thesis, 2D images are processed). These observations correspond to the pixels of the image and are described by their positions x_k on the grid Ω and associated vectors of values $s_k \in \mathbb{R}^m$. Since gray level images are considered, the dimension m is fixed to 1 and, so, the vector of values is just an intensity value. An image $\mathbf{u}$ can then be described by the following function:

$$u_k = \mathbf{u}(x_k), \quad k \in [1, \ldots, N]. \tag{2.1}$$

The values observed in an image could be deteriorated by various sources of disturbance, such as blurring, missing pixels, noise and so on. In practice, the image $\mathbf{u}$ hides another one $\mathbf{y}$ that is described by M observations and contains the actual information of interest. The relation between the images $\mathbf{u}$ and $\mathbf{y}$ is represented by the following model:

$$\mathbf{u} = \Phi\mathbf{y} + \sigma\varepsilon \tag{2.2}$$

where $\varepsilon \in \mathbb{R}^N$ models the errors due to the noise, $\sigma > 0$ is the noise level and Φ is a linear matrix $N \times M$ that describes linear operations such as, for example, blurring. The estimation of $\mathbf{y}$ from $\mathbf{u}$ is a restoration problem and aims at recovering a signal corrupted by known disturbances described by a given model.

In this thesis, no linear operations have to be taken into account and, so, $M = N$ and Φ is the identity matrix. In other words, restoration is, in this case, synonym of denoising. In order to solve this problem it is then necessary to properly characterize the noise properties and, for EOS images, this aspect is addressed in Section 3.1.1. In many works the noise level is assumed to be known, that is $\mathbf{s}$ is simulated from $\mathbf{y}$ with a given noise model and level. The objective of the research in such conditions is, for example, to evaluate the adaptability of different denoising techniques. This is not the case for the works conducted in this thesis as the images to process are not simulated, but obtained from real acquisitions. Therefore, σ needs to be estimated from a single image and Section 3.1.3 addresses this aspect. It could be discussed that the noise level can be easily predicted from the acquisition parameters. However, in digital radiology, the noise level also depends on the radiological thickness that cannot be properly estimated before the acquisition.

The additive white Gaussian noise (AWGN) is the most popular model. In this case, Equation 2.2 can be simply rewritten as follows:

$$\mathbf{u} = \mathbf{y} + \sigma\varepsilon \tag{2.3}$$

where ε is a realization of a stochastic process that follows a normal distribution.

The relation between $\mathbf{u}$ and $\mathbf{y}$ can be statistically represented by describing $\mathbf{u}$ as the realization of a probabilistic process U conditioned by $\mathbf{y}$ and, hence, using the probability density function $p(\mathbf{u} \mid \mathbf{y})$. Under the AWGN assumption, the probability density function $p(\mathbf{u} \mid \mathbf{y})$ follows a Gaussian distribution with standard deviation σ and mean $\mathbf{y}$. The statistical description is particularly useful for other types of noise for which it is not possible to decompose $\mathbf{u}$ as function of $\mathbf{y}$, e.g. the Poisson noise (Deledalle, 2011). As for EOS images, this aspect will be discussed in Section 3.1.1.

A filter h estimates a solution $\hat{\mathbf{y}} = h(\mathbf{u})$ from the noisy image $\mathbf{u}$. The optimal solution minimizes the square error with respect to the underlying ground truth signal. However, it is not possible to simply estimate the solution according to this principle because the problem is not deterministic. Indeed, two identical images could be produced from two different noise free ones or, similarly, from the same noise free image two different ones could be generated because the noise is stochastic. The solution can then be found only by minimizing a criterion such as the mean square error (MSE), i.e. by finding a solution $\hat{\mathbf{y}}$ that lies in the neighborhood of $\mathbf{y}$. Consequently the consistency of an estimation is described according to the bias-variance trade-off. The bias is the term that penalizes the loss of structures of interest and appearance of artifacts, while the variance penalizes the amount of residual noise. A filter aims at optimizing this trade-off, and good assumptions about the noise and the features of the noise-free image are necessary to achieve this goal. For instance, if the model is too limited, i.e. approximate, then the variance will be significantly reduced at the price of the introduction of a strong bias. On the contrary, if the model is too flexible, i.e. it tries to adapt to every possible situation, the bias will be low at the price of a high variance. Finally, the filter to be used should be chosen according to expected properties of the image $\mathbf{y}$ and type of noise. Therefore, an overview of different denoising techniques is provided in the following sections in order to identify which one could perform better on EOS images.

2.1.2 Filtering in image space

A filter that estimates $\hat{\mathbf{y}}$ directly in the domain of the image can operate according to linear or non linear operations. The following sections review linear and non-linear filters.

2.1.2.1 Linear filters

Blurring filter. The principle of spatial coherence can be applied to address the denoising problem. This relies on the simple assumption that neighbor pixels share the same information and eventual variations are only due to noise. A sliding window W is then moved along the image space and each intensity value is replaced by the average value of the gray level distribution at the window. These filters are also called linear filters as the estimated image $\mathbf{y}$ is the result of linear operations and, in particular, of a convolution of $\mathbf{u}$ with a kernel function of given shape and size. The advantages of these filters lie in their conceptual simplicity and strong noise reduction because by averaging K values the noise level is reduced $\sqrt{K}$ times.

Pixels contain less and less likely the same information as the distance to the window center increases and, hence, the contributions to the estimate from neighbors can be weighted depending on spatial distances. This idea is applied by replacing the square window with, for example, a Gaussian kernel. Formally, the value of $\hat{\mathbf{y}}$ at pixel x_i is defined as follows:

$$\hat{\mathbf{y}}(x_i) = \frac{\sum_j \omega(x_i, x_j)\mathbf{u}(x_j)}{\sum_j \omega(x_i, x_j)} \tag{2.4}$$

where $x_j \in W$ and $\omega_{ij} = \omega(x_i, x_j)$ is a term that weights the contribution of x_j for the estimation of x_i. When a Gaussian kernel of bandwidth equal to h is used, the

weights are defined as follows:

$$\omega_{ij} = \exp\left(\frac{-\|x_j - x_i\|^2}{2h^2}\right). \tag{2.5}$$

Then, the greater the spatial distance between the pixels x_i and x_j the lower the contribution of x_j for estimating $\hat{\mathbf{y}}(x_i)$ is.

Wiener filter. The image can be denoised by maximizing the probability $p(\mathbf{y} \mid \mathbf{u})$, i.e. the solution is the maximum a posteriori (MAP) estimation. The Wiener (or Bayes) filter is based on this principle. In practice, some assumptions about the noise and image models need to be made to get a solution. The noise and the image are both described by a Gaussian normal distribution of standard deviations σ_η and σ_y, respectively. Moreover, the space of possible solutions is constrained to linear combinations of the observed data. Under these conditions, the solution that maximizes the MSE is obtained as follows:

$$\hat{\mathbf{y}} = \left(\frac{{\sigma_y}^2}{{\sigma_y}^2 + {\sigma_\eta}^2}\right)\mathbf{u} \tag{2.6}$$

The hypotheses made to obtain this solution are, however, quite restrictive and rarely observed as natural images do not typically have constant mean and variance. Some improvements can be obtained by applying the Wiener filter in local windows where the distribution of gray levels is more stable. Nevertheless, the assumptions are still not valid in areas that present strong discontinuities (e.g. edges).

2.1.2.2 Edge preserving filters

Linear filters provide good results on smooth images, i.e. without strong edges. However, this is rarely the case when dealing with natural images and, since linear filters are basically low pass filters, the estimated $\hat{\mathbf{y}}$ will suffer from significant resolution loss. Edge preserving filters address this issue by imposing that the images are piece-wise constant, i.e. regular but with some strong variations. In this case, the image is processed according to its content and, so, these filters are adaptive. They are also addressed with the name of non-linear filters to highlight that the pixels are not all processed in the same way. This principle has been introduced by Perona and Malik (1990) and is known as anisotropic diffusion. The information located on opposite sides of an edge should not be merged or, in other words, the diffusion should not be homogeneous. Formally, Perona and Malik (1990) have used the heat equation with spatially varying coefficients depending on the gradient. Therefore, the diffusion relies on a parameter that quantifies the parallelism between strong diffusion and low gradient magnitude.

An alternative way to adapt to image content consists in extending linear filters by adapting shapes, sizes and/or bandwidths of the used kernels (Katkovnik et al., 2002; Takeda et al., 2007) according to local statistical analysis or gradient magnitude and orientation.

Total variation. Rudin et al. (1992) have introduced the total variation (TV) that formulates the denoising as an inverse problem where the solution is the result of a compromise between two terms:

$$\hat{\mathbf{y}} = \arg\min\left(\|\mathbf{u} - \mathbf{y}\|_2 + \lambda \sum_i |\nabla(y(x_i))|\right) \tag{2.7}$$

where the l_2 norm controls the consistency with the data and the other term is the TV that penalizes variations and transitions. The higher the parameter λ the more regular the solutions is, i.e. for $\lambda \to +\infty$, $\hat{\mathbf{y}}$ is constant. If the l_1 norm is used in place of the l_2 norm, then, the algorithm is called TV-l_1 (Alliney, 1992). Nikolova (2002) has shown that by using the l_1 norm the method is more robust to the presence of outliers. However, a drawback of this solution is that structured artifacts (*staircase-effects*) at the side of strong edges may appear.

Bilateral filter. Tomasi and Manduchi (1998) have introduced the bilateral filter that is a non-iterative edge preserving filter. The image is blurred according to spatial distances, as for linear filters, but only similar intensity values are taken into account. In practice, within a local window, the weights are defined according to both closeness and similarity by using two Gaussian functions (see Equation 2.5). Therefore, the bilateral filter depends on two parameters σ_s and σ_r that are, respectively, the spatial and range kernel bandwidths. Intuitively, for high σ_r values the filter behaves like a Gaussian filter, but for lower values the edges are better preserved. Nevertheless, how to define these two parameters is not trivial especially for noisy images where the intensity level is not a consistent descriptor.

In conclusion, adaptive filters in the image space improve the results compared to linear ones, but they fail in restoring high frequency structures (e.g. textures) as the amount of discontinuity is defined as a function of parameters that are difficult to set.

For EOS images, these methods do not seem to be well adapted because the images are rich in high frequency patterns such as bone texture. Moreover, because of the low contrast of the images to process, the gradient based information is not sufficient to capture the discontinuities.

2.1.3 Filtering in transform domain

The features of an image can be represented by means of a dictionary that is composed by structuring elements called atoms. A dictionary is a set of bases and the image is described as their linear combinations. A denoised image is then defined by finding an optimal combination of coefficients by exploiting that, in a proper base, natural signals are sparse, i.e. they can be represented by a limited number of bases (Candes and Wakin, 2008). Since the noise does not share this property, as shown by Donoho and Johnstone (1994), the denoised image can be obtained by analyzing the representation of the noisy one provided by an appropriate dictionary. Formally, the dictionary based filters look for a sparse decomposition of $\mathbf{y}$ by fitting the observations $\mathbf{u}$, which requires to solve the following constrained minimization problem:

$$\min_{D,\alpha} \|D\alpha - \mathbf{u}\|, \quad \|\alpha\|_0 < \epsilon \tag{2.8}$$

where D is the dictionary represented as a matrix of size $P \times Q$ and α is the vector of Q coefficients used to represent the denoised image. The variable ϵ controls the amount of sparsity of the solution since the pseudo-norm l_0 is the number of non-zero coefficients. This problem can be minimized by using Orthogonal Residual Matching Pursuit (ORMP) (Cotter et al., 1999). The pseudo-norm l_0 is often replaced by the l_1 norm leading to the following formulation:

$$\min_{D,\alpha} \|D\alpha - \mathbf{u}\|, \quad \|\alpha\|_1 < \epsilon \tag{2.9}$$

This choice still guarantees a sparse representation of a signal (Candes and Wakin, 2008), whereas the l_2 norm does not preserve sparsity.

Orthogonal bases. These problems can be solved in practice by imposing some restrictions on the properties of the dictionary. Historically, orthogonal bases have been used first. In this case, $P = Q$ and the matrix D is orthogonal.

For example, the discrete cosine transform (DCT) is obtained by using orthogonal cosine functions as bases. This dictionary is used in the famous compression standard JPEG. Since DCT captures only frequencies of the signal, it is difficult to separate it from noise. Therefore, Mallat (1989) has introduced the discrete wavelet transform (DWT) that captures scale and space. The dictionary is defined according to dyadic translations and rescaling of a mother wavelet $\Psi(t)$ as follows:

$$b_{j,k}(t) = \sqrt{2^j}\Psi(2^j t - k) \tag{2.10}$$

where j and k are two integers. The mother wavelet $\Psi(t)$ is chosen in order to well encode the features of natural images, e.g. discontinuities. Besides, it has to respect some conditions in order to guarantee the inversion of the transform (Daubechies, 1992). For example, Haar, Daubechies and Gabor wavelets follow these conditions. Some researchers have adapted the wavelets in order to better represent some specific shapes. For example, Candes (1998) has introduced ridgelets and then curvelets (Candes and Donoho, 2001) to represent ridges and curves, respectively. Please refer to Mallat (2008) for a complete review on these and other representations.

In denoising applications, the solution to the problem in Equation 2.8 is obtained by applying hard thresholding to the noisy coefficients, that is:

$$\hat{c} = \begin{cases} 0 & \text{if } |c| < \lambda \\ c & \text{otherwise} \end{cases} \tag{2.11}$$

where λ is the value of the threshold, c are the input wavelet coefficients and $\hat{c}$ are the corresponding output coefficients. Similarly, the solution to the problem in Equation 2.9 is obtained by applying soft thresholding to the noisy coefficients, that is:

$$\hat{c} = \begin{cases} 0 & \text{if } |c| < \lambda \\ sign(c)(|c| - \lambda) & \text{otherwise} \end{cases}. \tag{2.12}$$

The choice of the value of λ significantly influences the quality of the result. Rather than fixing a global value, Donoho and Johnstone (1995) have shown that it is better to spatially adapt the threshold. This procedure is driven by the Stein unbiased risk estimation (SURE) that, under AWGN conditions, estimates the MSE while not requiring to know $\mathbf{y}$ (Stein, 1981).

Over-complete dictionaries. Some oscillations around edges, known as Gibbs effect, can be produced by shrinking the coefficients as the orthogonal bases are not shift-invariant, that is, coefficients obtained by decomposing a shifted image do not correspond to shifted original ones. Over-complete dictionaries address this issue by making the image representation nearly shift-invariant. For example, Selesnick et al. (2005) have proposed the dual tree complex wavelet transform that mirrors the properties of the Fourier discrete transform to obtain smooth and non-oscillating coefficients magnitudes. This method still relies on prior shapes to build the dictionary, while other researchers have proposed to learn the dictionary from a set of noisy

data as it allows to better encode image features. The task is practically impossible when the whole image space is considered because of the high dimension. On the other hand, this becomes feasible if small sub-images, i.e. the patches, are used. Aharon et al. (2006) have exploited this idea in the formulation of the K-single values decomposition (K-SVD) algorithm. This is just an example of how classical filters have been in the last recent years adapted to the representation in the patch space that is used in the formulations of the current state of the art denoising algorithms. We provide an overview of this type of representation and related filters in the following sections.

2.1.4 Filtering in patch space

An image can be represented as a collection of overlapping patches P_x. The interest for such a description of the images is motivated by the redundancy that is often present in natural images. For example, homogeneous and textured regions repeat themselves in different areas of the image. By exploiting this regularity, information of interest can be detected even if the observed data are deteriorated by noise. The patches consider the spatial context of each pixel and, then, allow for a consistent description of similar image features. Formally, an observed image $\mathbf{u}$ is represented by means of the following relations

$$\{(x, \mathbf{u}(P_x)) \mid x \in \Omega\} \quad \text{where} \quad P_x \in \Omega \quad \text{and} \quad \mathbf{u}(P_x) = \{u_k \mid x_k \in P_x\} \tag{2.13}$$

then, each pixel $x \in \Omega$ is associated with a set of $|P|$ values $\mathbf{u}(P_x)$ where $|P|$ is the size of the patch $P_x \in \Omega$. A patch can have different shapes, but the most used is the square one. In this case, a patch P_x of half-size p is a sub-image of size $|P| = (2p+1) \times (2p+1)$ centered at the pixel x.

This representation is redundant because the patches are overlapped and projects the image in another space, that is the *patch space*. According to how a filter quantifies the similarities among the patches, different denoising techniques are defined and the main ones are described in the following sections. From the discussion, it will become clear that all these methods are extensions of conventional filters to the patch space.

2.1.4.1 Non Local-means filter

The Non Local-means (NL-means) filter is conceptually the easiest patch-based filter as it could be seen as the extension of the Gaussian filter in the patch space. Indeed, the estimation is obtained as a weighted mean, however the weights are not assigned according to distances in the spatial space but rather in the patch one. As long as it is possible to find several similar patches in the same image, the self similarity boils down to a local image regularity assumption (Lebrun et al., 2012). This concept has been first applied to the synthesis of textures (Efros and Leung, 1999). Then, Buades et al. (2005a) have adapted this idea to denoising. The filtered image $\hat{\mathbf{y}}$ is obtained as a weighted sum (see Equation 2.4) where the weights are computed as follows:

$$\omega(i,j) = \varphi\left(\frac{d(P_{x_i}, P_{x_j})}{2h^2|P|}\right) \tag{2.14}$$

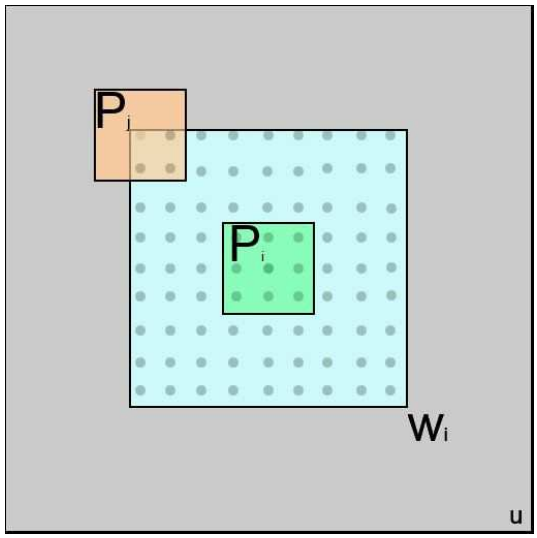


Figure 2.1: *Search for similar patches. The green patch P_{x_i} centered at pixel x_i is compared to all the other orange patches P_{x_j} centered at pixels $x_j \in W_i$. In this example, $|P| = 3 \times 3$ and $|W| = 9 \times 9$.*

where φ is a kernel decay function $\mathbb{R}^+ \to [0, 1]$ and in this thesis corresponds to $\exp(-t)$. Then, d is the distance function that quantifies the affinity between patches and h is a smoothing parameter that drives the decay of the function φ. The search domain for similar patches could be the whole image, but for both efficiency and better quality of the results it is preferable to limit the search to a small window W of size $|W|$ (Buades et al., 2005a). Figure 2.1 explains how the research is conducted. All these parameters influence the quality of the estimation of $\mathbf{y}$ and need then to be reviewed:

- **Window size**. A too small window decreases the probability of finding similar patches and, so, the filtered image could still present some residual noise because the hypothesis of redundancy, on which the NL-means is based, is not respected anymore. On the contrary, a too wide window increases the number of elements to average, which can cause over-smoothing. Duval et al. (2011) have suggested to locally adapt the window size according to SURE measures by increasing in this way the result quality but the computational load as well.

- **Patch size**. A too small patch size makes unstable the comparison between patches. Note that the degenerate case $|P| = 1$ returns a solution that boils down the NL-means filter to the bilater filter. Patches that are too large may not find a sufficient amount of replicas as the redundancy principle is locally inconsistent. Visually, this causes the *rare patch effect* (Deledalle, 2011), i.e. residual noise artifacts in regions with a low amount of redundancy. Some improvements are achievable by locally adapting patch size and shapes. For example, Deledalle et al. (2011) refine patch parameters according to local risk estimation, which however increases the computational load.

- **Kernel bandwidth**. The smoothing parameter h is often defined as a function of the noise strength σ (Buades et al., 2005b). This demands to estimate

σ, that in real cases is not known in advance, to avoid manual setting of this parameter. Moreover, h has to be locally adapted when the noise is signal dependent as discussed in Section 3.2.1.

- **Distance function**. Buades et al. (2005a) have defined d as a weighted Euclidean distance, i.e. an Euclidean norm convolved with a Gaussian kernel. However, in more recent works (Darbon et al., 2008; Coupé et al., 2008), the NL-means filter uses a standard Euclidean norm with no noticeable qualitative difference, while avoiding setting the smoothing parameter of the Gaussian kernel. Buades et al. (2005b) have demonstrated that this distance provides an optimal estimation under AWGN model. Deledalle et al. (2012) have extended the patch comparison principle to other noise models by proposing other distance functions.

In this thesis, these parameters have been set according to the properties of digital X-ray images and the associated noise model. As these aspects are strictly related to the NL-means extension to radiography images that we propose, they are detailed in Section 3.2.1.

Compared to classicals filter Buades et al. (2005b) have shown that the NL-means filter allows capturing different image features such as straight edges, flat and textured regions. The filter is not specifically designed for these cases, but it well represents these features by exploiting the spatial context encoded by the patches. Figure 2.2 shows the distribution of weights associated with the estimation of the central pixel of the window in various cases encountered on natural images. Moreover, under stationary conditions, i.e. with a sufficient amount of redundancy, the NL-means filter converges to the Bayesian optimal solution (Buades et al., 2005b).

The order of computational complexity of the NLM filter in its naive formulation is equal to $\mathcal{O}(N|W||P_i|)$, which is quite high. Darbon et al. (2008) have addressed this issue by making the estimation independent on the patch size: the weights are computed from the discrete integration between the difference of the observed image $\mathbf{u}$ and its shifted versions by using integral images; the computational load is then reduced to $\mathcal{O}(N|W|4)$. Moreover, considering that the NL-means filter works separately window by window, the process can be decomposed in multiple parts and, then, easily parallelized. Note that non-naive implementations of the NL-means filter are consistent only when the Euclidean distance is used. Then, the extension of efficient NL-means filter to non-homoscedastic noise cases is not trivial. In this thesis, the NL-means is efficiently implemented in spite of signal-dependent noise as explained in Section 3.2.1.

Since the patches are overlapped, $|P|$ estimations are associated with each pixel and, so, they need to be aggregated to retro-project the result from the patch to the image space. In detail, the $|P|$ estimations may simply be averaged. However, strong edges affect the weights associated with surrounding pixels, entailing an effect called *noise halo*, i.e. residual noise in flat regions near strong edges. This effect can be significantly reduced by weighting the contribution of each vote by preferring low variance ones, i.e. by giving more importance to estimations from pixels that are near the edges but do not meet them. Various works have addressed this specific issue. For example, Kervrann and Boulanger (2008) have proposed to decompose the image according to the bias-variance principle and then select the estimations at

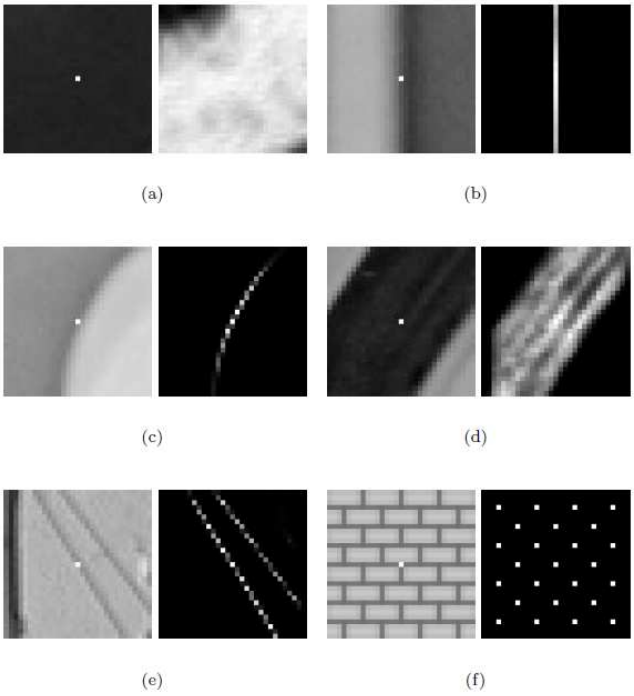


Figure 2.2: *Distribution of the NL-means weights (right-hand side of each sub-image) associated with the estimation of the central pixel in a window (left-hand side of each sub-image) in case of: (a) Homogeneous region: all the patches look alike and so the NL-means acts as a mean filter; (b) Straight edge: only patches centered at the discontinuity are averaged; (c) Curved edge: even if the discontinuity is not straight, the NL-means filter catches the similarity; (d) Flat neighborhood: case similar to (a) but only pixels with similar intensity values are considered; (e) and (f) similar patterns: the spatial distance does not constraint the NL-means in finding similar patches. These images are extracted from the paper by Buades et al. (2005b)).*

low variance. However, this approach is not exploitable in practice as it requires to know the ground truth image $\mathbf{y}$. The approach proposed by Deledalle et al. (2011) overcomes this drawback by adapting patch size and shape according to SURE estimation. In general, all these approaches cannot be efficiently implemented as they demand expensive local analysis. Moreover, this problem is actually secondary in the context of this thesis because the images to be processed have low contrast.

A commonly used pre-step that increases the effectiveness of patch-based filter consists in estimating the weights from regular versions of the observed image $\mathbf{u}$ (Lebrun et al., 2012). In detail, a second term that represents the estimation from the smoothed input image (e.g. linear blurring filter) could be added to the argument of the function φ in Equation 2.14 to balance the contributions from corrupted observed data and regular ones. However, this implies the introduction of another variable that further increases the already not trivial task of defining a good parameter setting. Alternatively, recent approaches propose to estimate the weights in an iterative way (Lebrun et al., 2012): after estimating $\hat{\mathbf{y}}_\mathbf{1}$, the image is filtered a second time and the weights $\omega(x_i, x_j)$ that quantify the distance between patches in the regular image $\hat{\mathbf{y}}_\mathbf{1}$ are used to restore the observed data $\mathbf{u}$. In practice, the

image $\hat{\mathbf{y}}_1$ plays the role of an *oracle* that is closer to the ground truth solution $\mathbf{y}$ than $\mathbf{u}$ and from which it is easier to capture image features. Note that, as shown in some works (Chatterjee and Milanfar, 2012; Lebrun et al., 2013), more than one iteration does not improve much the results. In this thesis, this idea is not directly applied to denoise the image, but rather to estimate noise containment maps that are then used in the proposed contrast enhancement algorithm (see Section 3.3).

2.1.4.2 Other patch-based filters

Given the good performances of the NL-means filter, the representation of the image by means of patches has raised increasing interest in the research community. Therefore, more elaborated concepts than simple weighted mean according to Euclidean distances have been elaborated to compare patches. Here follows a list of some of these methods and related main ideas. Please refer to the provided references for the mathematical formulations and details.

BLS-GSM. Classical denoising methods in transform domain separately process coefficients, but neighbor coefficients are dependent. Therefore, Portilla et al. (2003) have modeled these relations by using Gaussian scale mixture (GSM), i.e. the product of a Gaussian random vector and an independent hidden random multiplier. From this model, a Bayes least-square (BLS) estimator is used to locally denoise the coefficients. This algorithm is patch-based as it considers a patch composed by nearby coefficients plus another one at the next equally oriented scale. The solution is obtained by retro-projecting the denoised scales into the image space with the inverse wavelet transform.

K-SVD. An algorithm introduced by Aharon et al. (2006) and then extended by Mairal et al. (2009) combines the sparse coding principle to patch representation. It estimates the sparse representation of noisy patches from an initial dictionary. Then, the dictionary is updated and optimized by performing a K single values decomposition (K-SVD). Note that the K-SVD algorithm does not execute any research in local windows but it rather computes a sparse representation of each patch in the image by using ORMP to solve the non-polynomial hard problem in Equation 2.8.

BM3D. Dabov et al. (2007) have proposed the block matching 3D (BM3D) algorithm that is also known under the name of collaborative filter as it combines sliding window DCT coefficient shrinkage (Yaroslavsky et al., 2001) and NL-means. In detail, similar patches are first grouped in 3D piles, then each of them is encoded with a 3D wavelet transform and the resulting coefficients are shrinked. After applying the inverse 3D transform, the filtered patches are replaced in their original positions and aggregated according to the number of preserved coefficients per patch. Moreover, BM3D uses a first denoising step as oracle in order to make the clustering of patches more robust. This method has also been extended to the use of patches of other shapes than the square one (Dabov et al., 2009).

PLOW. The patch-based locally optimal Wiener filter (PLOW) proposed by Chatterjee and Milanfar (2012) is an example of empirical adaptation of Bayes principle by relying on patch representation. The K-means algorithm is used to group the patches into 15 different clusters according to geometrical similarity. Then, for each cluster, the denoised patches are obtained by optimizing the linear minimum mean-squared-error (LMMSE). As for the BM3D, in order to improve the quality of clustering in the first step, the algorithm is run in two iterations.

NL-Bayes. The non local Bayes (NL-Bayes) proposed by Lebrun et al. (2013) is another Bayesian based filter that differs from PLOW in the way similar patches are defined. In detail, NL-Bayes starts by computing the Euclidean distance between a target patch P_{x_i} and other ones in a search window of size $|W|$. Then, the patches are ordered according to the computed distances and the smallest E ones are retained, where E is fixed at a given value. Then, the mean patch $\bar{P}$ and co-variance matrix $C_{\bar{P}}$ are computed in order to approximate the respective ground truth ones. Also in this case, the procedure is executed twice to obtain better results. This model is then used to estimate denoised patches from observed data by applying the following formula which resembles a local Wiener filter (Lebrun et al., 2013):

$$\hat{P} = \bar{P} + C_{\bar{P}}[C_{\bar{P}} + \sigma^2 I]^{-1}(P - \bar{P}) \tag{2.15}$$

Compared to other methods, NL-Bayes creates piles of patches as the PLOW and BM3D algorithms, but it uses a more robust similarity principle than the global geometrical clustering used in PLOW. Besides, it is more flexible than BM3D as the fixed 3D orthogonal transform is replaced by adaptive bases composed by eigenvectors of the local Gaussian model (Lebrun et al., 2013)

2.1.4.3 Comparison of patch-based filters

These algorithms are compared among others by Lebrun et al. (2012), showing some interesting properties under the AWGN assumption. Here we summarize some of the most important aspects.

The first tool used in the comparison is the analysis of the residual noise image. By using the patch-based techniques, the residual image is similar to the AWGN, which is not a property shared by classical methods like wavelet shrinkage and total variation (Buades et al., 2005b). The magnitude of the residual obtained from K-SVD and NL-means are stronger than those of the two other methods. Moreover, while these two filters tend to uniformly denoise the image, the NL-bayes, BM3D and the BLS-GSM filters remove little noise around edges and in textured regions, because they are essentially threshold methods and, thus, coefficients larger than the predicted noise level are preserved.

The second aspect studied by Lebrun et al. (2012) evaluates the noise to noise principle, i.e. by filtering an image of AWGN noise the resulting image should still be an image of noise. The results are not satisfying for any of the compared algorithms as they produce low frequency oscillations and eventually artifacts in the case of BM3D. This happens because while the noise equally corrupts all frequencies, patch-based methods remove only the high frequency component of noise. Therefore, the filter should be extended to a multiscale framework to remove the residual low-frequency noise component. A recent approach based on this principle (Delbracio et al., 2014) shows that a multiscale analysis can solve this problem of conventional patch-based filters.

Finally, a simple qualitative analysis is performed. The filter BLS-GSM presents strong Gibbs effect (i.e. ring) around edges, which is due to the orthogonal bases that are used to decompose the image. This effect is present in the result given by the K-SVD filter too, but it is considerably reduced because of the over-complete representation provided by the dictionary. The NL-means filter has no artifacts, but returns an image which is slightly more blurred than the ones estimated with BM3D

and NL-Bayes. While the BM3D filter still shows some ring artifacts for high noise levels ($\sigma > 30$) due to the orthogonal 3D transform, the NL-Bayes does not present this problem and offers a slightly better noise reduction.

In general, the comparison shows that the NL-Bayes is quantitatively and qualitatively the best algorithm. Nevertheless, the computational load is important as it demands to search for similar patches, build the Gaussian model and apply the Bayes principle. Besides, all this has to be multiplied by a factor two since similar patches and Gaussian model steps are repeated a second time because an oracle is used. Considering that the search window size is fixed according to the patch size, the overall complexity to denoise a single patch is equal to:

$$\frac{19}{3}|P|^6 + |P|^4 \left(\frac{13}{4} + 9\log\left(\frac{9}{8}|P|^4\right)\right) \tag{2.16}$$

This is considerably higher than, for example, the computational load of the NL-means filter, which limits the application of this method in the medical context where images have a high number of pixels and short execution time is demanded.

The patch-based representation adapts well to digital X-ray images given the high amount of redundancy and presence of difference image features such as: regular regions corresponding to soft tissues, textured bone regions and straight or curve edges of the bones. Moreover, capturing the local variance in small windows is useful to limit the effect of low contrast in the images. We present a short review on patch-based methods used to restore medical images in Section 2.1.6 after providing some information on how to extend the patch-based method to more realistic cases, i.e. where the noise is not Gaussian.

2.1.5 Beyond the homoscedastic noise model

All the algorithms introduced in the previous paragraph describe the noise with the AWGN model. This assumption is not necessarily true in real applications and to deal with restoration, the noise model has to be well characterized. In particular, since radiography images are the object of this thesis, the Poisson noise and, more generally, the mix Gaussian-Poisson noise is considered.

In many approaches the noise is made homoscedastic by using Anscombe's variance-stabilizing transform (VST), which allows describing the noise model as Gaussian. It has been introduced by Anscombe (1948) and then extended by Murtagh et al. (1995) to adapt it to contexts where Gaussian and Poisson noises are both present. The final goal of this kind of method is to be able to exploit well known Gaussian denoising methods like those described in Section 2.1.4. However, there are multiple arguments against the VST. The main drawback lies in the non-linear distortion of noise-free elements introduced by the transformation. Besides, a more fundamental limit is the non-existence of a perfect variance stabilizing transform for some distributions, like for example the Poisson one (Deledalle et al., 2012). Finally, once the image is denoised, the VST needs to be inverted to obtain the estimation of the signal of interest. The simple algebraic inversion could introduce a bias and, thus, the operation needs to be optimized (Makitalo and Foi, 2011).

The alternative to the Anscombe transform consists in dealing directly with Poisson noise. In this case, the denoising parameters need to take into account the signal-dependent noise. Luisier et al. (2010) proposed a wavelet-based method to

denoise elements affected by Poisson noise, based on a powerful data-driven technique that exploits an unbiased estimate of the MSE, called Poisson Unbiased Risk Estimate (PURE). This work has been later improved (Luisier et al., 2011) by generalizing to the mix Poisson/Gaussian noise. Deledalle et al. (2010) have proposed to replace the Euclidean distance used to compare patches in the NL-means filter with a likelihood estimate that considers Poisson noise model. This solution improves the quality of the result compared to the VST based-approaches when the noise level is high (Deledalle et al., 2012). Salmon et al. (2012) have proposed another approach that uses a sparse learning technique to reconstruct an image from small size dictionaries built from the noisy image. In particular, an adaptation of the principal component analysis (PCA), named Poisson-PCA, is combined with patch analysis.

2.1.6 Reduction of noise in medical applications

Restoration of medical and, in particular, of digital radiography images is a very challenging task as, even by perfectly describing the problem, it is practically impossible to clearly separate noise and anatomical information. As a consequence, even a slight loss of resolution is interpreted as a potential loss of relevant data. For this reason, denoising methods are not very popular in this domain. Nevertheless, some recent works that are here reviewed show that patch-based methods may be useful in this challenging domain too.

(a) (b) (c)

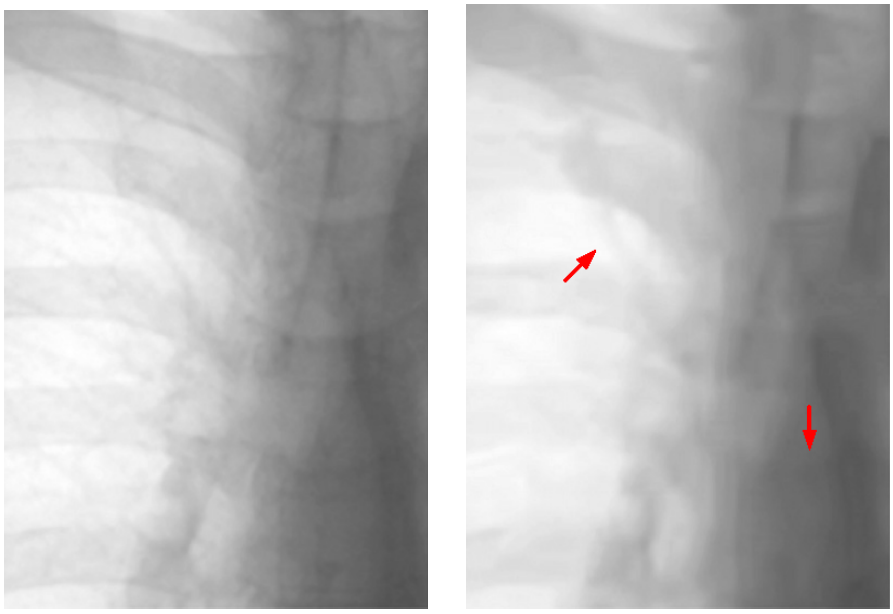
Figure 2.3: *Application of NL-means filter to MRI brain images: (a) Input image; (b) Noisy image adding AWGN noise; (c) Image (b) restored by using the NL-means filter. The noise is sensibly reduced and the structures quite well preserved. However, in some areas indicated by red arrows a significant resolution loss is perceivable. The images are extracted from the paper by Coupé et al. (2008) and the arrows are added to facilitate our discussion.*

Coupé et al. (2008) have applied the NL-means filter to the restoration of 3D MRI images. Their contributions consist in showing that manual tuning of parameters can be avoided as long as σ is correctly estimated, in aggregating patch estimates according to the most significant voxels, and in efficient implementation

by using a block-wise representation and parallelization. The results show a higher capacity in preserving anatomical structures while reducing noise compared to TV, which is coherent with the discussion presented by Buades et al. (2005b). However, Figures 2.3 clearly shows a loss of resolution with respect to the ground truth image, which is reasonable given the high amount of noise but may raise some issues in clinical evaluation.

Patch-dictionary learning have been applied to the restoration of low dose CT scans. Xu et al. (2012) have shown a significant improvement with respect to TV regularization methods that use gradient based measures to drive the reconstruction, which, on images highly corrupted by noise, are not robust. On the other hand, the dictionary learns the shapes of structures of interest in order to both preserve them and reduce the noise. Nevertheless, Xu et al. (2012) give warning about the choice of the dictionary: if a pre-defined dictionary does not describe well the structures of interest in the target image, then some information may be lost and artifacts introduced. Adaptive learning that adds atoms to the global ones from the target image itself, addresses this issue. However, if the sinograms are too noisy, the dictionary learning process fails in capturing high quality information about the structures of interest, causing a quality degradation. Other warnings are given about parameters selection and computational load due to the iterative scheme, but the uncertainty about the possibility of learning consistent dictionaries is the major matter. In planar 2D X-ray images this task is expected to be even more complex given rotational issues related to the projection of 3D volumes on a plane. Finally, an approach based on dictionaries built from patches could also be used to restore 2D radiography images, but it seems very difficult to define a dictionary general enough for this application.

As for digital radiography, the most considerable amount of work has probably been conducted on fluoroscopic images as it is an intrinsic low dose X-ray modality. Cerciello et al. (2012) compare different filters making both AWGN and signal-dependent noise assumptions. The experiments are conducted by denoising computed radiography (CR) images, that are intended to represent noise free images, corrupted with simulated mix Gaussian/Poisson noise. The results are evaluated according to signal to noise ratio (SNR), pick signal to noise ratio (PSNR), MSE and structure similarity image measure (SSIM, Wang et al. (2004)). The best results are given by the BM3Dc filter that is an adaptation of BM3D for signal-dependent noise after applying a VST proposed by Foi et al. (2008). In detail this method outperforms average spatial filters adapted to the properties of the noise (Gilboa et al., 2006), K-SVD, classical BM3D, wavelet-based methods that model the relation between coefficients by using hidden Markov models (Crouse et al., 1998) and other filters. Figure 2.4 is extracted from Cerciello et al. (2012) and compares the ground truth image with the input restored by using BM3Dc. We remark that while the edges of the ribs and the lateral borders of the spine are well preserved, the fine structures in the lungs and internal structures and edges of the vertebrae are smoothed. This may not represent a problem for fluoroscopic images that are used for good positioning and navigation of medical tools, but it surely is an issue for osteo-articular diagnosis for which EOS images are used. This problem of the BM3D filter has also been pointed out by Lebrun (2012) in his analysis of the method on natural color images: micro-textured regions are flattened out giving bad visual rendering. Moreover, note that this problem cannot be easily quantified by image quality measures.



(a) (b)

Figure 2.4: *Image restoration in digital radiography: (a) detail of a CR image assumed to be noise-free; (b) image (a) restored with BM3Dc filter after simulating noise on to obtain a SNR equal to 15 dB. The red arrows point at the loss of fine lung structures and attenuation of horizontal inter-vertebral edges. The images are extracted from the paper by Cerciello et al. (2012) and the arrows are added to facilitate our discussion*

Finally, restoration of digital X-ray images is very challenging, which probably explains the limited amount of works in the literature that directly address this issue. Moreover, X-ray images need to be enhanced to facilitate the diagnosis. Therefore, the bibliography on X-ray image processing mainly focuses on contrast enhancement where the noise is just considered as a variable that further complicates the problem. In the following sections we formalize the contrast enhancement problem, give the background theory and, specifically, in Section 2.2.3 we provide an overview of the state of the art methods on contrast enhancement of noisy X-ray images.

2.2 Contrast enhancement techniques

2.2.1 Introduction to the contrast enhancement problem

The contrast enhancement aims at increasing the visibility of structures of interest. It differs from restoration because it does not improve the image quality by correcting the observed data from disturbances that can be modeled, but rather consists in estimating from a low contrast image $\mathbf{u}$ another image $\mathbf{e}$ by means of a function ξ that equalizes the energy levels of $\mathbf{u}$. Formally, the contrast enhancement can be represented as follows:

$$\mathbf{e} = \xi(\mathbf{u}) \tag{2.17}$$

The definition of the function ξ is not trivial because the reduction of contrast of $\mathbf{e}$ with respect to $\mathbf{u}$, in medical imaging, changes from patient to patient. For example, digital radiography is used for the osteo-articular diagnosis because human tissues have different associated density levels that influence how X-rays are attenuated. However, due to strong attenuation, the shape of a bone, for example, may not be clearly visible. Another example is given by the slight variations of absorption within the same anatomical structure, which is the bone texture. For this reason, contrast enhancement can be used to increase the variance of signal level and, hence, facilitate the diagnosis. However, note that the low contrast results from a non-stochastic process, i.e. the attenuation has a strict physical meaning that exclusively depends on the relation between the signal strength and anatomical properties of the imaged region. Therefore, the function ξ can be chosen mainly according to empirical observations. Typically, the bias with respect to the low contrast image $\mathbf{u}$ needs to be minimized, that is, the mean brightness has to be preserved and no artifacts may appear on $\mathbf{e}$, while the variance needs to be maximized, that is, the visibility of discontinuities has to be increased.

The contrast enhancement is mainly applied to 2D X-ray image processing and, more rarely, to MR and CT imaging (Dippel et al., 2002). Such images are affected by noise, which significantly complicates the task of defining a good function ξ. Indeed, by maximizing the variance, the image $\mathbf{e}$ would have a bad quality because the noise would also be increased. This problem can be addressed through two different approaches. First, the restored image $\hat{\mathbf{y}}$ could be enhanced in place of $\mathbf{u}$ by computing an image $\hat{\mathbf{e}}$ as follows:

$$\hat{\mathbf{e}} = \xi(\hat{\mathbf{y}}) \tag{2.18}$$

The problem is that $\hat{\mathbf{y}}$ is just an estimation of $\mathbf{y}$ and, thus, it could contain residual noise, artifacts or some information could be missing, i.e. $\hat{\mathbf{e}} \neq \bar{\mathbf{e}}$, where $\bar{\mathbf{e}} = \xi(\mathbf{y})$. As a consequence, a second approach consists in limiting the increase of contrast in regions that are affected by noise while not erasing any information from the observed data $\mathbf{u}$. Formally, the result is obtained as follows:

$$\hat{\mathbf{e}} = (1 - \mathbf{n})\mathbf{e} + \mathbf{n}\mathbf{u} \tag{2.19}$$

Where the image $\mathbf{n} \rightarrow [0, 1]$ is called noise containment map and defines whether the observed image can be fully enhanced (i.e. $\mathbf{n}(x) = 0$) or not changed because only associated with noise (i.e. $\mathbf{n}(x) = 1$). Since noise and signal cannot be clearly separated, the image $\hat{\mathbf{e}}$ is the result of a pixel-wise weighted sum between the input and enhanced images. This approach has the advantage of preserving all information of interest, but the definition of the noise containment map $\mathbf{n}$ is not trivial. Indeed, while an excessive containment would produce a low contrast output image, in the opposite case noise residuals would appear on $\hat{\mathbf{e}}$.

Section 2.2.2 describes how to define the contrast enhancement function ξ and Section 2.2.3 presents the state of the art approaches for contrast enhancement of noisy X-ray images.

2.2.2 Multiscale image processing for contrast enhancement

A simple approach for enhancing the contrast in an image consists in equalizing the full range of intensity levels, which is defined according to the number of bits used

to encode the image. In practice, the histogram of the gray levels is transformed in order to follow a uniform distribution, which is why this method is called histogram equalization (HE). This enhances the global contrast but fails in boosting fine details and often introduces artifacts, such as saturation, that cause loss of the natural outlook of the image. Therefore, other methods have been introduced to solve these issues. Pizer et al. (1987) have proposed the adaptive histogram equalization (AHE) that equalizes the histogram within local windows. While this effectively enhances fine details, artifacts are also introduced and the noise in homogeneous areas is excessively increased. The contrast limited AHE (CLAHE) prevents noise amplification by constraining the contrast enhancement according to user-defined threshold values that clip the local histograms. However, the definition of these parameters are not trivial for X-ray images because of the signal-dependent noise. Please refer to Menotti et al. (2007) for a quite recent overview on HE-based contrast enhancement methods.

Rather than enhancing the contrast only in the image space, the information can be encoded according to different spatial frequencies, and, then, the visibility of precise image features increased by changing the energy distribution at each level. This idea is used in the unsharp masking (UM) method proposed by Ramponi et al. (1996). The image is divided into two frequency channels, and the high frequency one is boosted and then added back to the low frequency one. Nevertheless, the structures of medium size are not enhanced and the presence of noise strongly influences the results. Therefore, this method has been extended by employing a multiscale (MS) representation: the image is split up into $N_c > 2$ channels, that are separately processed. Contrast enhancement methods based on the MS analysis and synthesis are often used to process radiography images as they can increase the visibility of different types of details and, hence, adapt to multiple clinical needs or anatomical regions. For example, if the fine texture is relevant for the diagnosis, the channels that need to be mainly enhanced are high frequency ones, whereas medium spatial frequency can be important to delimit the envelop of big structures such as the spine.

2.2.2.1 Multiscale analysis

The MS decomposition can be linear or non-linear according to the type of filters used to progressively separate the input image in different levels. This is strictly related to the discussion on filters in image space (see Section 2.1.2).

Linear decompostion. In linear decomposition methods, the channels are obtained by iteratively applying linear filters to the input image. The Laplacian Pyramid (LP) (Burt and Adeldon, 1983) is a linear MS algorithm that encodes an image $\mathbf{u}$ in N_c band-pass images $\{\mathbf{d_0}, \mathbf{d_1}, \ldots, \mathbf{d_{N_c-1}}\}$. Formally, band-pass and low frequency residual images are obtained as follows:

$$\begin{cases} \mathbf{l_t} = \downarrow (g(\mathbf{l_{t-1}})) \\ \\ \mathbf{d_t} = \mathbf{l_{t-1}} - \uparrow (\mathbf{l_t}) \end{cases} \qquad t \in [1, N_c] \tag{2.20}$$

where $\mathbf{l_0} = \mathbf{u}$, $\downarrow$ and $\uparrow$ are the down-sampling and up-sampling operators, respectively, and $g(.)$ is a Gaussian-like filter, e.g. a binomial filter. The dyadic down-sampling doubles the spatial scale at each iteration to represent coarser and coarser

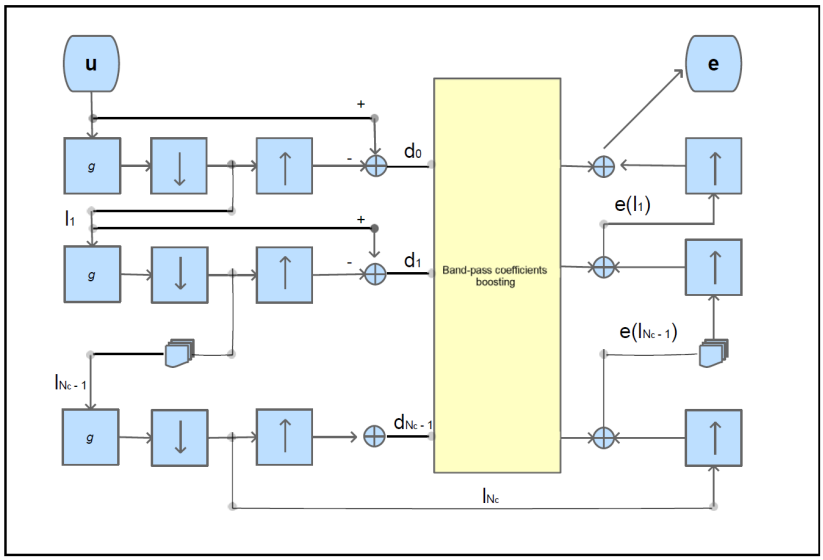
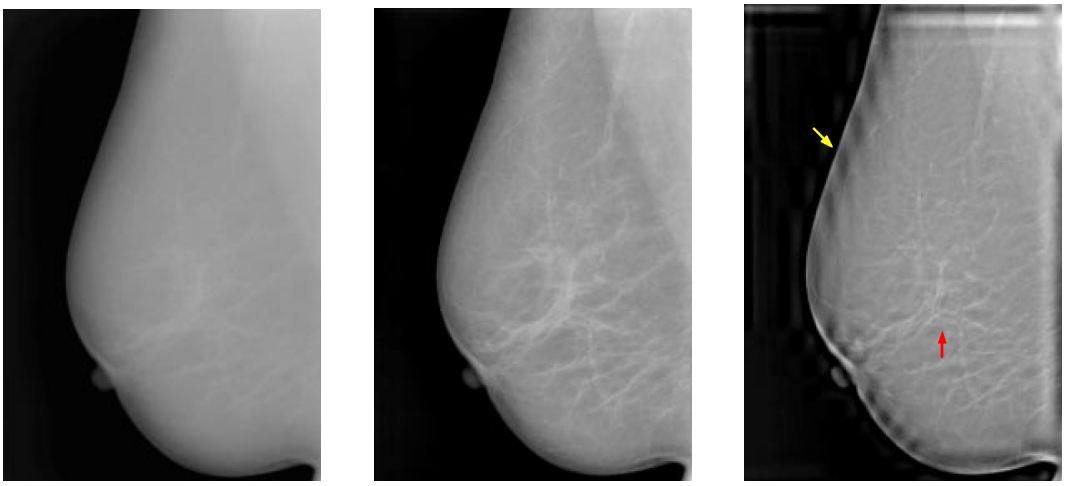


Figure 2.5: *Contrast enhancement with the multiscale LP representation: the left-hand side of the diagram is the analysis that builds the pyramid; the right-hand side is the synthesis that rebuilds the image from the scales after boosting the band-pass coefficients (central block).*

details as the number of iterations increases. The smooth transition between scales causes redundancy. Figure 2.5 shows the diagram of the LP algorithm and, specifically, the analysis process in its left-hand part. Stahl et al. (1999) have shown that LP efficiently represents X-ray images for the contrast enhancement of X-ray 2D images.

The fast wavelet transform (FWT) is another linear decomposition technique that presents some potential advantages compared to the LP such as: perfect decomposition due to the orthogonality of the wavelet bases, directional decomposition and better representation of noise. This method has been used for contrast enhancement of digital mammography images by Laine et al. (1995). The LP and FWT have been deeply compared by Dippel et al. (2002) with application to digital radiography. The authors show that, despite the good theoretical properties, the FWT is not well adapted to this application for two reasons. First, evident Gibbs artifacts appear near the edges that are due to the non-shift invariance of the orthogonal transform as presented in Section 2.1.3. While this problem could be solved by using an over-complete wavelet representation, there is another issue that is more constraining: the enhanced images do not preserve the good balance between the energies in the different frequency channels, creating an unwanted visual effect called *overshooting*. This effect does not depend on the choice of the boosting parameters, i.e. on how the details bands are modified to enhance the contrast (see Section 2.2.2.2), but it is rather due to the inverse transform that, by filtering the processed detail levels, enhances even more the details. Figure 2.6 shows the two different outlooks on a mammography exam: excluding the Gibbs effect, the FWT gives an unnatural rendering as edges are boosted too much, whereas LP involves only smooth operations in the decomposition which attenuate strong edges giving,



(a) (b) (c)

Figure 2.6: *Contrast enhancement of a mammogram: (a) Low contrast input image; (b) Contrast enhanced image with non linear remapping of detail coefficients obtained by encoding the input image with LP. The fine details are more visible while avoiding the introduction of artifacts, i.e. preserving the natural outlook of the input image; (c) Contrast enhanced image with non linear remapping of detail coefficients obtained by encoding the input image with FWT. The yellow arrow points at the Gibbs artifacts and the red one points at the fine details that give an overshot effect as excessively boosted. The images are extracted from the paper by Dippel et al. (2002) and the arrows are added to facilitate our discussion.*

hence, a more balanced effect (Dippel et al., 2002).

Therefore, LP is usually preferred to wavelets for processing X-ray images. However, if non-linearities are introduced in the reconstruction, which actually happens as detail coefficients are modified to increase the contrast, aliasing artifacts may appear (Li et al., 2005). In order to avoid this issue, the MS can be formulated in an undecimated framework, i.e. to avoid the sub-sampling and replace it by zero padding the filter g. This undecimated version of the LP is also known under the name of isotropic undecimated wavelet transform (IUWT, Starck et al. (2007)) and is used in this thesis.

Edge preserving decomposition. The main problem of linear MS is identified in halo artifacts that appear around strong edges because linear filters badly respond to discontinuities in the images. In detail, the smoothing operator g causes a frequency leakage that is more important as the response to the filter is strong. This effect can be compensated by adding the detail levels $\mathbf{d}_t$ in the synthesis, but if the coefficients are non linearly modified the leakage becomes visible under the form of strong halos around edges.

Some works applied to computational photography have addressed this issue by replacing the smoothing filter g with edge-preserving filters. For example, Fattal et al. (2007) have used the bilateral filter to combine in a multiscale framework images acquired under different lighting conditions. Farbman et al. (2008) have used weighted-least-squares (WLS) instead of the bilateral filter, which improves the results but

requires solving multiple linear systems. However, in both approaches, the images are not encoded in different spatial frequencies as the applications addressed are tone mapping and details exaggeration that are out of the scope of the medical context. As an alternative, some researchers have proposed to adapt classical linear decomposition techniques to avoid the frequency leakage. Fattal (2009) has used wavelet bases specific to each image to avoid the correlation between coefficients at edges and surrounding flat regions. Paris et al. (2011) have used LP and studied the correlations between coefficients to preserve it during the filtering. Basically, a local analysis is performed in order to avoid boosting coefficients of values larger than an user-defined intensity variation. Therefore, this could be interpreted as a patch-based analysis associated with the classical LP decomposition, which sensibly reduces halo artifacts, but is computationally expensive.

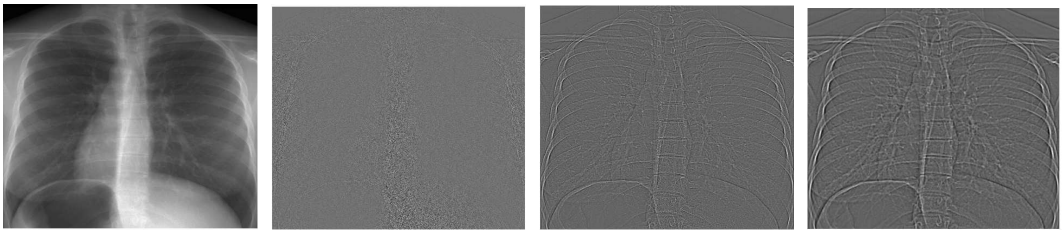
In spite of the advantages given by edge-preserving decomposition techniques that have been quantified on digital photography images, these methods have not been used for X-ray image processing, which can be explained by multiple reasons. First, in radiography it is important to precisely represent the information according to the spatial frequency because it allows adapting the algorithm to different clinical needs as previously pointed out. As for edge-preserving analysis, the decomposition depends on how the discontinuities are quantified. Since this depends on gradient features and parameters (see Section 2.1.2), it is not trivial to define edge-preserving analysis that provides the same completeness of analysis as given by linear decomposition. Secondly, these techniques address applications such as detail exaggeration in photography which can pose problems in medical applications where it is important to preserve the original natural outlook of non-processed images and the presence of noise cannot be neglected. In radiology, the halo artifacts are a problem mainly in presence of radio-opaque objects, e.g. prostheses, as shadows may appear around the prosthesis making difficult for the radiologist to evaluate the health of the surrounding tissues. Excluding this case, the halo artifacts are limited and can, hence, be tolerated (Dippel et al., 2002; Paris et al., 2011). In detail, the boosting functions need to take into account this aspect, which is why the non-linear detail enhancement is the most common solution (Stahl et al., 1999). Therefore, in this thesis linear decomposition techniques are used because coherent with the state of the art. Nevertheless, it would be interesting to study the contribution of edge-preserving techniques for processing images with wide metallic objects in the field of view. This aspect is not addressed in this thesis because not related to the low dose, but it may be worth to be taken into consideration in further works.

2.2.2.2 Multiscale synthesis

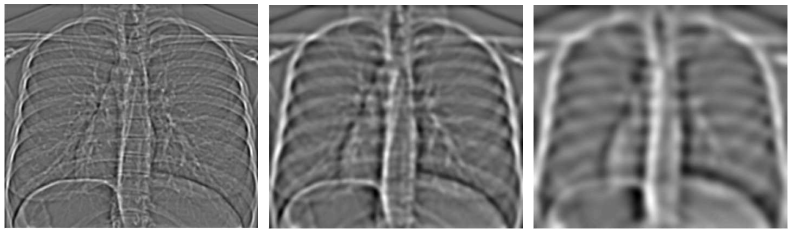
The retro-projection of an image encoded using IUWT simply consists in adding all the detail images $\mathbf{d_t}$ to the low frequency residual $\mathbf{l}_{N_c}$, that is:

$$\mathbf{u} = \mathbf{l}_{N_c} + \sum_{t=0}^{N_c-1} \mathbf{d}_t. \tag{2.21}$$

The right-hand side of the diagram in Figure 2.5 shows this process with the difference that the up-sampling operators are avoided as the decomposition is undecimated. Moreover, Figure 2.5 shows that the contrast enhanced image is obtained



(a) (b) (c) (d)



(e) (f) (g)

Figure 2.7: *Multiscale decomposition of an EOS image: (a) input, (b)* $\mathbf{d}_0$*: very fine texture and noise, (c)* $\mathbf{d}_1$*: fine lung structures (e.g.), (d)* $\mathbf{d}_2$*: medium lung structures (e.g.), (e)* $\mathbf{d}_3$*: edges of the vertebrae (e.g.), (f)* $\mathbf{d}_4$*: rib cage (e.g.), (g)* $\mathbf{d}_5$*: envelop of the spine (e.g.).*

by individually boosting the band-pass images $\mathbf{d}_t$, which is formally expressed as follows:

$$\mathbf{e} = \mathbf{l}_{N_c} + \sum_{t=0}^{N_c-1} \xi_t(\mathbf{d}_t) \tag{2.22}$$

where the contrast enhancement function ξ_t depends on the level t because its parameters are defined according to the targeted clinical need. For example, Figure 2.7 shows the MS decomposition of a chest EOS exam. In this case, if the lungs represent the ROI, then the band-pass images $\mathbf{d}_t$, $t = 0, 1, 2$ need to be enhanced while the others should not be modified. On the other hand, all the band-pass images should be enhanced if both the lungs and the thoracic spine are of interest.

The boosting functions ξ_t are defined according to a contrast measure that describes the activity, i.e. energy, in a given sub-band. In many works on X-ray images (e.g. Dippel et al. (2002)), the contrast measure is simply the magnitude of the coefficients. Li et al. (2005) have shown that it is however preferable to use average coefficient magnitudes in local windows rather than a coefficient-wise measure as it reduces the impact of strong edges by smoothing them and, consequently, limiting halo artifacts. Besides, it is more robust to possible outliers. Peli (1990) has discussed that the contrast in complex images does not only depend on the local energy at the band-pass levels, but also on the local background luminance, i.e. the low frequency residual of the band-pass levels. Formally, Peli's contrast is defined as follows:

$$\mathbf{a}_t = \frac{\mathbf{d}_t}{\mathbf{l}_{t+1}}, \quad t \in [0, N_c - 1]. \tag{2.23}$$

These are the main activity measures used to enhance the contrast of X-ray images. Please refer to Mantiuk et al. (2006) for a complete review on contrast measures

beyond medical applications.

Once a contrast measure is defined, the boosting function ξ_t has to restrict the enhancement in strong activity regions while increasing it in low activity ones in order to equalize the energy level in a sub-band image $\mathbf{d}_t$. Therefore, non-linear remapping functions are preferred to linear ones as they increase the visibility in low activity regions while limiting halo artifacts. In this thesis, the following model of boosting function is used:

$$\xi_t(\mathbf{d}_t) = \begin{cases} \mathbf{d}_t \left[(\gamma_t - 1) \left(1 - \frac{\mathbf{a}_t}{\bar{a}_t} \right)^2 + 1 \right] & \text{if} \quad \mathbf{a}_t < \bar{a}_t \\ \mathbf{d}_t & \text{otherwise} \end{cases} \tag{2.24}$$

where γ_t is the maximal gain at scale t and $\bar{a}_t$ is the activity measure threshold beyond which the band-pass $\mathbf{d}_t$ should not be modified.

The HE methods have been extended to the MS framework (Fan and Han, 2011; Jin et al., 2001). For example, Jin et al. (2001) have introduced the multiscale AHE (MAHE) that applies the classical AHE at the levels of a MS decomposition by setting the local window size according to the scale content (e.g. smaller sizes at coarser scales). This method improves the results compared to classical HE-based methods but is applied to CT chest scans and does not address, for example, cases of images strongly affected by noise. Fan and Han (2011) have combined band-pass coefficients boosting and HE of the scales, but the method involves a high amount of parameters which limits its flexibility.

Finally, the contrast enhancement by means of MS representation depends on the definition of an activity measure and boosting functions that change the activity distribution in the band-pass images. However, how to define optimal parameters and consistent activity measures from noisy decomposition is not deeply addressed in the literature. An overview of methods that take into account the presence of noise in contrast enhancement of X-ray images is provided in the following section.

2.2.3 Noise in contrast enhancement

The noise contaminates equally all frequencies. Therefore, both fine and coarse band-pass images contain unclean data. However, the iterative algorithm that encodes the image $\mathbf{u}$ entails a progressive noise reduction and, hence, the fine levels are more concerned by this issue (see Figure 2.7b). However, a naive enhancement of band-pass images generally increases the noise. By following the discussion introduced in Section 2.2.1, the problem can be addressed in two ways.

The first option consists in restoring the detail coefficients in order to enhance only signal information, which can be formulated in two different ways: to restore the image $\mathbf{u}$, i.e. to obtain the image $\mathbf{y}$, before encoding it with a MS framework or to apply a function ξ that jointly shrink noisy coefficients while boosting significant ones. For example, Sakata and Ogawa (2009) have proposed a method that denoise and enhance in sequence low dose digital radiography images. This method is divided into two steps: the image $\mathbf{u}$ is denoised with translation-invariant wavelet coefficients shrinkage and, then, enhanced by decomposing the resulting image with FWT and boosting the band-pass coefficients. The signal-dependent noise model is considered by applying a correction factor to a global threshold according to neighbor coefficients. Instead, Loza et al. (2014) have proposed to jointly enhance and

denoise dual-tree complex wavelet coefficients by modeling their local dispersion with a Cauchy distribution and apply it to light microscopy image processing. This method is more efficient compared to the method proposed by Sakata and Ogawa (2009) as it avoids the double multiscale decomposition. However, both methods present the drawback that eventual information could be loss or significantly attenuated at the denoising step. Besides, eventual residual noise artifacts surrounding strong edges would be significantly increased after the enhancement. In general, given the low contrast and strong noise, the WT decomposition does not allow perfectly separating noise from signal. Thus, it is very challenging to define consistent thresholds and/or boosting functions that both reduce the noise and enhance the contrast while avoiding artifacts.

The noise containment maps take into account the impossibility to clearly distinguish noise and information and, hence, the strength of the enhancement is controlled by means of a weighted average as presented in Equation 2.19. This idea is extended to the MS framework by applying a coefficient-wise weighted sum between the input band-pass level $\mathbf{d}_t$ and its fully enhanced version $\xi_t(\mathbf{d}_t)$:

$$\hat{\mathbf{d}}_t = (1 - \mathbf{n}_t)\xi_t(\mathbf{d}_t) + \mathbf{n}_t\mathbf{d}_t \tag{2.25}$$

where the noise containment maps show a dependency on the scale t because the presence of noise and the distribution of the information change through the levels as shown in Figure 2.7. The advantage of noise containment maps in comparison to filters is that no information is lost, as nothing from $\mathbf{d}_t$ is actually erased. For this reason, in X-ray image processing this approach is often preferred to deal with noise. However, it is not trivial to define correct $\mathbf{n}_t$ maps. While an overestimation of noise would cause a reduction of visibility of anatomical structures, an underestimation would overshoot the detail coefficients giving an unnatural effect. In the literature, Stahl et al. (1999) define $\mathbf{n}_t$ by using local density (gray level) and activity (e.g. local standard deviation) maps. Fan and Han (2011) implicitly estimate noise containment maps by comparing Peli's contrast measure before and after boosting. The main drawback of these methods lies in their dependency on some global user-defined parameters, e.g. the level of activity associated with noise and structures. This is a significant issue in our context for two reasons. First, there is a high intra-patient variability because the noise level significantly changes in an EOS full-body image and therefore globally defined parameters are sub-optimal. Secondly, there is a high inter-patient variability due to changes in patient age and morphotype, and acquisition conditions. Consequently, different parameter settings should be defined to face this large heterogeneity, which is very difficult to validate.

2.3 Evaluation of the image quality in digital radiography

In clinical routine the ground truth image $\mathbf{y}$ is not available and, hence, the evaluation of denoising and contrast enhancement techniques is not trivial. Therefore, specific validation processes are considered. Moreover, the techniques of evaluation differ according to the type of image, i.e. phantom or clinical images. In the following sections we present the retained techniques for the quantification of the image quality in digital radiography.

2.3.1 Phantoms

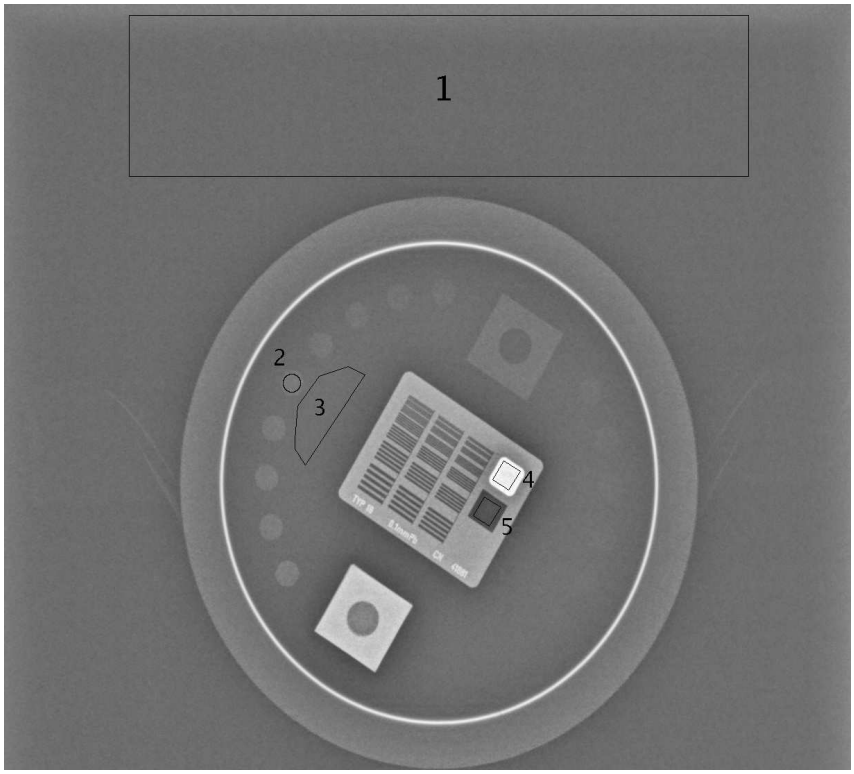


Figure 2.8: *Standard PHD5000 fluoro phantom with annotated ROIs used to compute the following measures: SNR from region 1; CNR from regions 2 and 3; DYN from regions 4 and 5. The high resolution grid at the center of the phantom is used to quantify the spatial resolution.*

The standard PHD5000 fluoro phantom (Figure 2.8) has been used to conduct experiments on phantom images. A set of tests can be built according to different signal strengths and different polymethyl methacrylate (PMMA) block thicknesses to be placed between the X-ray tube and the phantom to simulate different levels of absorption. The strength of the signal is defined as a function of the acquisition parameter setting and the tests in this thesis (see Section 3.4 for details) are conducted with the default parameter settings used in clinical routine according to the anatomical ROI and patient morphotype.

Figure 2.8 shows five ROIs that are used to compute the following quantitative measures: signal to noise ratio (SNR), contrast to noise ratio (CNR) and dynamic of the gray level distribution (DYN).

The SNR is defined as follows:

$$SNR = \frac{\mu\left(\mathbf{u}(\mathcal{R}_i)\right)}{\sigma\left(\mathbf{u}(\mathcal{R}_i)\right)} \tag{2.26}$$

where $\mathcal{R}_i$ is a homogeneous ROI (e.g. ROI number 1 in Figure 2.8), $\mu(.)$ and $\sigma(.)$ are two functions that compute the average and standard deviation values of the gray level distribution inside $\mathcal{R}_i$, respectively. This empirical definition of SNR

is consistent only if a constant region $\mathcal{R}_i$ is considered as, otherwise, $\sigma(\mathbf{u}(\mathcal{R}_i))$ would capture variations that are due to the signal rather than to noise. Given this definition, it is clear that the measure computed over images enhanced in contrast is reduced with respect to the input: while the standard deviation increases to optimize the contrast balance of the whole image, the average signal should not change compared to the input. The SNR performances of enhanced images can then be evaluated by computing how much the SNR drops with respect to the one computed over the input image. Nevertheless, this measure does not give a feedback on how structures of interest are processed as it simply takes into account a void region. Therefore, the evaluation needs to be completed by other measures.

The CNR is defined as follows:

$$CNR = \frac{|\mu(\mathbf{u}(\mathcal{R}_i)) - \mu(\mathbf{u}(\mathcal{R}_j))|}{\sigma(\mathbf{u}(\mathcal{R}_j))} \tag{2.27}$$

where $\mathcal{R}_i$ identifies the position of a homogeneous object of interest such as the disk in $\mathcal{R}_2$, and $\mathcal{R}_j$ is a void region used as background, e.g. $\mathcal{R}_3$. The CNR quantifies the compromise between the increase of the visibility of an object of interest and the boost of noise. The output CNR can be compared to the initial one: a decrease of CNR means that the gain of contrast is not sufficient to compensate the increase of noise and can be associated with the concept of *overshooting* (see Section 2.2.2.1) with the difference that the unnatural final outlook is due to excessive noise enhancement and not to detail exaggeration.

The DYN is defined as follows:

$$DYN = |\mu(\mathbf{u}(\mathcal{R}_i)) - \mu(\mathbf{u}(\mathcal{R}_j))| \tag{2.28}$$

where $\mathcal{R}_i$ is the most absorbing region (e.g. $\mathcal{R}_4$ in Figure 2.8) and $\mathcal{R}_j$ is the least absorbing one (e.g. $\mathcal{R}_5$ in Figure 2.8). The DYN represents how well the available gray levels are exploited, and is expressed as percentage of the gray level range, i.e. 65535 in EOS images. Therefore, it must not be confused with the dynamic of the detector, also called dynamic range, which is expressed in decibels. This measure is not as conventional as the SNR and CNR, but the DYN is important to evaluate EOS images because a full-body scan covers all the types of tissues, from cartilage to bones in strong absorption regions. Therefore, the DYN allows verifying the capability in representing this wide range of variations. All contrast enhancement algorithms increase the dynamic compared to input low contrast images and, hence, their performances can be evaluated according to the relative increase of DYN value.

The image quality measures introduced so far, do not quantify if fine details are well enhanced. In particular, this aspect is very important to study the effect of a noise reduction that precedes the band-pass detail enhancement. The high resolution grid at the center of the phantom (Figure 2.8) can be used for this purpose. In detail, it is used to evaluate the spatial resolution of a system which is measured in line pairs per millimeter (lp/mm)[1]. The resolution is computed at the finest set of visible lines on the grid and depends on the signal strength and PMMA block thickness to pass, in addition to detector properties. Finally, by comparing with the input image, it can be assessed whether the post-processing algorithms entail or not a loss in spatial resolution.

[1]This measure has been chosen in place of the modulation transfer function (MTF) because easier to read on display images and conceptually equivalent.

2.3.2 Proposed protocol for clinical images

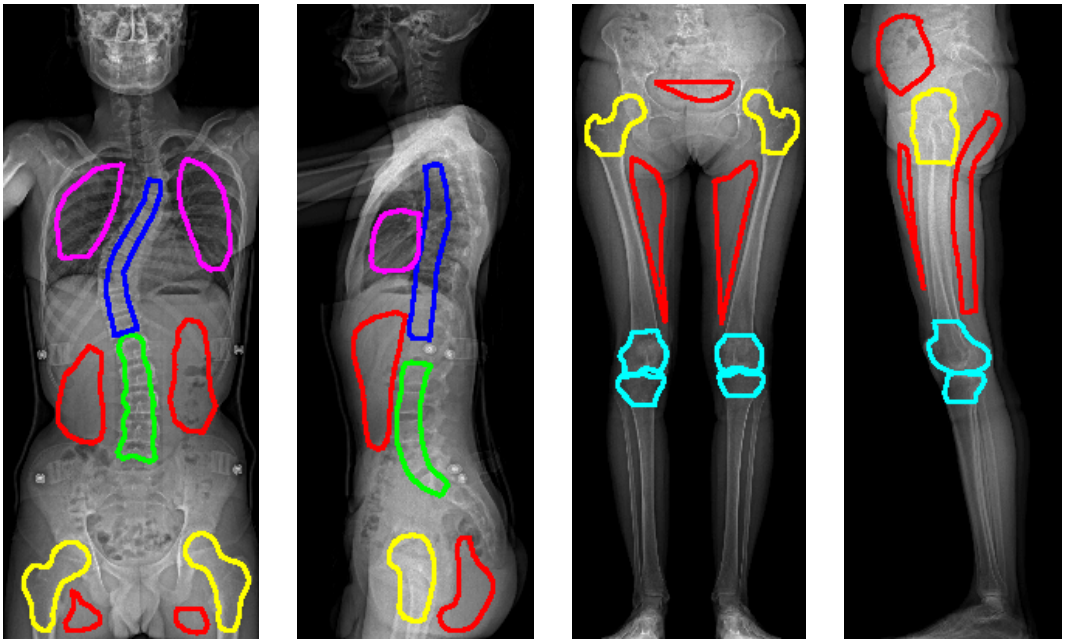
The measures used on the phantom rely on the presence of ideal homogeneous regions, which is not a valid hypothesis on clinical images as different tissues superpose each other. In order to address this point and due to the lack of knowledge of the ideal image $\mathbf{y}$, two measures are often used to quantify the performances of contrast enhancement algorithm in radiography: the average local variance (ALV) (Chang and Wu, 1998) and the contrast improvement index (CII) (Laine et al., 1995).

The ALV is the average of variance values computed in local windows (e.g. 5×5 pixels). A global evaluation is not significant as multiple image features are present in the pixel space and the amount of enhancement should not be the same everywhere (see Section 2.2.2.2). Therefore, the pixel space is divided into three disjoint regions and an ALV measure is associated with each one of them. In particular, the regions represent smooth (ALVS), detail (ALVD) and edge (ALVE) features. The ALVS measure indirectly quantifies the growth of noise because smooth regions are intended to represent regions void of pertinent information for the diagnosis. On the other hand, the ALVD measure aims at describing how much the visibility of structures of interest increases. The ALVE measure concerns the problem of halo artifacts as it is indirectly due to the enhancement of high activity features. Summing up, while low values of ALVS and ALVE are preferred because they quantify the noise enhancement and the presence of halos, high ALVD means good details processing. In the original approach, Chang and Wu (1998) define the regions by applying user-defined thresholds to the local variance image of the input. The thresholds should be manually adjusted exam by exam to capture anatomically relevant information. However, this concept may sometimes be ambiguous as X-ray images are very rich in different types of information. Besides, good thresholds are very difficult to define given the signal-dependent noise and, hence, the measures may depend on how the user interprets the correctness of the resulting ROI. As a consequence, the ALV relates to contrast but it is not a pure measure of it, and therefore should be used along with visual comparison.

As for the CII, it quantifies the improvement in terms of contrast on an enhanced image $\mathbf{e}$ with respect to the initial one. Therefore, it is a relative measure and only requires using the same techniques for quantifying the contrast in the input and enhanced images. The ALV measures can then be used too and, for example the relative CII can be expressed as follows to quantify the improvement in detail visibility:

$$CII = \frac{ALVD(\mathbf{e})}{ALVD(\mathbf{u})} \tag{2.29}$$

Rather than using a generic concept to characterize regions, we propose to use anatomically significant regions. In this way, joint denoising and contrast enhancement algorithms can be evaluated according to anatomical ROI and, hence, verify whether a given method behaves well in all the cases or its performances depend on the properties of the anatomical structures in the field of view. In practice, to each clinical image, regions associated with the following diagnostically significant structures are manually defined: lumbar spine ($\mathcal{A}_1$), thoracic spine ($\mathcal{A}_2$), proximal femur ($\mathcal{A}_3$), lungs ($\mathcal{A}_4$) and knees ($\mathcal{A}_5$). This approach permits to quantify the contrast in anatomical ROI, but these measures are influenced by the presence of noise too. Therefore, the noise contribution is quantified by measuring the ALV in



(a) (b) (c) (d)

Figure 2.9: *Examples of manually segmented regions used for computing anatomic ALV: red* $= \mathcal{A}_0$*; green* $= \mathcal{A}_1$*; blue* $= \mathcal{A}_2$*; yellow* $= \mathcal{A}_3$*; magenta* $= \mathcal{A}_4$*; cyan* $= \mathcal{A}_5$.

a void region ($\mathcal{A}_0$), i.e. with no signal of interest. Figure 2.9 shows four examples. Using these measures, another one, called anatomical contrast (AC), can be defined as follows:

$$\mathrm{AC} = \frac{1}{N_a}\left(\sum_{i>0}\mathcal{A}_i\right) - \mathcal{A}_0 \tag{2.30}$$

where N_a is the number of previously defined anatomic classes that appear in the image, and $\mathcal{A}_i = 0$ if the relative structure is outside the field of view. The first term of the equation coincides with the amount of contrast in the considered anatomical ROIs. However, since it is biased by the presence of noise, another term that only corresponds to noise, i.e. $\mathcal{A}_0$, is subtracted. It would be interesting to compute this measure for each considered anatomic ROI i, but this would require defining $\mathcal{A}_0$ in a void region of mean signal comparable to that of ROI i. Since such a region may not always be present, e.g. for the lungs in most of cases, this test has not been performed. The CII can be straightly used by taking the proposed anatomical ALV as contrast measure. First, the CII is computed for each anatomical ROI in the image to evaluate. Then a structure CII (SCII) is computed as follows:

$$SCII = \frac{1}{N_a}\sum_{i>0}\left(\frac{\mathcal{A}_i(\mathbf{e})}{\mathcal{A}_i(\mathbf{u})}\right) \tag{2.31}$$

Similarly, the Noise CII (NCII) is computed as follows:

$$NCII = \frac{\mathcal{A}_0(\mathbf{e})}{\mathcal{A}_0(\mathbf{u})} \tag{2.32}$$

Finally, we define an unbiased CII (UCII) as follows:

$$\mathrm{UCII} = \mathrm{SCII} - \mathrm{NCII} \tag{2.33}$$

This allows quantifying overshooting in clinical images. A negative value of UCII means that the noise is more enhanced than the signal, i.e. the higher visibility of structures comes at the price of an excessive boosting of noisy coefficients, whereas positive values of UCII indicate a relative higher enhancement of relevant information.

Visual comparison of differently post-processed images allows both completing the validation and verifying the consistency of the proposed anatomical ALV measures. The visual evaluation of clinical images is conducted in two ways. First, the magnitude of the gradients associated with the post-processed images are qualitatively analyzed to understand which method is more adapted for other further post-processing such as segmentation and registration. The Sobel operator is used to compute the gradient magnitude. Secondly, the feedback from clinicians is extremely valuable as it allows understanding if a given algorithm is interesting for use in routine practice. This feedback is quantified by means of questionnaires that are constructed to vote for the degree of visibility of anatomical structures or, if present, diseases.

The protocol here proposed will be used to evaluate our methods for processing EOS images in Section 3.5.

2.4 Conclusion

According to the analysis conducted in this chapter, we conclude that patch-based filters are a powerful tool for image restoration which, given the high amount of redundancy in diagnostic X-ray images, could be used to effectively reduce noise in EOS images. The application of the NL-means filter will be then studied in the following chapter and the choice for this particular method is supported by multiple reasons. First, it offers the best compromise in terms of efficiency and effectiveness, which is an important feature in medical applications as the image have to be processed in limited time. Secondly, the NL-means filter does not show artifacts as long as the patch size is properly fixed, which is not a property shared by other methods like BLS-GSM and, at high noise levels, BM3D. Nevertheless, the NL-means filter returns slightly smoother results compared to other patch-based filters, but this problem is shared by other methods too when processing X-ray images. For example, very fine details, such as micro bone structures, are very difficult to preserve because merged with noise. Therefore this aspect needs to be studied in detail and, in particular, the consequences on the contrast enhancement have to be identified. As for the latter aspect, a lot of works have been conducted with application to digital radiography, which has allowed concluding that the MS undecimated linear decomposition is the preferable representation to increase the contrast of this type of images. On the other hand, the effect of the presence of noise in contrast enhancement has been much less addressed. Noise containment maps are often the chosen option, but they depend on parameters that are very difficult to set given the heterogeneity of the content of the images. Therefore, we propose to combine patch-based filter and MS analysis to process EOS images in

order to robustly take into account the presence of noise in contrast enhancement algorithms. This aspect is extensively addressed in the following chapter and, to the best of our knowledge, represents an original approach for processing low dose X-ray images. Finally, note that we have also introduced in this chapter an evaluation protocol that is an original contribution of this thesis. This fulfills the need for clinical image quality quantification, which is not a trivial matter given the absence of ground truth. The comparison of the methods presented in the following chapter (Sections 3.4 and 3.5) builds on this evaluation protocol.

Chapter 3

Joint denoising and contrast enhancement of X-ray images

We discuss in this chapter how to post-process EOS images in order to facilitate the diagnosis for clinicians. It will be shown that, both denoising and contrast enhancement need to be considered to address this task. The contributions in this chapter include the description of the noise model, the formulation of the NL-means filter for X-ray images that are affected by signal-dependent noise, the definition of a contrast enhancement algorithm that takes into account the presence of noise while limiting manual parameter setting, and extensive study and comparison of different algorithms.

Organization of the chapter - Section 3.1 characterizes the noise that deteriorates the images to process and presents the percentile method that is used to estimate the noise levels in EOS images. Section 3.2 formulates the X-ray Non Local-means filter according to the noise model. Section 3.3 proposes a flexible contrast enhancement framework that is based on the introduced filter. Sections 3.4 and 3.5 present the validation on phantom and clinical images, respectively. Section 3.6 summarizes the contributions of this chapter.

3.1 Noise characterization

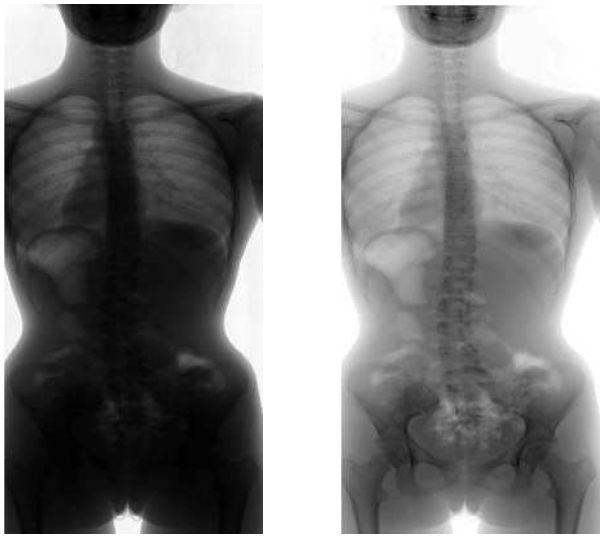
We study in this section how to model the noise, which is necessary to properly restore and enhance the contrast of an image. Moreover, the problem of the estimation of noise levels is addressed.

3.1.1 Noise model

Medical X-ray images are corrupted by signal-dependent noise. The noise at the detector is a combination of quantum and electronic noises that can be described by Poisson and Gaussian distributions, respectively. However, as shown by Damet et al. (2014), in EOS images the quantum noise contribution is predominant and, even at very low doses, the electronic noise is negligible. Therefore, the noise model can be approximated by a Poisson distribution. One of the properties of Poisson distributed random variables is that the expected value μ is equal to the variance σ^2. According to the SNR definition given in Section 2.3.1, the following relation is valid for images

corrupted by quantum noise:

$$SNR = \frac{\mu}{\sigma} = \sigma = \sqrt{\mu}. \tag{3.1}$$



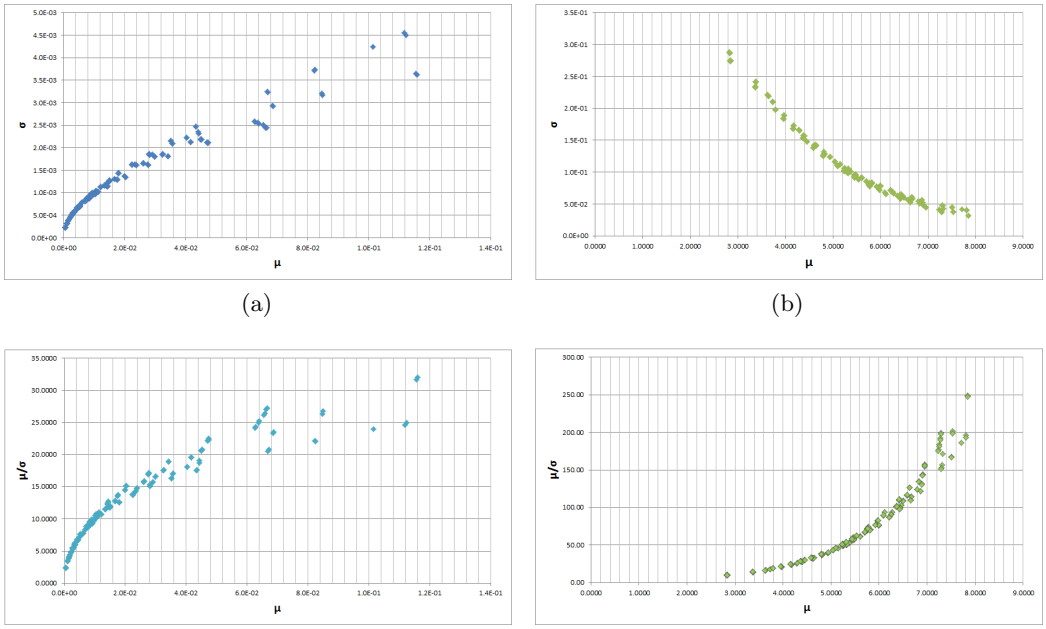
(a) (b)

Figure 3.1: *The acquired data are pre-processed to remove detector errors. EOS images can then be represented in the: (a) linear domain ($\mathbf{u}_r$); (b) logarithmic domain ($\mathbf{u}_l$).*

The relation in Equation 3.1 is rarely consistent for the X-ray images to be processed (Hensel et al., 2006), as the observed data are mapped into logarithmic domain to compensate for the exponential attenuation received by the X-rays passing through the body. This operation gives sense to the gray levels that are, then, linearly dependent on the matter thickness and density. Figure 3.1 shows the effect on the image after this conversion. The image in the linear domain is obtained by calibrating the acquired data, i.e. by applying offset, gain and dead-pixel corrections. However, in the resulting image (Figure 3.1a), the signal levels associated with different tissues are difficult to distinguish, whereas this is much easier after the logarithmic conversion (Figure 3.1b). Therefore, the image in the logarithmic domain is preferred for tasks such as denoising and contrast enhancement. For example, the features of the image can more easily be captured in this domain rather than in the linear one. However, the noise model also changes and, thus, it has to be characterized both before and after the conversion to the logarithmic domain.

A sequence of tests is then conducted to verify the Poisson hypothesis on observed images and to study the effect on noise after the logarithmic conversion. In these tests, $\mathbf{u}_r$ is the image in the linear domain and $\mathbf{u}_l = log(\mathbf{u}_r) + \epsilon$ is the corresponding logarithmic conversion. The value of ϵ is set to 10 for all the tests and is used just to avoid attributing negative values to the signal[1]. In total, 259 images of flat fields, i.e. homogeneous regions, have been acquired. Different combinations of kV,

[1]Since the values of $\mathbf{u}_r$ are mainly below 1, the corresponding values of $\mathbf{u}_l$ are negative.



(a) (b)

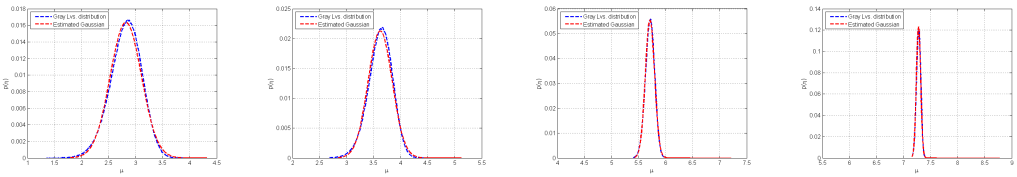
(c) (d)

Figure 3.2: *Study of the noise model before and after the logarithmic conversion: (a) $\sigma(\mathbf{u}_r)$; (b) $\sigma(\mathbf{u}_l)$; (c) $SNR(\mathbf{u}_r)$; (d) $SNR(\mathbf{u}_l)$. The noise follows a Poisson-like distribution in the image $\mathbf{u}_r$, but this property is lost after the logarithmic conversion.*

tube current, and thicknesses of a polymethyl methacrylate (PMMA) block placed between the X-ray tube and the flat field have been considered in order to simulate various examination conditions. On the contrary, the scan speed has been set to 4 because, as pointed out by Damet et al. (2014), it does not influence the noise model. The following parameter settings have been tested:

- kVp $= \{70, 90, 120\}$
- mA $= \{10, 20, 40, 50, 60, 63, 80, 100, 160, 200, 280\}$
- PMMA thickness in cm $= \{15, 20, 25, 30\}$.

Figure 3.2a shows the relation between signal and noise levels computed over each image $\mathbf{u}_r$. The resulting curve well approximates a root square function, which confirms the consistency of the assumption of Poisson noise (Equation 3.1) on images $\mathbf{u}_r$. On the contrary, after the logarithmic conversion (Figure 3.2b) the curve does not exhibit this trend. Indeed, the function is approximately monotonically decreasing, which implies that the lower the signal at the detector the higher the noise is. In other words, the logarithmic function causes an inversion of the noise curve, i.e. the opposite behavior with respect to the noise response in the linear domain, which is coherent with the results presented by Hensel et al. (2006). Note that the noise levels increase in the images $\mathbf{u}_l$, but this is simply due to the logarithm that, being monotonically increasing and concave, boosts the variance at low signal levels while flattening it at strong ones.



(a) (b) (c) (d)

Figure 3.3: *Verification of the consistency of the assumption of local Gaussian distribution. The blue line is the normalized histogram of the gray levels within the flat field. The red line is a Gaussian function with mean and variance defined according to the maximum like estimation. The signal level increases from left to right and the SNR computed from* $\mathbf{u}_l$ *images are, respectively, equal to: (a) 2.37; (b) 3.90; (c) 11.58; (d) 26.12.*

The noise curve in Figure 3.2b does not follow a Poisson law anymore, but it can still be easily modeled. In particular, at a given signal level, we can assume that the noise follows a Gaussian distribution with signal-dependent standard deviation. Figure 3.3 shows the consistency of this hypothesis: if the mean and variance parameters of a Gaussian function are fixed to the empirical mean and variance, the resulting function perfectly matches with the distribution of gray levels inside the flat field. Note that this is valid regardless the value of the signal, even when it is extremely low (Figure 3.3a). Therefore, the following relation between the ground true image $\mathbf{y}$ and the observed image $\mathbf{u}_l$ is assumed:

$$\mathbf{u}_l = \mathbf{y} + \sigma(\mathbf{y})\varepsilon = \mathbf{y} + \eta \tag{3.2}$$

where the probability density function associated with the random variable η is defined as follows:

$$p(\eta) = \frac{1}{\sigma(\mathbf{y})\sqrt{2\pi}} \exp\left(-\frac{\eta^2}{2(\sigma(\mathbf{y}))^2}\right) \tag{3.3}$$

This conclusions are coherent with the observations made by Hensel et al. (2006).

We complete our analysis by computing the SNR values on both $\mathbf{u}_r$ and $\mathbf{u}_l$, which are reported in Figures 3.2c and 3.2d, respectively. The noise level and the SNR curve have a comparable trend, but the scales of the values are different due to the calibration and, in particular, to the gain and offset corrections that follow the image acquisition. This effect can be balanced by means of a linear transform, $\tilde{\mathbf{u}}_r = a\mathbf{u}_r + b$, where a and b are two parameters that enforce linear dependency between signal and variance.

The logarithmic conversion does not change the monotonically increasing behavior of the SNR, but the function is convex rather than concave, as the more the signal increases, the lower the noise gets.

Finally, our study shows that the logarithmic conversion invalidates the assumption of Poisson-distributed noise. However, the noise is locally distributed as a Gaussian variable whose variance decreases as the signal increases. It remains to be established which image is more adapted for joint denoising and contrast enhancement, and how to estimate the signal-dependent noise levels on clinical images.

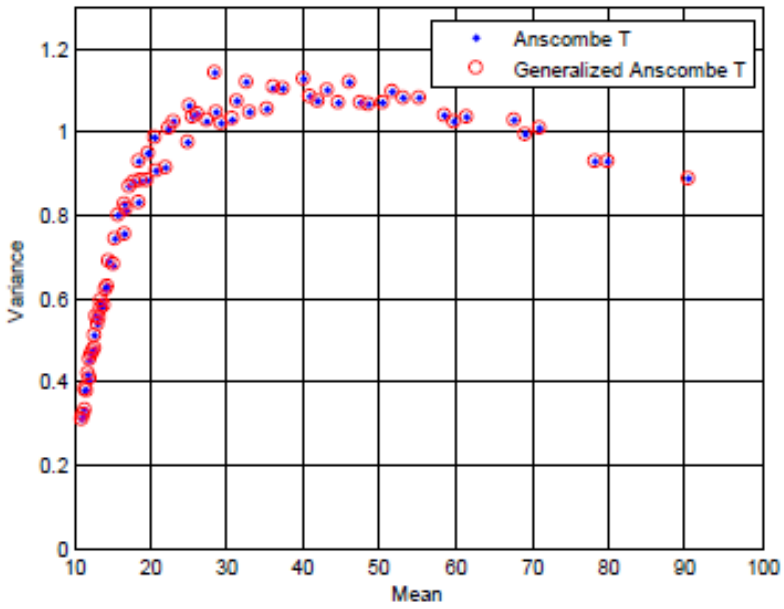


Figure 3.4: *Results obtained by applying the standard and the generalized Anscombe transform to EOS images. The transform fails in stabilizing the variance at low/medium signal levels.*

3.1.2 The drawbacks of the Anscombe transform on EOS images

Given the Poisson-distributed noise in $\mathbf{u}_r$ images, the Anscombe transform could be used to stabilize the variance. The images could then be denoised with classical methods that are theoretically optimal under AWGN assumptions. In particular, the parameters of the Anscombe transform can be estimated by fitting the following model:

$$\arg\min_{a,b,c} \left[\left(1 - \sigma^2 (a\sqrt{b\mathbf{u}_r + c}) \right)^2 \right]. \tag{3.4}$$

This transform estimates an image whose variance is equal to one and also corrects the linear difference in scale between mean and variance observed in the previous section. More generally, by assuming that a linear transform has been applied to the input image $\mathbf{u}_r$ in order to obtain the image $\tilde{\mathbf{u}}_r$ such that $\sigma(\tilde{\mathbf{u}}_r) = \sqrt{\mu(\tilde{\mathbf{u}}_r)}$, the following variance stabilizing transform can be applied (Anscombe, 1948):

$$\tilde{\mathbf{u}}_r^a = 2\sqrt{\tilde{\mathbf{u}}_r + \frac{3}{8}}. \tag{3.5}$$

When both quantum and read-out noises affect the input image, the generalized Anscombe transform can be used (Murtagh et al., 1995):

$$\tilde{\mathbf{u}}_r^{ag} = \frac{2}{a}\sqrt{a\tilde{\mathbf{u}}_r + \frac{3}{8}a^2 + \sigma_e^2 - a\mu_e}. \tag{3.6}$$

where a is a parameter to estimate, and μ_e and σ_e^2 are the mean and the variance of the read-out noise, respectively. Figure 3.4 shows that $\tilde{\mathbf{u}}_r^a \approx \tilde{\mathbf{u}}_r^{ag}$, which is logical

because the electronic noise is negligible in EOS images. Besides, it reveals that the Anscombe transform applied to EOS images does not stabilize well the variance that, at low/medium signal levels, is still signal-dependent. For high values the estimation is approximately correct, but the function decreases at high signal values due to the saturation of the detector where the absorption is absent or very low.

This analysis is sufficient to avoid applying the Anscombe transform to process EOS images, but intuitive reasons further support this choice. The limits related to the Anscombe transform (see Section 2.1.5) invite to use with caution this technique on digital X-ray images applied in diagnostics. On these images, it is indeed important to avoid any sort of bias as gray levels have a physical meaning, i.e. X-ray attenuation. The non-linear modification of data and potential errors due to the inverse transform could introduce a bias in this relation and, then, distort the diagnosis.

Finally, we do not use Anscombe transform on EOS images according to the aforementioned inability in stabilizing the variance and the presented empirical risks. Therefore, the image needs to be directly denoised by taking into account the signal-dependent noise. Section 3.2 addresses this aspect, that requires to know the noise level. In the following section, we explain how to estimate σ from clinical images.

3.1.3 The percentile method for noise level estimation

A noise level can be associated with images that come directly from the imaging system by estimating some parameters (Foi et al., 2008). This task is often considered simple, but this depends on the complexity of the detector and, mainly, may not be achievable in practice because the information in the data could be lost due to image processing steps such as sampling, denoising, contrast changes and compression (Lebrun et al., 2012). In the case of interest, in Section 3.1.1 we have pointed out that, after the pre-processing steps that correct detector errors, the noise follows a Poisson-like distribution but the noise variance is not equal to the mean signal and, besides, after the logarithmic conversion the noise model completely changes. As a consequence, the knowledge of the acquisition configuration is not sufficient to correctly estimate σ values.

This problem has been recently highlighted by Lebrun et al. (2012) who have also reviewed the methods for noise estimation from a single image that, in general, rely on a local statistical analysis. We use the percentile method (Ponomarenko et al., 2007; Colom and Buades, 2013) that, according to the literature, is a very robust technique for measuring the noise standard deviation from a single image. Please refer to Lebrun et al. (2012) for a comparison of different techniques.

The percentile method has been introduced by Ponomarenko et al. (2007) and it represents an image as an histogram of variance values computed at small windows all over the image. High frequency features such as edges and texture will be more probably located in the right-hand side of the histogram and, thus, the noise level can be estimated from low energy blocks, i.e. those at the left of a low percentile of the histogram.

The first step of the percentile method consists in high-pass filtering the observed image $\mathbf{u}$, which returns an image that is denoted $\mathbf{u}_h$. By applying this filter, the deterministic component due to the signal is not considered or, in other words, given the sparsity of the image, the energy is concentrated at strong variations. On the

other hand, the noise is uniformly distributed over the pixel space, i.e. it is not sparse. Therefore, the majority of the windows of half-size b will present a dominant noise component. Some examples of suitable high-pass filters are the discrete Laplacian, the Discrete Cosine Transform (DCT) and the Wavelet coefficients.

A measure of energy is then associated with each window, for example this measure may be the variance of the gray level distribution. Then, the histogram of variance values h_{σ^2} is obtained and the biased noise level is computed as a low percentile of this histogram. The percentile is low because only the most regular, i.e. those at lower energy, blocks are taken into account to avoid affecting the measure with information associated with image features. Since the MSE estimation would be obtained by fixing the percentile at 50% (Colom and Buades, 2013), the noise level estimation is biased, but this bias can easily be corrected by a factor that depends on the percentile value, the window half-size b and the chosen high-pass filter.

We use a 10% percentile, $b = 7$ and the first detail level of the multi-scale decomposition (see Section 2.2.2.1) as high pass filter in the tests presented in the following sections. Given this configuration, we have computed the optimal linear correction factor by using a training set of flat field images like those that compose the database used in Section 3.1.1. According to these tests, the correction factor has been set to 1.3805, i.e. the median of factors obtained to correct the bias of σ^2 computed over the images in the training set. In order to validate this value, we have computed the difference between the estimated SNR values and the ground truth ones over another set of flat field images disjoint with the training set[2]. The root-mean-square error (RMSE) and the root-maximum-square error have resulted equal to 0.35 and 1.82, respectively, on the images $\mathbf{u}_r$. This proves the good quality of the noise level estimation regardless the change in X-ray signal strength and degree of absorption. This calibration technique is different from the percentile with learning (Lebrun et al., 2012). In this case, a set of training images is used to learn the bias according to the features of the image. This would be potentially interesting in our case as we process images that share the same properties. However, this requires the noiseless images and to simulate the noise, which is not possible in our application context. In any case, Lebrun et al. (2012) have shown that when the noise level is higher than 10 (for images coded on 8 bits) the noise component is clearly predominant in local windows and, hence, the classical percentile method is enough. In other words, the advantage of the percentile with learning consists in better taking into account edges and texture in the estimation when the noise is very low and, hence, in avoiding σ^2 overestimation at low noise levels. Considering the trends of the functions in Figures 3.2a and 3.2b, the noise level risks to be overestimated in strong and low absorption regions for, respectively, $\mathbf{u}_r$ and $\mathbf{u}_l$. Intuitively, the first case is more dangerous because in strong absorption regions there is a high amount of diagnostic information. On the other hand, the overestimation of the noise level in low absorption regions is less problematic because they are mainly associated with soft tissues at the border of the envelop of the patient or background where eventual over-smoothing is not an issue. However, the lungs have also low density and in this region all the information must be preserved.

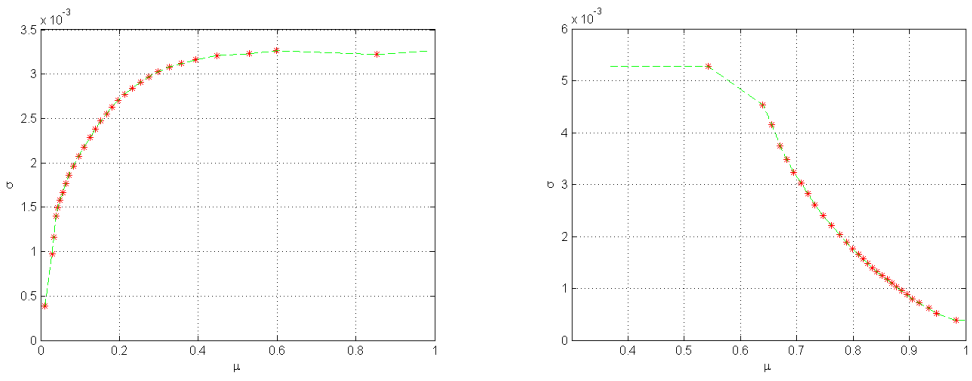
In a clinical image there are different degrees of absorption due to different

[2]We use the SNR because the scale of values is more intuitive to interpret.

tissue densities and, hence, multiple noise levels need to be estimated according to changes of the signal. Colom and Buades (2013) have recently proposed a noise estimation method that does not require the assumption of homoscedastic (i.e. signal-independent) noise and consists in a block-based formulation of the percentile method. First, a value μ that is equal to the average of the intensity levels of the pixels $x_i \in B$ is associated with each block B of half-size b in the image as it estimates the underlying signal level. Then, the intensity values are classified into n non-overlapping bins $i = 1, 2, \ldots, n$ of equal size. The sub-division into these bins is simply achieved according to the gray levels and, then, a given bin can concern non-connected regions of the image. In each bin i, the noise is assumed to follow a Gaussian distribution and signal-independent. The classical percentile method is used to determine the noise standard deviation σ_i associated with each bin i and the signal level of the bin μ_i is also computed as median$(\bar{\mu}_i)$, where $\bar{\mu}_i$ is the set of average intensity values belonging to the bin i. The n points $\{(\mu_1, \sigma_1), (\mu_2, \sigma_2), \ldots, (\mu_n, \sigma_n)\}$ are associated with disjoint regions and, hence, can be linearly interpolated to obtain a noise curve, i.e. a function that defines the noise standard deviation depending on the signal.

3.1.4 Application to clinical EOS images

The block-wise implementation of the percentile method can be used to estimate Poisson noise. The underlying assumption is that a Poisson process can be simulated as Gaussian at a fixed amount of signal, which is true as long as the number of photons that reach the detector is high enough (Colom and Buades, 2013). Therefore, this method can be used to estimate noise levels from $\mathbf{u}_r$ images, but it should be taken into account that the estimation may not be completely reliable at very low amount of signal, whereas, according to Section 3.1.1, on images $\mathbf{u}_l$ the noise is locally Gaussian-distributed regardless the signal strength.



(a) (b)

Figure 3.5: *Estimation of the noise levels σ on: (a) a clinical image before the logarithmic conversion (Figure 3.1a); (b) after the logarithmic conversion (Figure 3.1b).*

Figures 3.5a and 3.5b show the σ curves computed from the images in Figure 3.1a and 3.1b, respectively. There are 29 bins (red marks in Figures 3.5a and 3.5b) and each one contains 2.5×10^5 pixels. The trends of the two curves correspond to those of the functions in Figures 3.2a and 3.2b, which shows that the percentile method is

a valid approach to estimate noise levels on clinical X-ray images. This assessment has been validated by a sequence of tests conducted on about 200 images.

Finally, both images could be taken into account for denoising as the noise model can be characterized in both cases. However, considering that the image needs to be converted to the logarithmic domain to enhance the contrast, it is in practice more convenient to process the image $\mathbf{u}_l$, which from now on in this chapter is denoted by $\mathbf{u}$, in order to jointly denoise and enhance the contrast of the image. In the other case, we would have to denoise the image, pass to the logarithmic domain and then enhance the contrast. Nevertheless, given the results presented in Section 3.1.2, it is not convenient to use the Anscombe transform and, besides, the noise model does not perfectly correspond with the theoretical Poisson distribution. These observations on the noise properties lead us to avoid directly denoising the raw data. The image $\mathbf{u}_r$ suits better for estimating information relative to the acquisition, e.g. exposure levels, and it will be used in Chapters 4 and 5.

According to the conclusions from the previous chapter (see Section 2.4), in the following section we propose a formulation of the NL-means filter adapted to the noise model affecting the image after logarithmic correction.

3.2 X-ray Non Local-means Filter

3.2.1 Formulation

The noise model described in the previous section enables a straightforward extension of the NL-means filter for digital radiology. The noise is signal dependent and the same parameters cannot be used to denoise the whole image. However, according to the analysis presented in Section 3.1.1, the σ_η values are approximately constant in small windows because the tissues therein have similar densities, i.e. the map σ changes smoothly. By considering that the NL-means filter smoothing parameter h is proportional to the noise standard deviation σ (Buades et al., 2005b), i.e. $h^2 = 2k\sigma^2$, the distance between patches can be modified as a function of the underlying signal strength. The proportional factor k adjusts the automatic estimation of σ. Coupé et al. (2008) have pointed out that the optimal value of k depends on the strength of the noise affecting the data, e.g. $k = 0.5$ for low noise levels and $k = 1$ for medium-high noise levels. However, this is mainly due to the tendency in over-estimating σ at low noise levels (Colom and Buades, 2013). Considering that the percentile method gives a robust estimation (Lebrun et al., 2012) and that, in the worst case, the only region affected by over-smoothing would be the background where the X-rays are not attenuated, k is fixed to 1 in our tests.

The distance between patches takes into account the dependency on the X-ray absorption according to the following definition:

$$d_X(P_i, P_j) = \frac{\|\mathbf{u}(P_i) - \mathbf{u}(P_j)\|_2}{2{\sigma_i}^2|P|}. \tag{3.7}$$

Note that we consider the noise level associated at the window centered at the pixel x_i, which explains the dependency of σ on i. Therefore, this equation implies that $\sigma_i \simeq \sigma_j \ \forall j \in W$, where W is the window of small size where the patches are compared. This hypothesis is in general valid because the tissue density values change smoothly, i.e. strong edges are rare in X-ray images (Dippel et al., 2002).

Some exceptions may occur, for example, in presence of a metallic objects or at the borders of the lungs. Nevertheless, even in these cases we have not observed a significant drop of the filter performances.

The weights are then simply computed as follows:

$$\varsigma_X(i,j) = \exp\left(-d_X(P_i, P_j)\right). \tag{3.8}$$

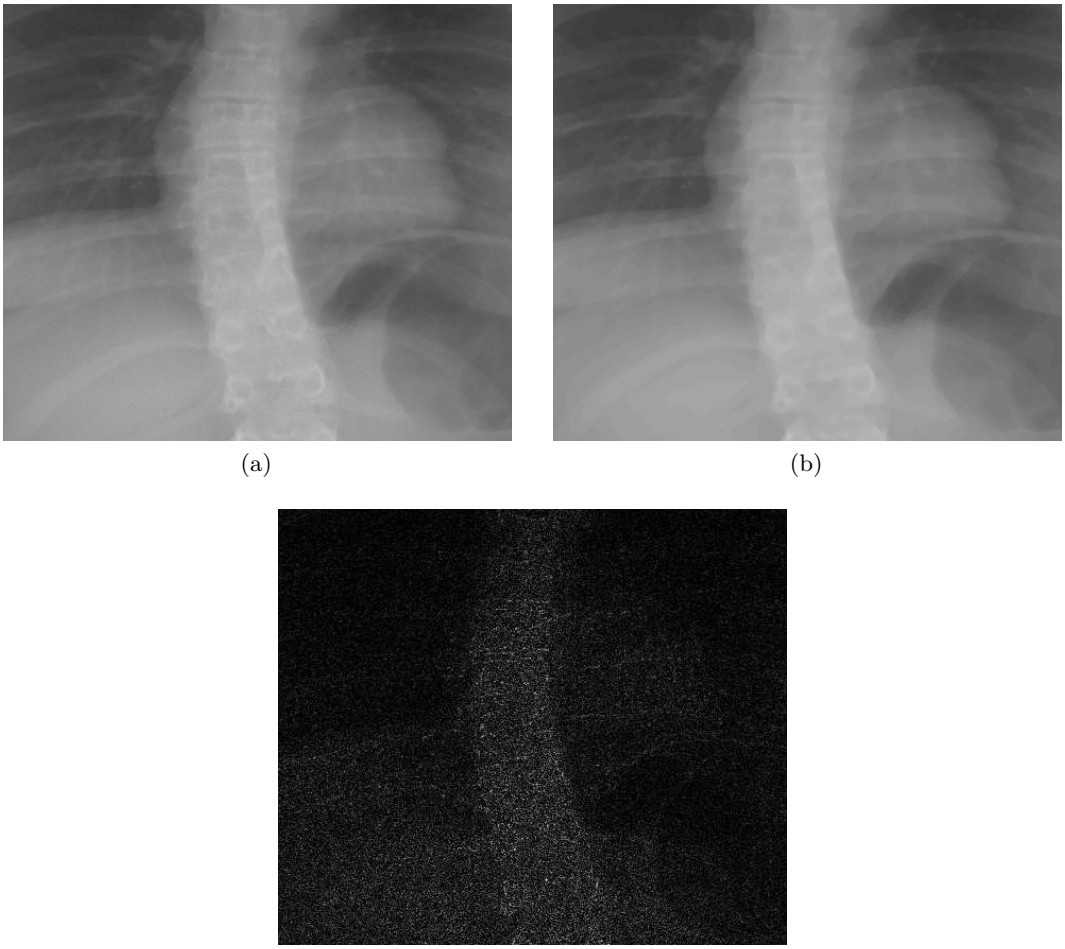
In conclusion, by considering the signal dependent nature of the noise (Section 3.1.1), a new formulation of the NL-means filter is introduced to denoise X-ray images. The proposed filter is called XNL-means, which stands for X-rays Non Local Means filter. It is worth noting that the automatic definition of σ avoids the manual parameter setting and, hence, it is more robust to changes in X-ray acquisition settings and patient morphotypes. However, the patch and the window sizes still need to be defined by the user. In our work patch and window half-sizes are fixed to $p = 2$ and $w = 7$, respectively. These parameters have been set according to empirical observations as the best compromise between computational time and quality of the results, mainly verifying the absence of artifacts (patch effect). Furthermore, the Euclidean distance is used to compare patches and, hence, the XNL-means filter can be efficiently implemented by using integral images, which is necessary given our context. Besides, a further reduction of time can be obtained by combining parallel execution to the integral images-based formulation (Darbon et al., 2008).

3.2.2 Qualitative evaluation of the X-ray Non Local Means Filter

In this section we briefly present the qualitative performances of the proposed filter in terms of noise reduction and preservation of the anatomical structures.

Figure 3.6 reports an example of restoration of the thoracic spine. Figure 3.6c shows the absolute difference between the observed image $\mathbf{u}$ and the restored one $\hat{\mathbf{y}}$ and highlights the good properties of the filter as it removes a large amount of noise while preserving anatomical structures such as the envelops of the vertebrae. Moreover, the filter adapts to different levels of absorption as the noise reduction is more important in high density tissues, e.g. the spine and the soft tissues in the abdomen, than in the low ones, e.g. the lungs. Nevertheless, with respect to the observed image $\mathbf{u}$ (Figure 3.6a), some vertebral structures in the denoised image $\hat{\mathbf{y}}$ are slightly attenuated. Besides, it is not trivial to tell if only noise is removed from bone tissues or if some micro-texture is also lost because the two are mixed up. Even for a clinician it is difficult to distinguish the noise from fine bone texture. Therefore, it is not trivial to evaluate the XNL-means filter on its own and, hence, its performances need to be quantified by considering the contrast enhancement too, which is provided in Sections 3.4 and 3.5.

In a second example we discuss the capability of the proposed filter to adapt to different ROIs by exploiting the noise level estimation and, hence, to avoid manual parameter setting. We consider two ROIs extracted from the same image that contain the femoral head and the knee. In order to study this aspect, we just enhance the restored image $\hat{\mathbf{y}}$, i.e. Equation 2.18 is used. Figure 3.7 shows the two input images (Figures 3.7a and 3.7e) and the corresponding outputs $\hat{\mathbf{e}}$ obtained by estimating $\hat{\mathbf{y}}$ with the NL-means filter with a low h value (Figures 3.7b and 3.7f) and a high h value (Figures 3.7c and 3.7g), and with the XNL-means filter (Figures 3.7d and 3.7h). As for the femoral region, a too low value of h leaves a significant



(a) (b)

(c)

Figure 3.6: *Properties of the XNL-means filter: (a) Observed image* $\mathbf{u}$*; (b) Denoised image* $\hat{\mathbf{y}}$*; (c) Absolute difference* $\|\mathbf{u} - \hat{\mathbf{y}}\|$*.*

amount of noise residual, whereas both a higher h value and the automatic method reduce the noise much more while preserving image features such as the borders of the femur and the cotyloid cavity, i.e. where the femur meets the pelvis. However, high h values completely smooth the bone texture in the knee, whereas this does not happen if low h values are used or if the XNL-means filter is applied. As a consequence, this example shows that the XNL-means offers a good compromise between noise reduction and preservation on anatomical structures, while avoiding manual parameter setting.

By looking at the images $\mathbf{u}$ it is clear that the contrast of the images has to be enhanced to facilitate the diagnosis. For example, the coarse borders of the spine in Figure 3.6a are easy to distinguish, but the fine structures of the vertebrae are almost invisible. Therefore, it is necessary to understand how the XNL-means can be used to optimize the balance between increase of structure visibility and noise. For instance, Figures 3.7d and 3.7h show the outputs obtained after increasing the contrast of the filtered image, but we need to assess if this quality is convenient for a

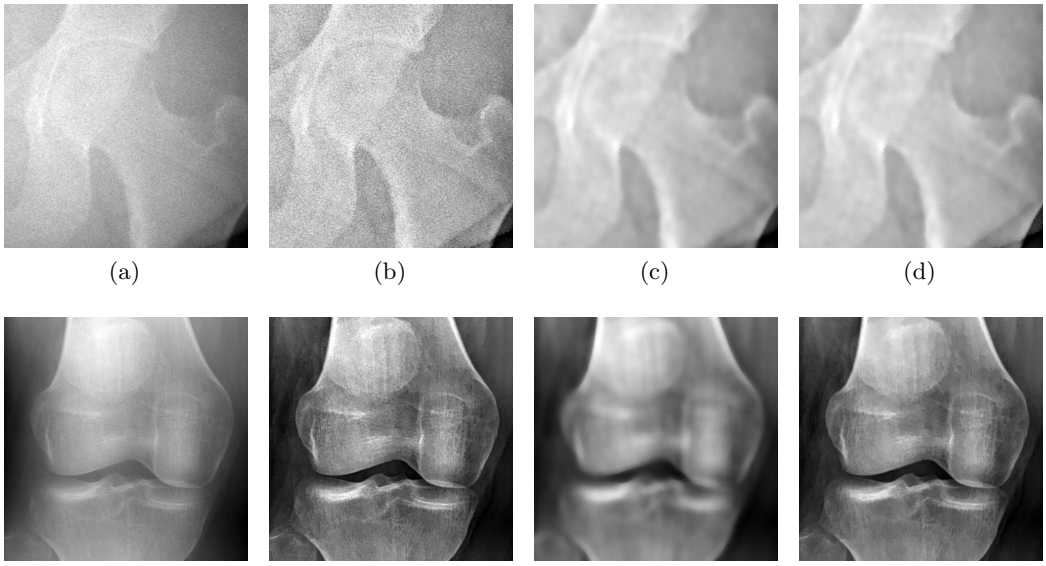


Figure 3.7: *XNL-means automatic parameter setting according to the estimated noise level. Two ROIs are considered as examples: a strong absorption ROI, i.e. femoral head (a-d), and a low absorption one, i.e. knee (e-h). The observed images (a,e) are filtered and, then, enhanced with the technique proposed in Section 3.3.3. The parameter smoothing parameter h is set: manually low (b,f); manually high (c,g); automatically (e,h). The proposed XNL-means filter adapts well to both regions by relying on the noise level estimation and, hence, allows avoiding manual parameter setting.*

given clinical goal and, otherwise, which solution to propose. The evaluation of the method on clinical images is completely discussed in Section 3.5, but we first clarify the role of the XNL-means filter with respect to the multiscale (MS) analysis and the contrast enhancement.

3.3 Contrast Enhancement of images affected by noise

3.3.1 Multiscale analysis of the X-ray Non Local Means Filter

According to the discussion in Section 2.2.1, a way to enhance the contrast of noisy digital X-ray images consists in fully boosting the band-pass images $\mathbf{d}_t^{\hat{y}}$ that encode the output of the XNL-means filter $\hat{\mathbf{y}}$. Nevertheless, this option raises several questions. First, some signal components may be erased by the filter leading to an incomplete MS analysis and, hence, some relevant clinical information may be attenuated or even missing in the output. This aspect can be observed on the images previously given as examples to study the performances of the XNL-means filter: Figure 3.7d shows an excessive regularization of the region inside the femur and, even if this is justified by the low visibility of texture in the source image (Figure 3.7a), the final outlook may not be accepted for diagnosis; similarly some texture patterns inside the knee (Figure 3.7h) are considerably attenuated. Secondly,

the noise is uniformly distributed at all frequencies. Since patch-based filters remove only the high wavelength component of noise (Lebrun et al., 2012), the scales at low frequency are still corrupted and, therefore, some residual noise could appear on the output. Finally, the non-local mean filter can suffer of *noise halo* (Deledalle et al., 2011), i.e. residual noise caused by strong coefficients around edges that would be even more increased after the enhancement step. However, on our images this effect is less problematic as strong edges correspond to strong activity values and by using suitable boosting functions the noise halos should not be visible. This discussion remains purely qualitative and an exhaustive evaluation of the algorithm is provided in Sections 3.4 and 3.5, but it is sufficient to justify the search for another solution.

We propose to define non-parametric noise containment maps by comparing the band-pass images $\mathbf{d}_t^u$ and $\mathbf{d}_t^{\hat{y}}$ that encode the observed image and its restored version, respectively. In this section, we present the reasons that support this choice, whereas Section 3.3.2 formally defines these noise containment maps. The MS coefficients of the observed image $\mathbf{u}$ have both signal and noise components, while those of the encoded restored image $\hat{\mathbf{y}}$ contain only signal information. Therefore, the comparison between the two sets of coefficients can help detecting where the signal is predominant, and provides a way to control the strength of the enhancement.

In order to quantify this aspect, an energy measure is associated with the band-pass images. Given a detail image $\mathbf{d}_t$, only a few coefficients are significantly not null, i.e. the signal is sparse in the MS decomposition. This has led many researchers to use statistical models to describe detail coefficients. Please refer to Loza et al. (2010) and references therein for further information about statistical models associated with MS coefficients. The following discussion simply assumes that the coefficients can be described with a Laplacian distribution with zero mean and scale parameter β. Given the maximum likelihood estimation of β of a zero mean Laplacian distribution, the energy $\Xi(\mathbf{d}_t)$ of the detail level $\mathbf{d}_t$ distribution is:

$$\Xi(\mathbf{d}_t) = 2\left(\frac{1}{M}\sum_{i=1}^{N}|\mathbf{d}_t(z_i)|\right)^2 \tag{3.9}$$

where M is the number of detail coefficients at scale t and z_i are the coordinates of the coefficient i. The amount of lost energy at a given scale t is measured as follows:

$$\delta_t = 1 - \frac{\Xi\left(\mathbf{d}_t^{\hat{y}}\right)}{\Xi\left(\mathbf{d}_t^u\right)}. \tag{3.10}$$

The effect of the filter across the scales is studied in two ROIs: one in the right lung (Figure 3.8a) and the other in the lumbar spine (Figure 3.9a), which are taken as examples of a textured low absorption region and a strong absorption one, respectively. Figure 3.10 reports the δ_t values associated with these two ROIs and some important aspects can be deduced from the trends of the resulting functions. First, we remark that $\lim_{t\to\infty}\delta_t = 0$, i.e. the band-pass images $\mathbf{d}_t^u$ and $\mathbf{d}_t^{\hat{y}}$ are more and more similar at lower frequencies. This is due to the filter that, by removing the high frequency component of noise, has a greater influence at fine scales than at coarse ones. Moreover, the iterative smoothing filter used in the MS decomposition (Equation 2.20) progressively reduces the noise strength. Nevertheless, if we consider the values $\delta_t(ROI_r)$, $r = 1, 2$ that are associated with two different anatomical

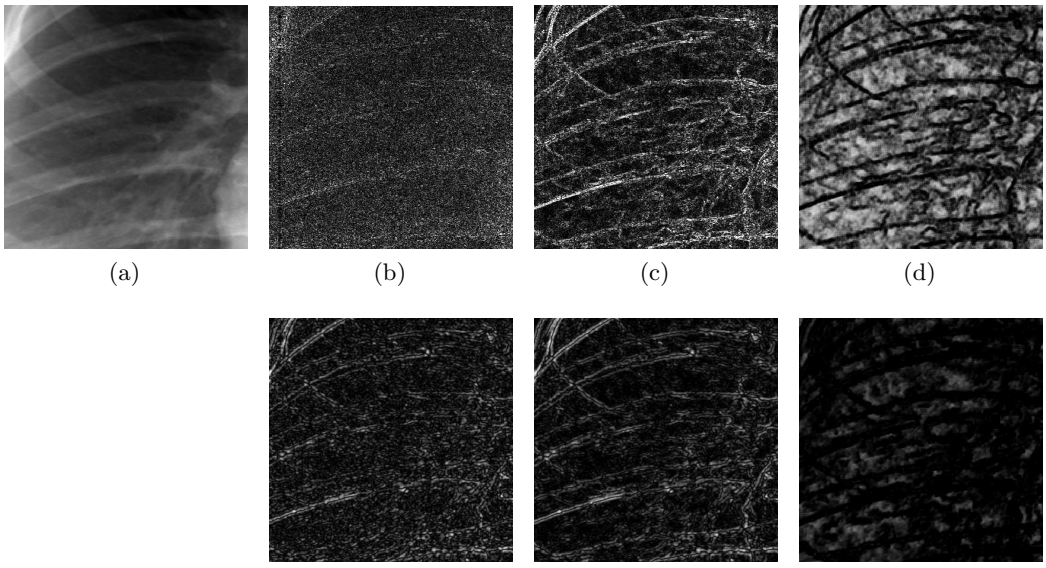


Figure 3.8: *Estimation of the LNC-maps from analysis of noisy and denoised images in the lung ROI: (a)* $\mathbf{u}$ *manually windowed to increase the visibility; (b)* $\mathbf{d}_0^u$*; (c)* $\mathbf{d}_0^{\hat{y}}$*; (d)* $\mathbf{n}_0 \to [0,1]$*; (e)* $\mathbf{d}_1^u$*; (f)* $\mathbf{d}_1^{\hat{y}}$*; (g)* $\mathbf{n}_1 \to [0,1]$*.*

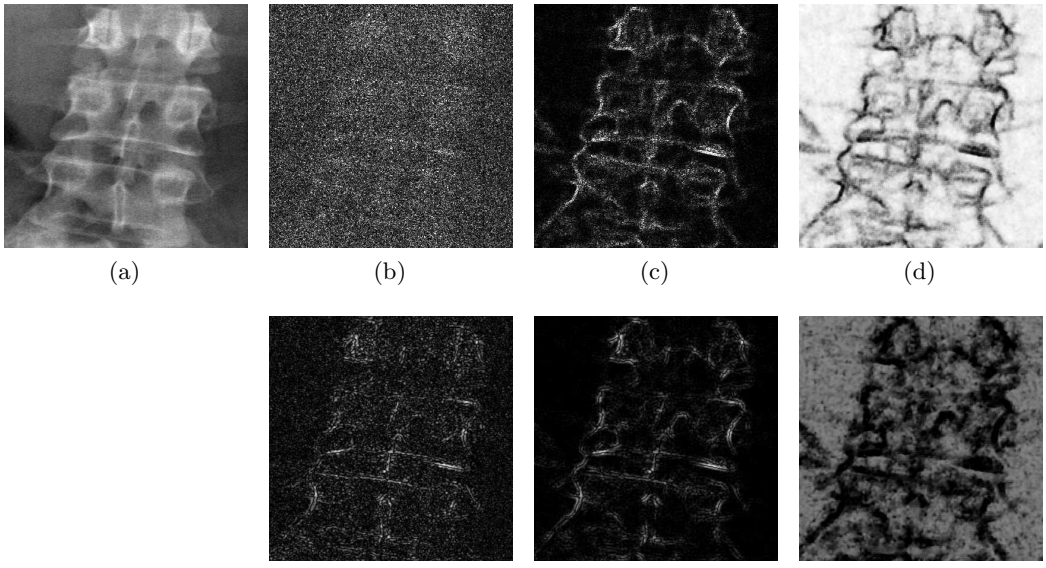


Figure 3.9: *Estimation of the LNC-maps from analysis of noisy and denoised images in the spine ROI: (a)* $\mathbf{u}$ *manually windowed to increase the visibility; (b)* $\mathbf{d}_0^u$*; (c)* $\mathbf{d}_0^{\hat{y}}$*; (d)* $\mathbf{n}_0 \to [0,1]$*; (e)* $\mathbf{d}_1^u$*; (f)* $\mathbf{d}_1^{\hat{y}}$*; (g)* $\mathbf{n}_1 \to [0,1]$*.*

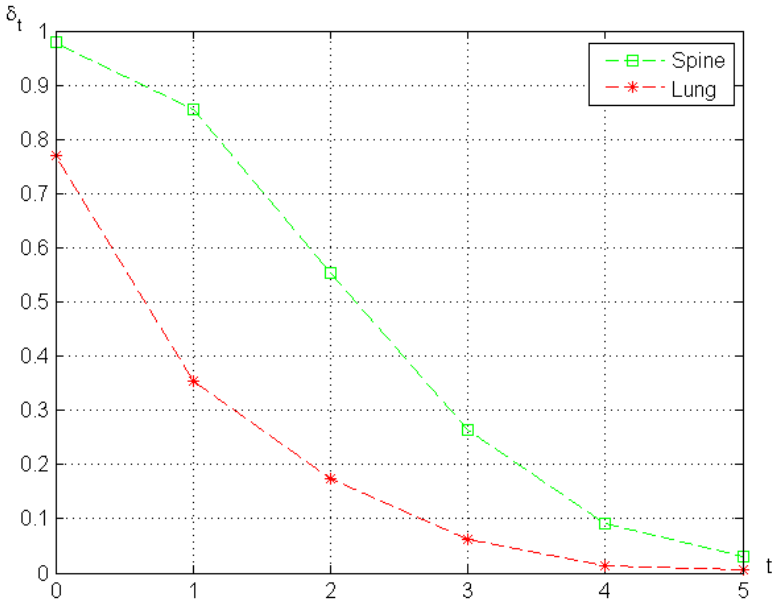


Figure 3.10: *Plots of δ_t at different scales t: red stars = lung ROI; green squares = spine ROI.*

regions ROI_j such that $\sigma(ROI_1) < \sigma(ROI_2)$, i.e. the X-ray absorption is higher in ROI_2 than in ROI_1 the following relation is valid:

$$\lim_{t\to\infty} \frac{\delta_t(ROI_1)}{\delta_t(ROI_2)} = 0, \quad \sigma(ROI_1) < \sigma(ROI_2). \tag{3.11}$$

Therefore, the decay of δ_t changes according to the ROI and is faster in low absorption regions where the noise level is lower. This implies that, for example, the noise needs to be more contained in the lumbar spine than in the lungs. As a consequence, the comparison between noisy and denoised detail coefficients quantifies the influence of the noise across the scales depending on region properties while avoiding relying on prior assumptions and user-defined parameters. Moreover, Figure 3.10 shows that the relative loss of energy is higher in strong absorption regions and, hence, the noise will be automatically contained in a different way depending on the region and patient morphotype.

This analysis is confirmed by comparing the band-pass images that encode the regions chosen as examples. Figure 3.8b shows the detail coefficients $\mathbf{d}_0^u$ in the lung region and it is very difficult to distinguish structures of interest from the noise. Hence, parametric noise containment maps are sub-optimal as the magnitude of the coefficients is not a good descriptor for structures in a noisy environment. On the other hand, the ribs and fine lung structures are much easier to see in the band-pass image $\mathbf{d}_0^{\hat{y}}$ (Figure 3.8c) and this is because of the capability of the XNL-means filter in restoring redundant information in the image by means of the patch representation. As a consequence, the XNL-means filter can be used to extract information merged with noise and, then, define consistent noise containment maps (Figure 3.8d). Passing to the coarser frequency level, the structures of interest appear both on $\mathbf{d}_1^u$ and $\mathbf{d}_1^{\hat{y}}$ (Figures 3.8e and 3.8f, respectively) as the noise becomes

practically negligible at this scale. This trend is well captured by the associated noise containment map $\mathbf{n}_1$ (Figure 3.8g) that has values very close to zero.

Similar observations come from the analysis of the region in the lumbar spine. However, the noise containment map $\mathbf{n}_0$ (Figure 3.9d) limits more significantly the coefficient boosting and is more regular compared to the one in the lung ROI (Figure 3.8d), because the noise level is higher and the details, i.e. the borders and internal structures of the vertebrae, are at coarser levels than fine pulmonary structures. Moreover, there is still a noticeable difference between $\mathbf{d}_1^u$ and $\mathbf{d}_1^{\hat{y}}$ (Figures 3.9e and 3.9f, respectively) given the high noise level, and the noise containment map $\mathbf{n}_1$ (Figure 3.9g) captures this difference.

Finally, the comparison of the encoded noisy and denoised images allows extracting important image features, and the noise containment maps are coherently estimated and allow optimizing the compromise between detail and noise enhancement. The major advantage of the proposed technique of analysis lies in its independence from user-defined parameters. This allows automatically taking into account the intra-patient and inter-patient variability, i.e. the changes in densities among tissues that are in the field of view and among patient size, respectively. Hereafter, we will call $\mathbf{n}_t$ local noise containment maps (LNC-maps), as defined according to the amount of redundant information in a neighborhood of coefficients. In the next section, we describe the steps to estimate the LNC-maps from the detail levels $\mathbf{d}_t^u$ and $\mathbf{d}_t^{\hat{y}}$.

3.3.2 Definition of local noise containment maps

The LNC-maps $\mathbf{n}_t$ are computed by comparing measures of local contrast extracted from $\mathbf{d}_t^u$ and $\mathbf{d}_t^{\hat{y}}$. These measures are images $\mathbf{c}_t$ where each pixel x_i is associated with a value equal to the average magnitude of coefficients in a patch of size $|P|$, i.e. the same patch size as the one used in the XNL-means filter, centered at the pixel x_i. The average within a small patch is more robust that coefficient-wise magnitude value (Li et al., 2005) and then is used here. Formally, the LNC-map $\mathbf{n}_t$ is defined as follows:

$$\mathbf{n}_t = \delta_t \left(1 - \frac{2\mathbf{c}_t^u \mathbf{c}_t^{\hat{y}} + \epsilon}{(\mathbf{c}_t^u)^2 + \left(\mathbf{c}_t^{\hat{y}}\right)^2 + \epsilon} \right) \tag{3.12}$$

where $\mathbf{c}_t^u$ and $\mathbf{c}_t^{\hat{y}}$ are the contrast measures associated with $\mathbf{d}_t^u$ and $\mathbf{d}_t^{\hat{y}}$, respectively, ϵ is a scalar factor used to avoid singularities, in this case fixed to 1, and δ_t is the energy factor computed by using Equation 3.10. Note that $\mathbf{n}_t$ is computed as the product of a local term that measures the correlation between local contrast measures, and a global one. The local term represents the probability of a coefficient being pure noise. The probability is low when the contrast values are similar because the XNL-means filter detects a structure of interest and, hence, preserves the most of its energy. The global term δ_t quantifies the amount of redundancy in the image and it is useful, for example to automatically take into account the difference between a small and a big morphotype. Indeed, a high amount of redundancy implies low amount of noise as pointed in Section 3.3.1 by means of the comparison between the lung ROI, highly redundant and at low noise level, and the lumbar spine ROI, not very redundant and at high noise level. This idea has been confirmed by gathering

the feedback from experts who concluded that, by defining the LNC-map using only the local term, the outcomes presented a slight loss of resolution. Finally, it is worth noting that the LNC-maps do not depend on user-defined parameters and, hence, the method does not require to manually set parameters according to the properties of the imaged patient or acquisition conditions.

3.3.3 Proposed boosting technique

The enhancement of contrast depends on how the detail coefficients are modified according to the activity maps $\mathbf{a}_t$. We propose to use activity maps defined as follows:

$$\mathbf{a}_t = \frac{\mathbf{c}_t^{\hat{y}}}{\mathbf{l}_t^{\hat{y}} + \tilde{\epsilon}} \tag{3.13}$$

where $\mathbf{c}_t^{\hat{y}}$ is the contrast measure associated with $\mathbf{d}_t^{\hat{y}}$ (Section 3.3.2), $\mathbf{l}_t^{\hat{y}}$ is the low frequency residual at scale t and $\tilde{\epsilon}$ is a scalar value that is used to avoid singularities and dependency from very low coefficient magnitudes, fixed in our tests to 100 by taking into account that the images are coded on 16 bits. This definition of activity is similar to Peli's (Peli, 1990) contrast that allows for a better description of the activity in complex images than coefficient magnitudes, because it is a measure of deviation compared to the mean signal and, therefore, is normalized. However, the contrast is computed within a small local window and not by considering coefficient-wise magnitude. This approach is more robust and, besides, Li et al. (2005) have shown that a smooth gain is preferable because it provides a good enhancement while limiting halo artifacts.

Note that $\mathbf{a}_t$ is defined as a function of $\mathbf{c}_t^{\hat{y}}$ and not of $\mathbf{c}_t^u$ and to justify this choice Figure 3.11 compares activity maps obtained by using $\mathbf{c}_t^u$ ($\mathbf{a}_t^u$) and $\mathbf{c}_t^{\hat{y}}$ ($\mathbf{a}_t^{\hat{y}}$) at different scales. While the image $\mathbf{a}_0^u$ (Figure 3.11b) associates with noise and image features very similar activity values, $\mathbf{a}_0^{\hat{y}}$ (Figure 3.11c) identifies very well the anatomical structures. The latter is then preferred as it allows properly estimating boosting values by distinguishing low and high activity features. This is similar to the idea of using $\hat{\mathbf{y}}$ as an oracle in order to more consistently quantify the distance between patches (Lebrun et al., 2012). Note that the difference between activity maps $\mathbf{a}_t^u$ and $\mathbf{a}_t^{\hat{y}}$ decreases at coarser scales as shown by comparing $\mathbf{a}_1^u$ (Figure 3.11d) with $\mathbf{a}_1^{\hat{y}}$ (Figure 3.11e).

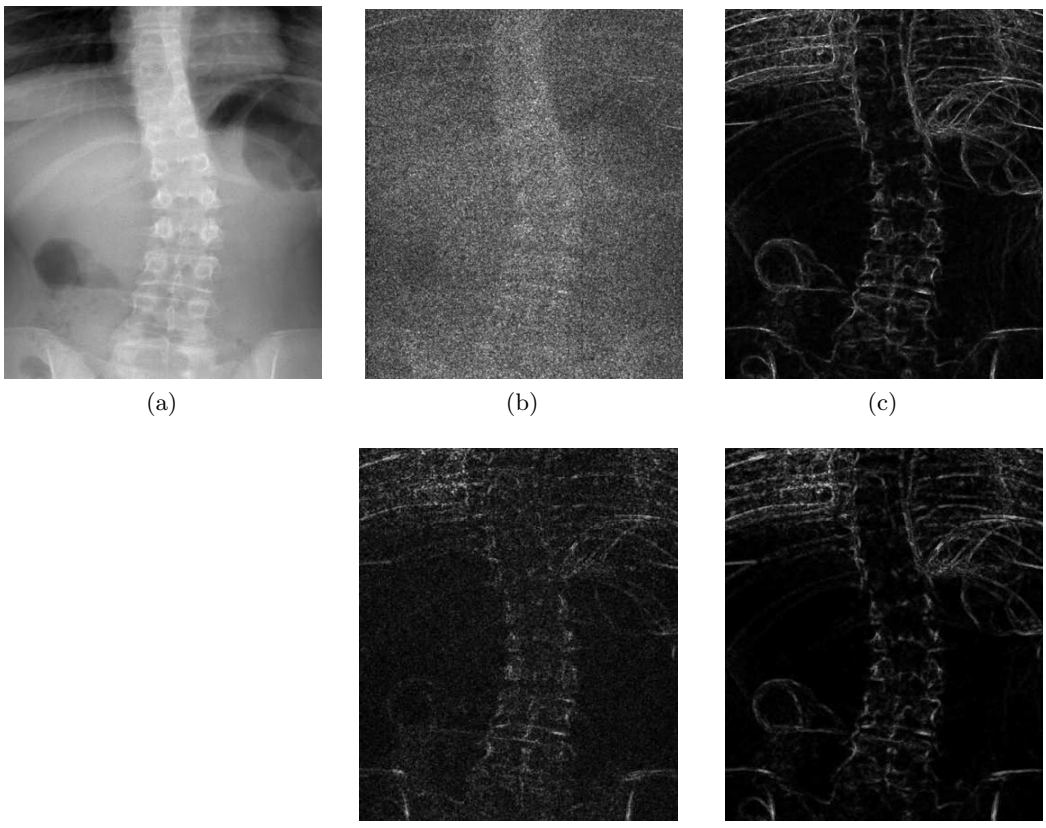
Given the activity maps associated with each scale, the enhanced band-pass images are obtained by using Equation 2.24 that we recall here:

$$\xi_t(\mathbf{d}_t) = \begin{cases} \mathbf{d}_t \left[(\gamma_t - 1) \left(1 - \frac{\mathbf{a}_t}{\bar{a}_t} \right)^2 + 1 \right] & \text{if} \quad \mathbf{a}_t < \bar{a}_t \\ \mathbf{d}_t & \text{otherwise} \end{cases} \tag{3.14}$$

where γ_t is the maximal gain at scale t and $\bar{a}_t$ is the activity measure threshold beyond which the band-pass $\mathbf{d}_t$ should not be modified. This is also called cut off value and is defined defined as follows:

$$\bar{a}_t = \min\left(\alpha \mu(\mathbf{a}_t), \max(\mathbf{a}_t)\right) \tag{3.15}$$

where $\mu(\mathbf{a}_t)$ is the average activity at scale t, and α is a constant value that avoids taking into account aberrant activity values and is set to 25 in all our tests. For



(a) (b) (c)

(d) (e)

Figure 3.11: *Comparison between activity maps defined as a function of noisy or denoised band-pass images at different scales: (a)* $\mathbf{u}$*; (b)* $\mathbf{a}_0^u$*; (c)* $\mathbf{a}_0^{\hat{y}}$*; (d)* $\mathbf{a}_1^u$*; (e)* $\mathbf{a}_1^{\hat{y}}$*. The activity maps* $\mathbf{a}_t^{\hat{y}}$ *are better than* $\mathbf{a}_t^u$ *because the low and high activity regions can be better distinguished, but the coarser the scale the more similar* $\mathbf{a}_t^{\hat{y}}$ *and* $\mathbf{a}_t^u$ *are.*

higher values of the latter parameter some halo artifacts may appear if strong edges are present in the image. On the other hand, for lower values of α the contrast may not be sufficiently increased.

As for the maximal gain parameters, they are initially set by linearly ranging their values from 6 at the finest scale to 3 at the coarsest one. Empirically, these are ideal parameters when the amount of noise is very low. However, in the context of this thesis, this case rarely occurs. The actual gain values are then semi-automatically computed as follows:

$$\gamma_t = g_t - (g_t - g_{k-1})\mu(\mathbf{n}_t) \tag{3.16}$$

where g_t are the initial ideal gain values, k is the number of scales (here $k = 6$) and $\mu(\mathbf{n}_t)$ is the average value of the LNC-map at t. The initial gain values tend to be preserved if the noise is absent at a given scale, i.e. $\mu(\mathbf{n}_t) \simeq 0$. On the contrary, when $\mu(\mathbf{n}_t) \simeq 1$, γ_t is set equal to the gain at the coarsest scale, i.e. the lowest one. The initial values g_t are defined according to the property of full-body X-ray images and they should probably be changed to process other types of images. In this way, the problem of defining the maximal gain parameters is simplified. Indeed, only

one set needs to be fixed for all the images and, then, the values are automatically adjusted according to the image content.

If the band-pass images $\mathbf{d}_t^{\hat{y}}$ are used in Equation 3.14, then the fully enhanced denoised image is computed by applying the formula in Equation 2.22. We will refer to this method with the abbreviation DE, which stands for denoised enhancement. If the band-pass images $\mathbf{d}_t^u$ are used in Equation 3.14, then two types of result can be obtained. A first option consists in fully enhancing the observed image $\mathbf{u}$ by applying the formula in Equation 2.22 and $\mathbf{a}_t = \mathbf{a}_t^u$ as in this case the XNL-means is not applied. We will refer to this method with the abbreviation NE, which stands for noisy enhancement. A more interesting result can be obtained by relying on a noise containment-based approach with $\mathbf{n}_t$ defined in Equation 3.12. We will refer to this method with the abbreviation LNCE, which stands for local noise containment map enhancement.

3.3.4 Overview of the proposed framework

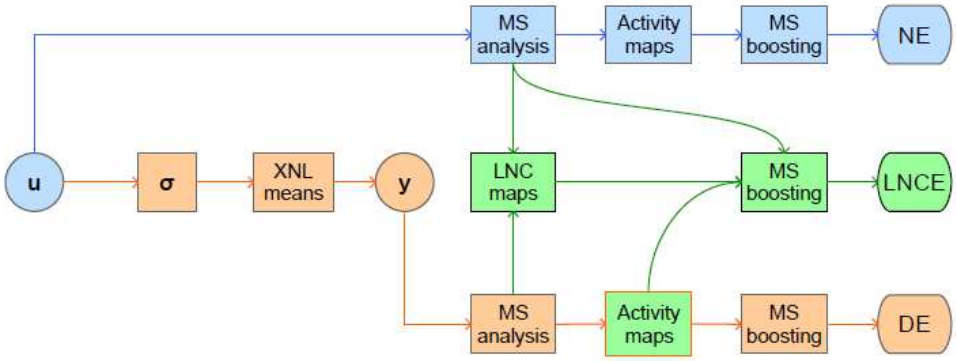


Figure 3.12: *Contrast enhancement framework.*

The flowchart in Figure 3.12 summarizes the main steps of the proposed framework that offers three different solutions.

The observed image is decomposed into the band-pass images $\mathbf{d}_t^u$ and low frequency residual $\mathbf{l}_k^u$. Then, the image is enhanced by boosting the detail coefficients according to the activity maps $\mathbf{a}_t^u$ computed by using Equation 3.13, but as a function of $\mathbf{c}_t^u$ rather than $\mathbf{c}_t^{\hat{y}}$. This sequence of steps provides the output NE. This is computationally and conceptually the easiest approach, but it does not take into account the presence of noise.

In order to overcome this drawback, the image $\hat{\mathbf{y}}$ is obtained by using the XNL-means filter (Section 3.2.1) according to the noise levels that are estimated from the input $\mathbf{u}$ by using the percentile method (Section 3.1.3). The image $\hat{\mathbf{y}}$ is then decomposed into the band-passes images $\mathbf{d}_t^{\hat{y}}$ and low frequency residual $\mathbf{l}_k^{\hat{y}}$. Then, the output with DE is obtained by exploiting the activity maps associated with the detail coefficients $\mathbf{d}_t^{\hat{y}}$. This solution directly addresses the presence of noise in the image. It is more complex as it requires to estimate the noise levels and to filter the image, which is why in a not optimized MATLAB code it takes about 2 minutes to get the solution for a full-body image that contains 14 millions of pixels. However, since the XNL-means filter is coded using integral images, an estimation

is that the proposed framework could be efficiently implemented in a C++ oriented environment and using parallel programming.

Finally, a noise containment-based solution, the LNCE, can be formulated at the price of decomposing both the observed image $\mathbf{u}$ and the filtered one $\hat{\mathbf{y}}$. The increase in computational time is negligible with respect to the algorithm DE because a full-body image can be decomposed in about 2 seconds. Both the blue and orange branches in Figure 3.12 are executed up to the MS analysis and, then, the two sets of band-pass images are used to get the LNC-maps (Section 3.3.2). Finally, the contribution of the fully enhanced band-pass images $\xi_t(\mathbf{d}_t^u)$ are weighted with the input ones according to the estimated signal contribution at each coefficient.

In conclusion, the proposed framework can accommodate different clinical needs by selecting different inputs and enabling or not the noise containment. Many relevant aspects for processing low dose X-ray images emerge from the comparison of these methods. Moreover, we add to the comparison an EOS proprietary algorithm that relies on the knowledge of the image processing chain to define the parameters of the noise containment maps. This algorithm is taken as example of parametric noise containment-based method. The details are omitted for confidentiality reasons and are not relevant for the analysis of the validation results. Hereafter, EOS enhancement is denoted by EOSE. This is the default option in clinical routine for processing EOS images, and, since one of our objectives is to certify the robustness to changes in the data, this algorithm is relevant to the validation. Consequently, the approach considered for the validation in Sections 3.4 and 3.5 is general as it does not focus on any specific clinical case study. The quality of the images post-processed with the aforementioned algorithms is evaluated in the following sections and we assume that the most performing solution is also the easiest one to adapt to specific instances.

3.4 Evaluation of the framework on phantom images

The standard PHD5000 fluoro phantom (Figure 2.8) is used to compare the four studied algorithms according to the image quality measures for a phantom introduced in Section 2.3. The set of cases tested on the phantom is built to take into account different signal strengths, and interposing between the X-ray tube and the phantom a polymethyl methacrylate (PMMA) block of various thicknesses. Table 3.1 lists the samples of the dataset, where the thickness of the PMMA block is indicated in cm and the signal strength represented by the entrance dose in μGy. The measures of image quality computed over the input images are also indicated. These tests simulate the acquisition conditions of typical clinical cases. For example, the samples: $\{10\ cm; 10\ \mu Gy\}$, $\{10\ cm; 71\ \mu Gy\}$ and $\{10\ cm; 215\ \mu Gy\}$ are acquired with, respectively, $\{60\ kV; 83\ mA; 4\ C\}$, $\{83\ kV; 200\ mA; 4\ C\}$ and $\{100\ kV; 280\ mA; 4\ C\}$. These parameters are used in the following exams: follow-up of the full spine, diagnostic full spine and pelvis. The samples at 10 cm, 15 cm, 20 cm, 25 cm and 30 cm correspond to the normal child, overweight child, normal adult, overweight adult and obese adult morphotypes, respectively.

Table 3.1: *Samples of the phantom dataset described by the thickness of the PMMA block in cm, the entrance dose in μGy and the SNR, CNR and DYN values computed over the input image.*

PMMA block thickness	Entrance dose	SNR	CNR	DYN
10 cm	10 μGy	47.25	1.35	49%
10 cm	67 μGy	116.47	4.29	53%
10 cm	71 μGy	126.02	4.23	53%
10 cm	75 μGy	110.73	4.24	53%
10 cm	215 μGy	190.71	7.47	52%
15 cm	65 μGy	108.51	2.44	40%
15 cm	87 μGy	115.30	3.01	41%
15 cm	89 μGy	116.84	2.98	42%
15 cm	100 μGy	123.53	2.80	42%
15 cm	102 μGy	132.75	2.95	41%
15 cm	136 μGy	165.43	3.78	41%
20 cm	16 μGy	37.63	0.53	30%
20 cm	109 μGy	117.90	1.96	33%
20 cm	454 μGy	180.01	2.59	33%
25 cm	109μGy	97.13	1.16	26%
25 cm	136 μGy	98.24	1.26	26%
30 cm	33 μGy	27.39	0.20	19%
30 cm	566 μGy	128.46	1.14	22%

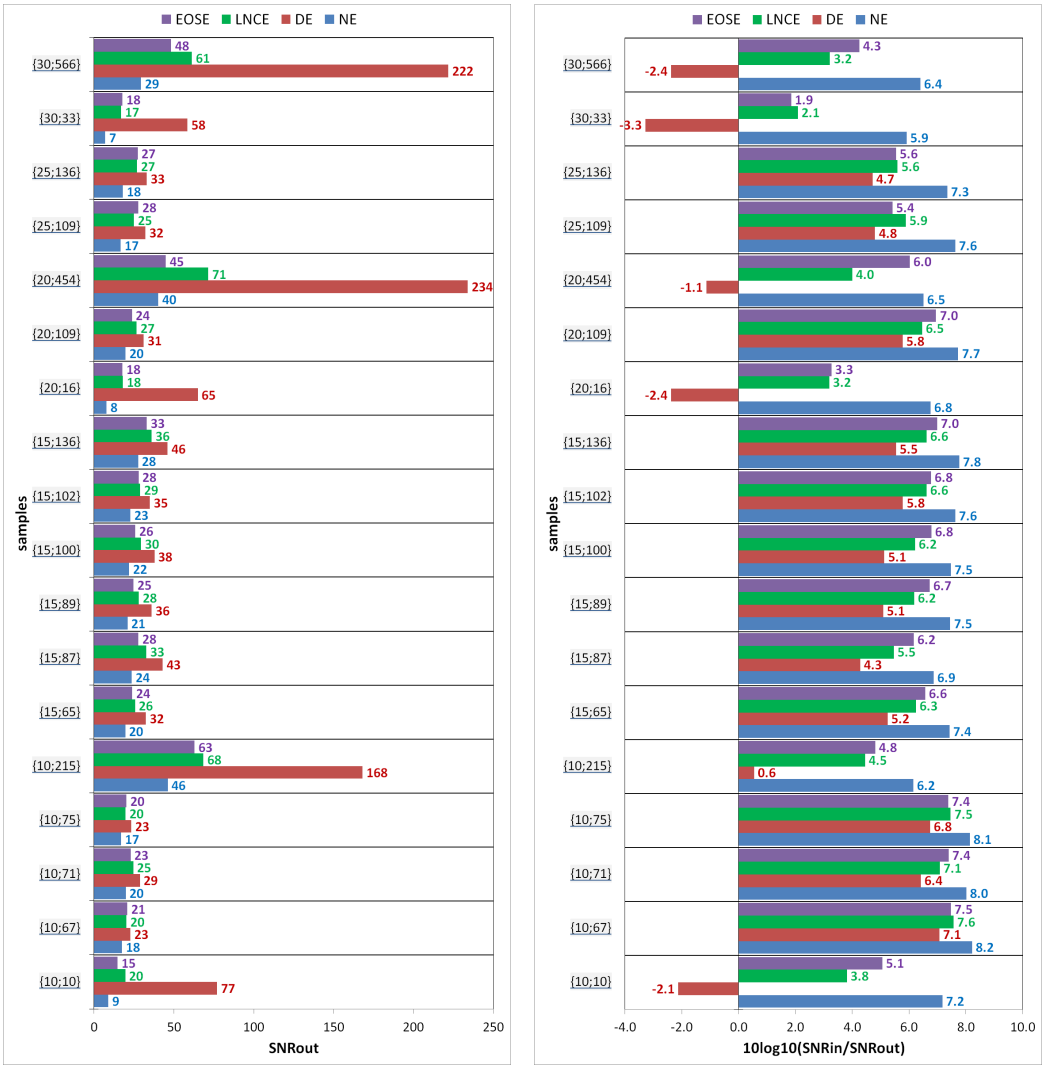
3.4.1 Signal to noise ratio

The SNR is defined in Equation 2.26 and is related to the amount of signal that reaches the detector: Table 3.1 shows that the lower PMMA thickness and stronger input signal (e.g. $\{10cm, 215\mu Gy\}$) the higher the SNR is.

Figures 3.13a and 3.13b show the output SNR values, SNR_e, computed over the enhanced images **e** obtained with NE, DE, LNCE or EOSE and the corresponding decrease of SNR compared with the input **u**, $10\log_{10}(SNR_u/SNR_e)$, respectively.

Figure 3.13a shows that DE performs better than other methods in all the cases, while NE is always the worst option, which is logical considering the definition of the algorithms. When compared to noise containment-based methods, DE is significantly better in two cases: when the ratio signal on thickness to cross is very low and very high. In the first case (e.g. $\{20\ cm, 16\ \mu Gy\}$), the noise level is very high and, hence, significantly present in image **u**. Therefore, it is not sufficient to contain the noise, but it should be reduced on the input image. In the opposite case (e.g. $\{10\ cm, 215\ \mu Gy\}$), the noise level is very low and noise containment-based methods tend to almost fully enhance the image. However, note that in this case LNCE performs better than EOSE ($\{20\ cm, 454\ \mu Gy\}$), while in general they have similar performances in terms of SNR.

Figure 3.13b shows that in most of the cases the SNR in the enhanced image is lower than the initial one, except for the previously described samples. Summarizing the results, NE, DE, LNCE and EOSE cause an average SNR reduction equal to



(a) (b)

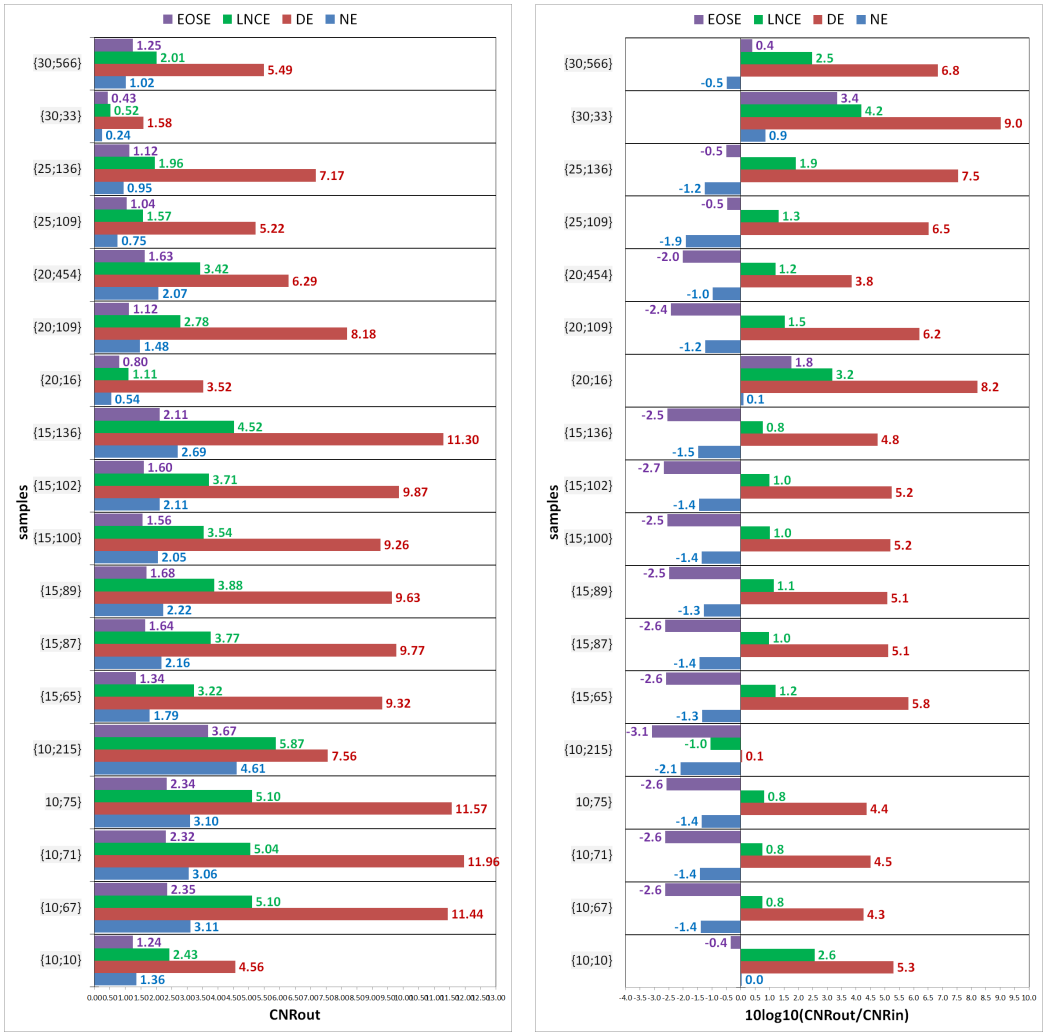
Figure 3.13: *Phantom image quality evaluated according to the SNR measure: (a)* SNR_e *where* **e** *obtained by applying the NE, DE, LNCE or EOSE algorithms; (b)* $10\log_{10}(SNR_u/SNR_e)$ *corresponding reduction of SNR compared with the input* **u**.

81.0%, 28.9%, 69.5% and 72.3%, respectively. Finally, DE is the best option in terms of SNR, which is logical because it fully enhances an image that has been filtered, while the other solutions only contain the noise (LNCE and EOSE) or simply do not take it into account (NE). However, since the SNR is measured in a void region, it gives no information on how regions of interest are processed.

3.4.2 Contrast to noise ratio

The CNR is defined in Equation 2.27 and it takes into account both contrast and noise level. Table 3.1 shows that, for a given PMMA thickness, the CNR increases with the entrance dose. However, the highest achievable CNR value depends on the

thickness: at 10 cm, 10 μGy are sufficient to go beyond a CNR level equal to 1[3], while, at 25 cm, the dose has to be increased up to 109 μGy.



(a) (b)

Figure 3.14: *Phantom image quality evaluated according to the CNR measure: (a) CNR_e where* **e** *obtained by applying the NE, DE, LNCE or EOSE algorithms; (b) $10\log_{10}(CNR_e/CNR_u)$ corresponding increase of CNR compared with the input* **u**.

Figures 3.14a and 3.14b show the output CNR values, CNR_e, computed over the enhanced images **e** obtained with NE, DE, LNCE or EOSE and the corresponding increase of CNR compared with the input **u**, $10\log_{10}(CNR_e/CNR_u)$, respectively.

Figure 3.14a shows that NE and EOSE have similar performances in terms of CNR, which implicitly indicates a higher increase of contrast with the proposed boosting technique (Section 3.3.3) as the results in terms of SNR suggest lower noise on images post-processed with EOSE. The proposed non-parametric noise containment-based method LNCE doubles the contrast with respect to the classical

[3]A disk of a phantom is empirically considered visible if the CNR is higher than 1.

parametric definition of noise containment maps. Nevertheless, DE clearly assures the best quality in terms of CNR, which is the logical consequence of the results on the SNR.

In Figure 3.14b positive values indicate a good balance between contrast and noise, while negative ones mean overshooting, i.e. the increase of contrast comes at the price of an exaggerated increase of noise level. According to this definition, it can be noted that the algorithms NE and EOSE present overshooting. While this is predictable with NE, because the noise is not contained at all, it is less expected for EOSE. In particular, EOSE avoids over-shooting in low signal/thick PMMA block instances (e.g. $\{20\ cm, 16\ \mu Gy\}$), but not in the other cases. This problem is common with methods based on parametric noise containment maps as they do not provide sufficient flexibility. In fact, LNCE clearly overcomes this drawback, because, except for one case ($\{10\ cm, 215\ \mu Gy\}$), there is no over-shooting. To sum up the results, NE, DE, LNCE and EOSE produce an average CNR improvement respectively equal to -20.9%, 281.9%, 43.7% and -21.5%. The fact that EOSE shows the least favorable average CNR score is not related to how the noise is contained, as we have shown in the previous section, but is due to the advantages of the proposed boosting technique (Section 3.3.3). The CNR measures on the disk are representative of how well medium/low frequency details can be enhanced. However, the analysis needs to be completed by studying objects at higher frequencies, which is performed in Section 3.4.4.

3.4.3 Dynamic

The DYN is defined in Equation 2.28 and it represents how well the gray level range is exploited. Table 3.1 shows that the dynamic depends on the thickness, whereas it practically does not depend on the amount of dose.

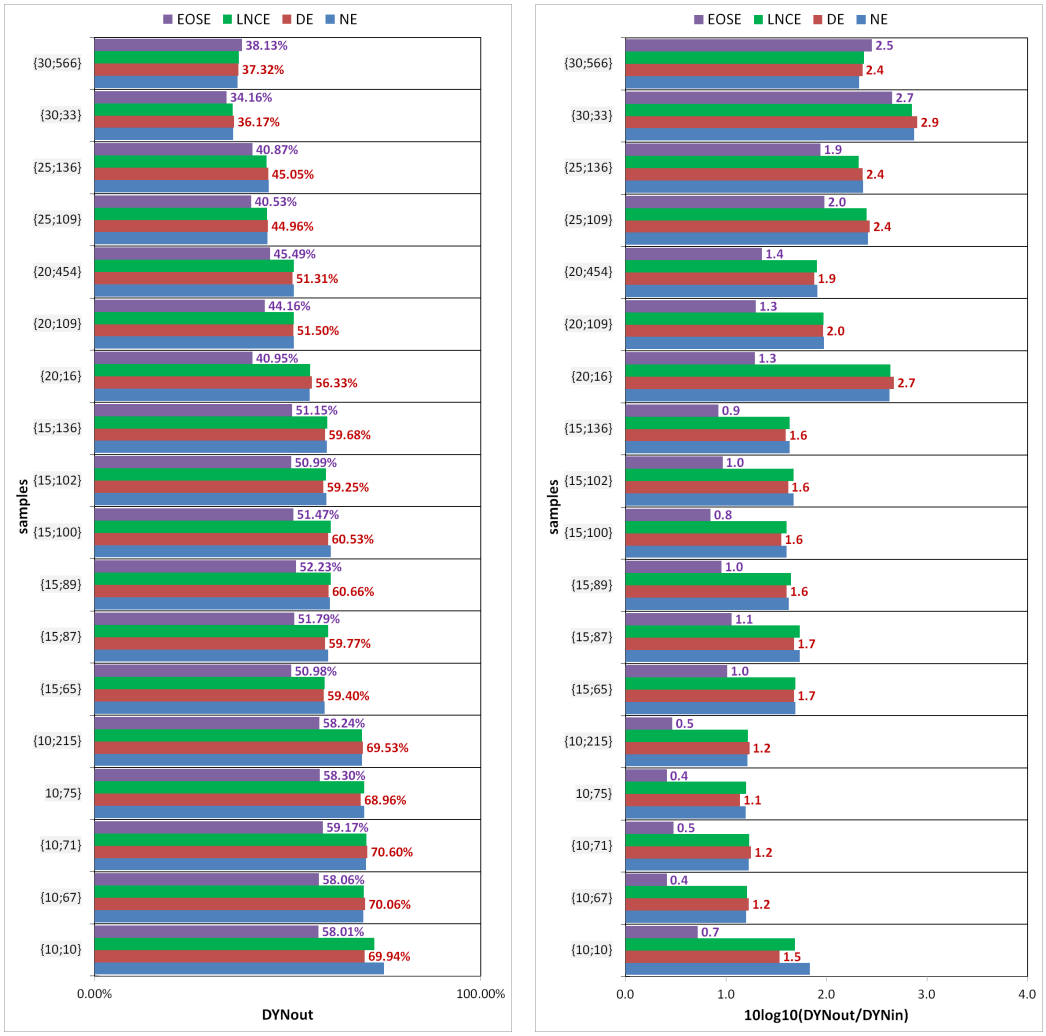
Figures 3.15a and 3.15b show the output DYN values, DYN_e, computed over the enhanced images $\mathbf{e}$ obtained with NE, DE, LNCE or EOSE and the corresponding increase of DYN compared with the input $\mathbf{u}$, $10\log_{10}(DYN_e/DYN_u)$, respectively.

Figure 3.15a shows that NE, DE and LNCE have very similar scores, because they use the same boosting technique (Section 3.3.3), and they perform better than EOSE. However, the degree of improvement depends on the absorption: $+11\%$, $+8\%$, $+6\%$, $+4\%$ and $+0\%$ at 10 cm, 15 cm, 20 cm, 25 cm and 30 cm, respectively. Therefore, the proposed boosting technique allows better exploiting the gray level range on children and adult normal size patients, whereas no sensible improvement is registered on obese patients.

As for the relative increase of dynamics with respect to the input (Figure 3.15b) NE, DE, LNCE and EOSE entail an average DYN improvement respectively equal to 53.8%, 53.0%, 53.5% and 32.8%.

3.4.4 Spatial resolution

The high resolution grid at the center of the phantom (Figure 2.8) is used to evaluate the spatial resolution of a system. The resolution is measured in line pairs per millimeter (lp/mm) and associated with the finest set of visible lines on the grid. It depends then on the signal strength and PMMA block thickness, in addition to detector properties. This section verifies whether the algorithms from the frame-



(a) (b)

Figure 3.15: *Phantom image quality evaluated according to the DYN measure: (a) DYN_e where* **e** *obtained by applying the NE, DE, LNCE or EOSE algorithms; (b)* $10\log_{10}(DYN_e/DYN_u)$ *corresponding increase of DYN compared with the input* **u**.

work preserve the spatial resolution or not. The input image manually windowed in the region of the resolution grid is compared to DE and LNCE. Note that the study of NE is unnecessary because, since the image is fully enhanced, it is not possible to reduce the resolution, and in some our tests we have verified that EOSE does not cause a resolution loss. Figure 3.16 presents three tests behind the 10 *cm* PMMA block: the columns, from left to right, show the resolution grid from the input image with manual windowing, DE and LNCE; the lowest, the medium and highest signal strengths of the sample at 10 *cm* are considered, i.e. 10 μGy (Figures 3.16a, 3.16b and 3.16c), 71 μGy (Figures 3.16d, 3.16e and 3.16f) and 215 μGy (Figures 3.16g, 3.16h and 3.16i), respectively. The arrows point at the finest set of distinguishable lines. The images show that DE causes a spatial resolution loss, while LNCE does not. They also show that the grade of the decrease depends on

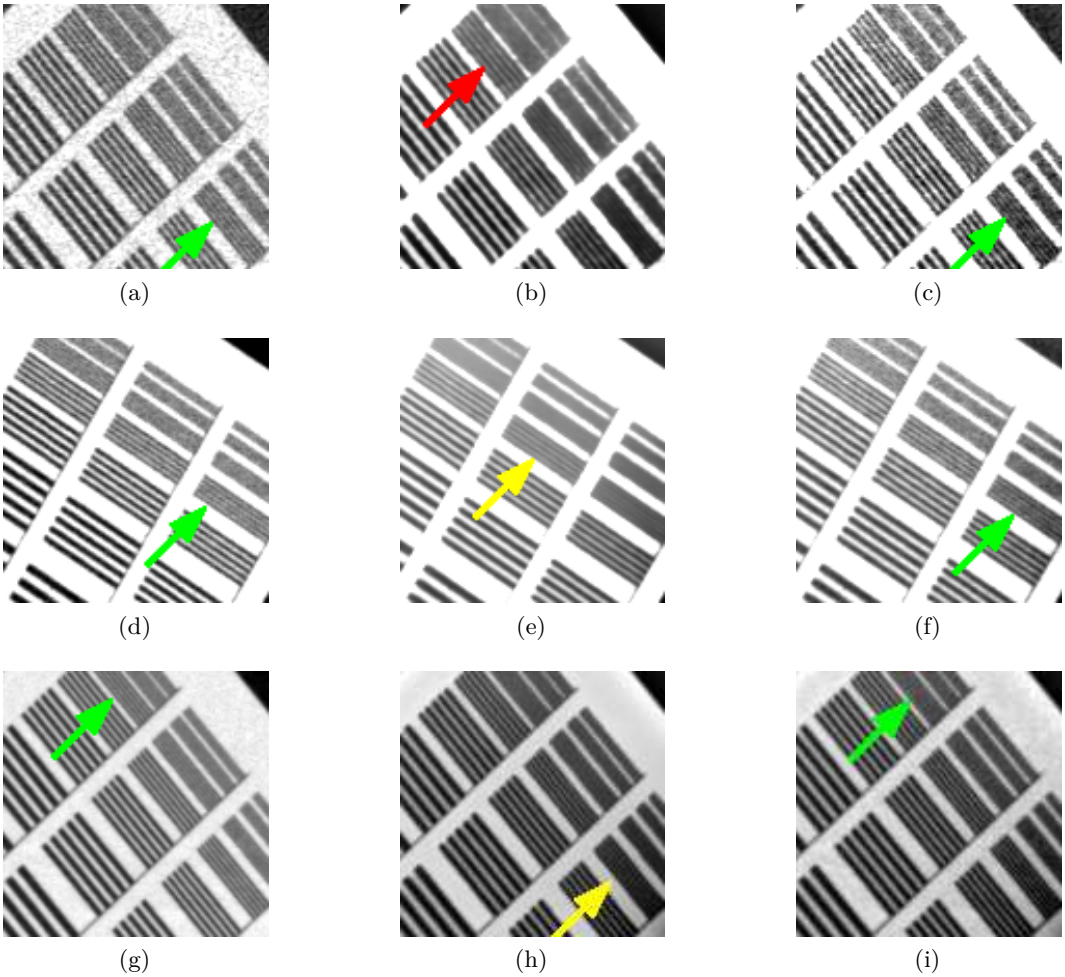


Figure 3.16: *Resolution grid of the phantom behind a* 10 *cm PMMA block. First row (a-c)* 10 μGy*; second row (d-f)* 71 μGy*; third row (g-i)* 215 μGy*. First column (a, d, g) input; second column (b, e, h) DE; third column (c, f, i) LNCE. The arrows indicates the finest set of visible lines: green = resolution preserved, yellow = resolution loss equal to* 0.2 *lp/mm, red = resolution loss equal to* 0.4 *lp/mm.*

the signal strength. Indeed, at 10 μGy the DE resolution is 1.4 lp/mm versus 1.8 lp/mm, at 71 μGy the DE resolution is 1.6 lp/mm versus 1.8 lp/mm and at 215 μGy 1.8 lp/mm versus 2.0 lp/mm. This is due to the XNL-means filter effectiveness in restoring the redundant information of the lines: the stronger the signal the easier is to capture the similarity between patches. Similar results are obtained by conducting the same tests with thicker PMMA blocks. For example behind a block of 20 *cm* the resolution visually assessed on input images is equal to 1.6 lp/mm, 1.8 lp/mm and 2.0 lp/mm at 16 μGy, 109 μGy and 454 μGy, respectively. The LNCE does not cause a loss of resolution, whereas with DE it is possible to see up to 1.4 lp/mm, 1.6 lp/mm and 1.8 lp/mm. In this case the degree of the loss is equal to 0.2 lp/mm regardless the signal strength, which implies that the XNL-means filter guarantees to preserve at least 1.4 lp/mm.

Finally, the results had so far indicated DE as the best option to process a low

contrast noisy image, but it emerges from the analysis in this section that it entails a resolution loss which could limit its exploitation for clinical diagnosis. Therefore, the design of the proposed framework (see Figure 3.12) becomes even more interesting as it is possible to easily switch to LNCE that does not present this drawback and, at the same time, offers a good compromise between detail enhancement and noise containment. In any case, both methods are relevant as justified in Section 3.5 that studies the application of the method to clinical images.

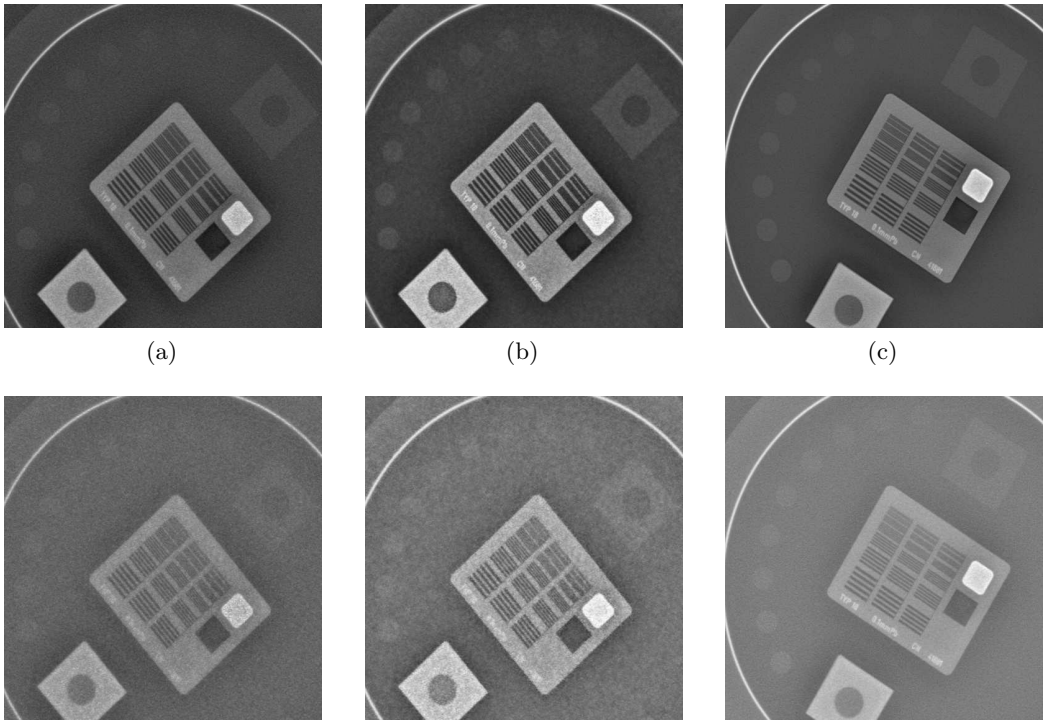
3.4.5 Gain in terms of dose reduction

Table 3.2: *Comparison of image quality as a function of the dose on a phantom acquired by simulating full spine pediatric (10 cm) and adult (20 cm) exams. The measures indicate the relative increase of performances in terms of SNR, CNR and DYN at diagnostic versus follow up dose. The diagnostic case is post-processed with EOSE and the follow-up is post-processed with EOSE and LNCE. By post-processing follow-up images with the proposed method the performances are almost identical (pediatric) and close (adult) to the diagnostic image quality.*

PMMA	$\frac{d_d}{d_f}$	$\frac{EOSE(d_d)}{EOSE(d_f)}$			$\frac{EOSE(d_d)}{LNCE(d_f)}$		
		SNR	CNR	DYN	SNR	CNR	DYN
10 *cm*	7.10	×1.56	×1.86	+1%	×1.17	×0.95	−13%
20 *cm*	6.62	×1.34	×1.40	+3%	×1.32	×1.01	−12%

The validation on phantoms allows also quantifying how much the dose can be reduced while preserving suitable image quality performances. This aspect is then analyzed by comparing the non-parametric (LNCE) and parametric (EOSE) noise containment-based methods. In detail, we measure how much the performances in terms of SNR, CNR and DYN change depending on the dose for pediatric (10 *cm*) and adult (20 *cm*) cases. Table 3.2 summarizes the results of the test: a phantom image acquired with diagnostic spine parameter settings (dose d_d, 71 μGy and 109 μGy for children and adult, respectively) is post-processed with EOSE and, then, compared with the same phantom image acquired with follow-up spine parameter settings (dose d_f, 10 μGy and 16 μGy for children and adult, respectively) post-processed with EOSE and LNCE. The relative increase of SNR, CNR and DYN values are computed from the measures in Figures 3.13a, 3.14a and 3.15a and has to be interpreted according to the increase of dose. The analysis is completed by a visual assessment (Figure 3.17).

As for the pediatric morphotype, Table 3.2 indicates that, by using EOSE, the SNR and the CNR are ×1.56 and ×1.86 higher on the corresponding diagnostic image while the DYN does not significantly change, which is already a good result considering that the dose is reduced by a factor 7.10. Nevertheless, LNCE allows obtaining almost a diagnostic image quality at dose d_f as the SNR only changes of a factor 1.17, and CNR and DYN even increases. The visual comparison (Figures 3.17a - 3.17c) supports these assessments. For example, the disks are slightly better visible on LNCE(d_f) than EOSE(d_d). Moreover, the resolution grid (e.g.) is certainly better contrasted which confirms the increase of DYN when using LNCE. Finally, despite the factor of dose reduction equal to 7.10, considering the results and that there is



(a) (b) (c)

(d) (e) (f)

Figure 3.17: *Visual assessment on image quality as a function of the dose: (a-c) Simulation on a phantom of full spine exam for pediatric morphotype; (d-f) Simulation on a phantom of full spine exam for adult morphotype. (a,d) Follow-up images post-processed with EOSE; (b-e) follow-up images post-processed with LNCE; (c-f) diagnostic images post-processed with EOSE.*

no difference in spatial resolution at 10 μGy and 71 μGy behind 10 cm PMMA (see Section 3.4.4), LNCE assures an image quality on follow-up exams that differs from the one required for diagnostic clinical goals only by a slight increase of noise level. Therefore, the proposed post-processing technique is quantitatively very interesting in the optic of dose reduction in pediatric exams.

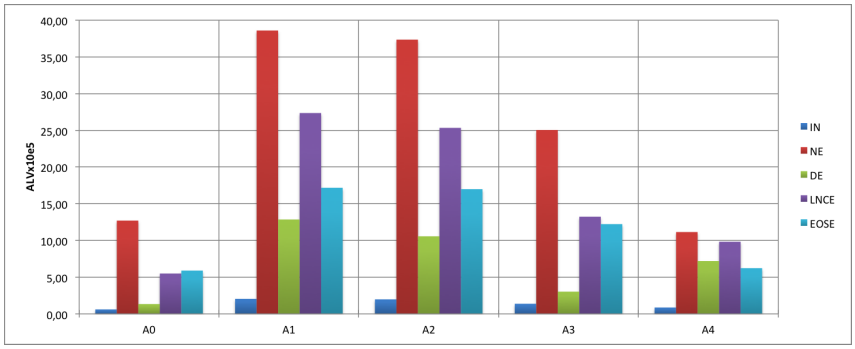
As for the adult morphotype, the assessments on CNR and DYN are still valid, whereas the performances in terms of SNR provided by LNCE and EOSE are practically identical. This is confirmed by the visual analysis of the image in Figure 3.17e that presents a non-negligible amount of residual noise. Therefore, on this morphotype, LNCE does not complete recover the image quality despite better performances in terms of CNR and DYN and, hence, a reduction factor lower than 6.62 should be investigated.

3.5 Evaluation of the framework on clinical images

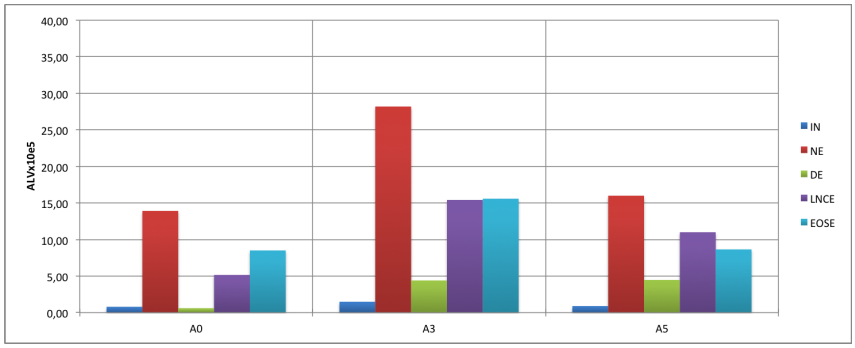
We compare the methods on clinical images by using the validation approaches introduced in Section 2.3.2. The clinical database consists of 130 images that have been randomly selected among anonymous data and it includes patients of different

ages and sizes as well as a wide range of type of exams (full spine, lower limbs, diagnostic, follow-up). This is the logical transposition of the tests conducted on the phantom to the clinical context.

3.5.1 Quantitative measures for clinical image quality evaluation



(a)



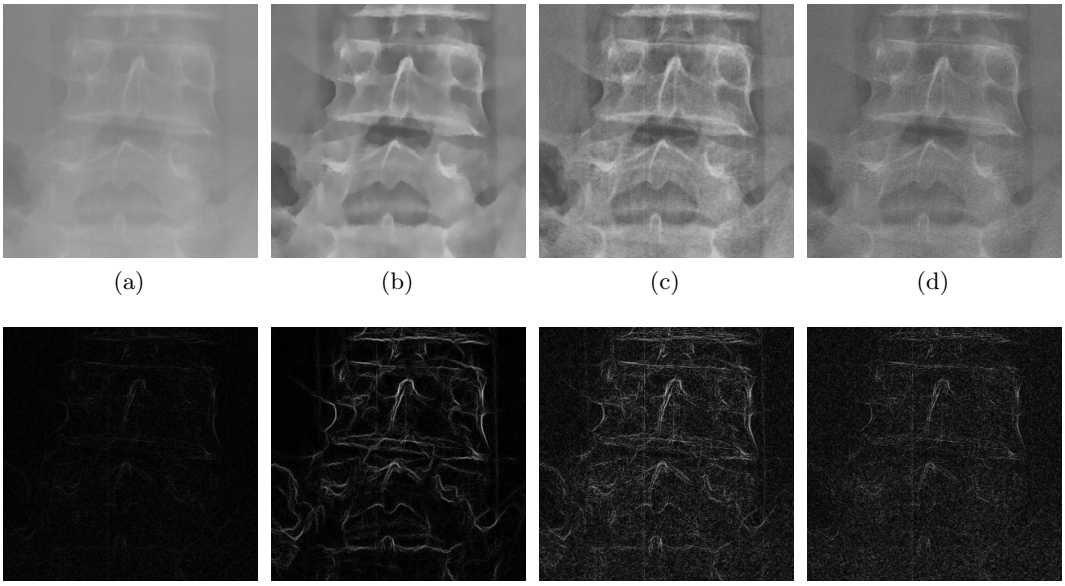
(b)

Figure 3.18: *ALV measures in anatomical ROIs computed from the images in (a) Figure 2.9a; (b) Figure 2.9c. For each of the anatomical ROI on the x-axis, the ALV measures are computed over IN, NE, DE, LNCE and EOSE.*

We analyze in detail the ALV measures (Figure 3.18) for the frontal acquisitions in Figure 2.9 and compare the performances of the studied post-processing algorithms according to the anatomical ROIs that are: void ($\mathcal{A}_0$), lumbar spine ($\mathcal{A}_1$), thoracic spine ($\mathcal{A}_2$), proximal femur ($\mathcal{A}_3$), lungs ($\mathcal{A}_4$) and knees ($\mathcal{A}_5$). The analysis is completed by visual comparison of the structures post-processed with different methods to assess the relative contributions in terms of signal and noise enhancement. Figures 3.19, 3.20, 3.21 and 3.22 show, in the first row, the visual rendering of the vertebrae L4 and L5, the proximal femur, the right lung and the knee, respectively, not enhanced and post-processed with DE, LNCE and EOSE, and, in the second row, the magnitude of the Sobel gradient computed over these images.

In the ROI $\mathcal{A}_0$, a significant increase of the noise with NE can be observed on both full spine (Figure 3.18a) and lower limbs (Figure 3.18b) images. Indeed, considering the full spine sample (Figure 3.18a) the ALV in region $\mathcal{A}_0$ increases with respect to the input by: $\times$20.84 (NE), $\times$2.20 (DE), $\times$9.02 (LNCE) and $\times$9.67 (EOSE). These results are coherent with the SNR obtained on phantom (Section 3.4.1): DE limits

the noise impact better than noise containment-based methods, whereas NE should be discarded because it excessively boosts the noise.



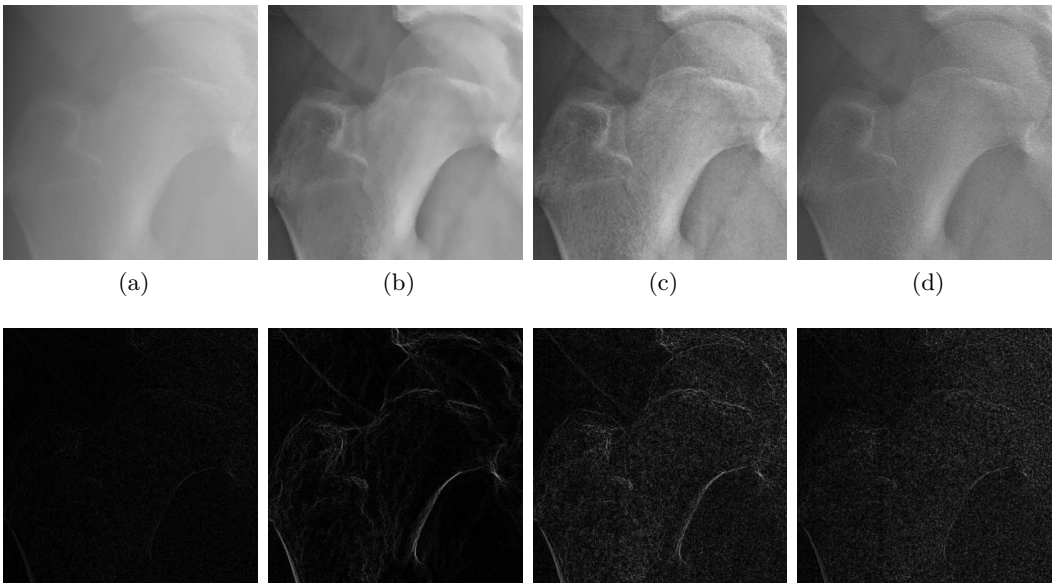
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.19: *Visual comparison of vertebrae L4 and L5 in $\mathcal{A}_1$: top row = images and bottom row = related Sobel gradient magnitudes; columns from left to right = IN, DE, LNCE and EOSE.*

In the lumbar spine (Figure 3.19), the ALV increases with respect to the input by: $\times$18.83 (NE), $\times$6.27 (DE), $\times$13.35 (LNCE) and $\times$8.38 (EOSE). The first observation is that the relative improvement of ALV given by NE is similar to the one revealed in region $\mathcal{A}_0$. Therefore, even if NE gets the highest score in $\mathcal{A}_1$, it is not reliable because it depends only on noise. On the contrary, the other algorithms do not show this trend. LNCE is better than DE and EOSE. Refining the input image demands extensive user interaction to actually make structures of interest appear therefore slowing down the process of diagnosis. This is also an issue from an image quality point of view because it is very difficult to find the optimal balance between noise boosting and structure visibility with manual contrast setting. Moreover, given the low visibility of the edges, tasks such as automatic segmentation of the spine become very challenging, whereas the post-processed images limit the need for manual interaction. These assessments are also valid for the other structures. The proposed algorithms (Figures 3.19b and 3.19c) have better brightness than EOSE (Figure 3.19d), which may demand user interaction to be properly studied. The medium-low frequency structures of the spine are preserved with DE, but high frequencies are lost in the vertebral body, as shown by magnitude of gradient of DE and LNCE in Figures 3.19f and 3.19g, respectively. While DE returns a quite regular image that could be more adapted for instance to the automatic detection of the vertebral body, LNCE suits better for diagnosis because it reflects the textured nature of the bones. The higher ALV value obtained with LNCE is a further demonstration in this discussion. Finally, note that the gradient magnitude is extremely

low in the tissues surrounding the vertebral body, while it is not inside it. This means that the LNCE enhances information of real interest and the contribution of noise to the obtained ALV value is negligible. In the thoracic spine, the ALV increases with respect to the input by: ×18.87 (NE), ×5.34 (DE), ×12.80 (LNCE) and ×8.58 (EOSE). These values are very similar to those obtained in the lumbar spine region because the structures of interest, i.e. vertebrae, are the same.



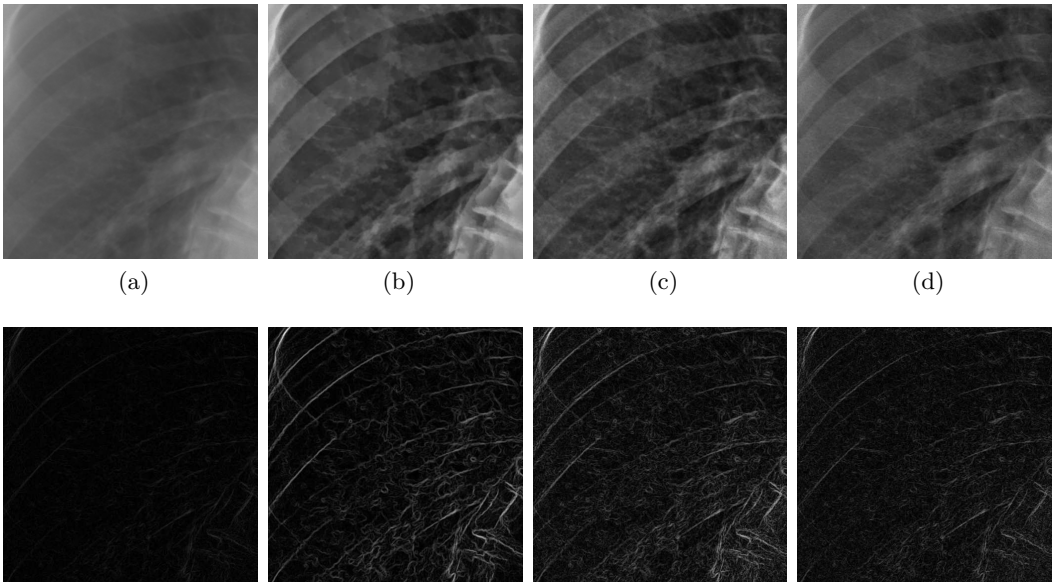
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.20: *Visual comparison of the proximal femur: top row = images and bottom row = related Sobel gradient magnitudes; columns from left to right = IN, DE, LNCE and EOSE.*

In the proximal femur (Figure 3.20) the ALV increases with respect to the input by: ×18.15 (NE), ×2.20 (DE), ×9.59 (LNCE) and ×8.86 (EOSE). The same trend can be noted as for region $\mathcal{A}_1$, but the ALV score obtained with DE is significantly lower than LNCE. Moreover, Figure 3.20b shows that there is a slight loss of resolution inside the femur, which justifies the low ALV value. Two aspects explain this result. First the proximal femur has a more limited amount of medium-low structures than the vertebrae as the only structures of this type are the femur edges and the cotyloid cavity. Secondly, as observed by analyzing the high resolution grid of the phantom (Section 3.4.4), the low amount of signal makes it more difficult to detect redundancy and, so, to preserve high frequency details. The image post-processed with LNCE (Figure 3.20c) is fully appropriate for the diagnosis as, in addition to the femur and cotyloid cavity boarders, it gives a rendering of the bone texture that is superior to that obtained by using EOSE (Figure 3.20d). Moreover, by comparing the gradient magnitude values, it is possible to remark that, while EOSE enhances bone texture and noise in soft tissues in approximately the same way (Figure 3.20h), the two are clearly processed in different ways with LNCE (Figure 3.20g), but some noise residual in soft tissues is still present. Finally, $\mathcal{A}_3$ is a typical region where LNCE is preferred if a diagnostic image quality is desired.

In the lungs (Figure 3.21), the ALV increases with respect to the input by:



(a) (b) (c) (d)

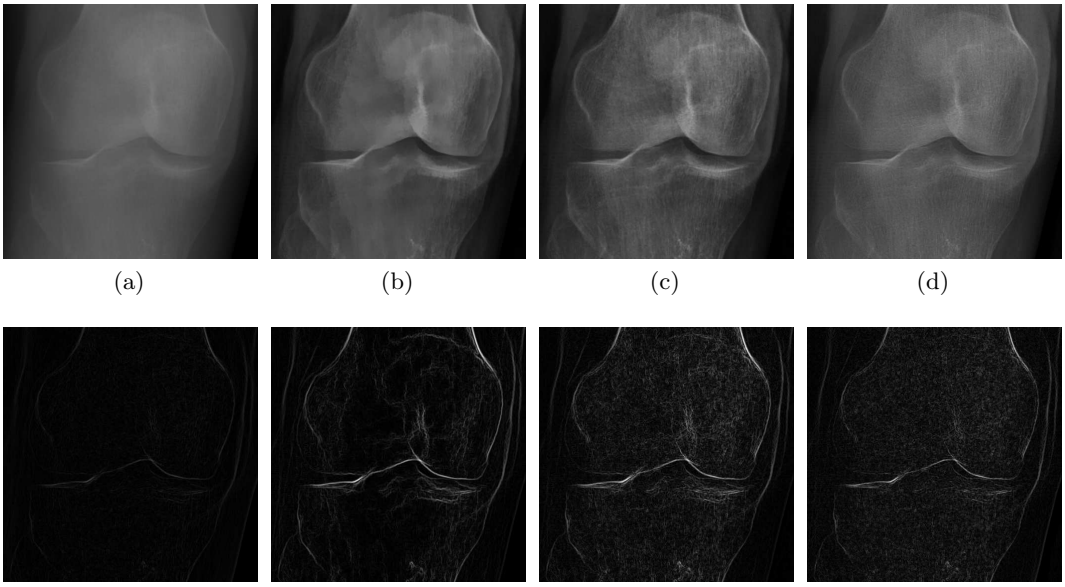
(e) (f) (g) (h)

Figure 3.21: *Visual comparison of right lung in $\mathcal{A}_4$: top row = images and bottom row = related Sobel gradient magnitudes; columns from left to right = IN, DE, LNCE and EOSE.*

×12.82 (NE), ×8.28 (DE), ×11.30 (LNCE) and ×7.16 (EOSE). Since the lungs are a low density region and rich in medium low frequency structures, DE provides a good outlook (Figure 3.21b). However, by comparing it with LNCE (Figure 3.21c) and considering the corresponding gradient magnitude values (Figures 3.21g and 3.21h), a slight attenuation of fine lung structures is observed, which explains the lower AVL value associated with DE. However, when compared to EOSE (Figure 3.21d), both DE and LNCE show a better capacity to distinguish tissues associated with different densities, e.g. the ribs and the spine from the air in the lungs, which further gives merit to the proposed boosting technique (Section 3.3.3). Moreover, since the noise in this region is almost absent, it is logical that NE and LNCE have similar performances. This is not the case for the structures so far examined, which highlights the adaptability of the proposed non-parametric noise containment-based method to intra-patient variability.

Finally in the knee (Figure 3.22), the ALV increases with respect to the input by: ×17.78 (NE), ×4.97 (DE), ×12.22 (LNCE) and ×9.61 (EOSE). The most important information in this region is the bone texture, which is a high frequency image feature. The method DE is not adapted to process this region as justified by the corresponding low ALV value and visual rendering (Figure 3.22b) when compared with LNCE (Figure 3.22c).

The observations derived from the samples in Figure 2.9 are consistent for the whole database on which we have computed ALV measures as shown by the average AC and UCII measures on the totality of the dataset reported in Table 3.3. The highest value of AC is obtained with NE. However, this is mainly due to strong noise contribution in anatomical ROI. This is confirmed by the value of UCII that is much smaller than zero, which means that the overshooting is too important to



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.22: *Visual comparison of knee: top row = images and bottom row = related Sobel gradient magnitudes; columns from left to right = IN, DE, LNCE and EOSE.*

Table 3.3: *Average values of AC and UCII computed over the inputs of the clinical database and respective outputs post-processed with NE, DE, LNCE and EOSE.*

	IN	NE	DE	LNCE	EOSE
AC	2.01	25.14	3.79	11.45	5.69
UCII	0	−3.09	1.80	1.37	−1.16

actually take NE into consideration. The algorithms DE and LNCE allow avoiding overshooting and can be, in our opinion, alternatives to answer different clinical needs. Indeed, DE could be more adapted for tasks such as automatic bone segmentation and registration, because it is more regular than LNCE while preserving edges. However, a noise containment oriented method is more suitable to be presented to users for medical imaging aided diagnosis, which is confirmed by the higher AC value when using LNCE. Finally LNCE outperforms parametric noise containment methods, here represented by EOSE. Indeed, with respect to EOSE, the anatomical contrast doubles while avoiding over-shooting. Moreover, as shown by analyzing the regions $\mathcal{A}_i$, LNCE has the ability of adapting itself to the heterogeneous features of tissues present in a full body image, which proves that this method very flexible.

3.5.2 Quantitative feedback from clinicians

The last part of the evaluation addresses a practical and fundamental matter, i.e. the quantification of how much the proposed method can help clinicians to better see anatomical structures. Given the conclusions in the previous section, the study focuses on noise containment-based approaches and, therefore, we limit ourselves to compare LNCE and EOSE. The application of DE to object segmentation is not

studied as it would require to evaluate different segmentation techniques, which is not a subject of this thesis.

For this purpose we have worked in collaboration with a radiologist at the AP-HP, Cochin hospital in Paris (France). We have considered a set composed by 10 patients, between 54 and 81 years old, having Body Mass Index (BMI) between 20.28 and 28.93 corresponding to normal and overweight adults, respectively. None of the patients presents particular pathology. Therefore, only the visibility of a set of diagnostically relevant structures of the human skeletal apparatus has been evaluated.

In practice, the radiologist has assigned a vote to each structure that is representative of its degree of visibility. The votes go from 0, meaning that the structure is not visible at all, to 5, meaning that the structure is perfectly visible. Given the interest in a full-body diagnosis, different anatomical regions have been examined and the following list of structures retained:

- **Frontal cervical spine** ($\mathfrak{C}^{\mathfrak{f}}$): spinous process (1), vertebral body (2), intervertebral disc (3), zygapophyseal joint (4), transverse process (5), pedicle (6).
- **Lateral cervical spine** ($\mathfrak{C}^{\mathfrak{l}}$): median atlanto-axial joint (1), intervertebral disc (2), vertebral body (3), transverse process (4), uncus (5), pedicle (6), lamina (7), posterior interapophyseal joints (8), spinous process (9).
- **Frontal thoracic spine** ($\mathfrak{T}^{\mathfrak{f}}$): spinous process (1), vertebral body (2), intervertebral disc (3), transverse process (4), costovertebral joints (5), pedicle (6).
- **Lateral thoracic spine** ($\mathfrak{T}^{\mathfrak{l}}$): spinous process (1), vertebral body (2), intervertebral disc (3), transverse process (4), posterior interapophyseal joints (5), pedicle (6).
- **Frontal lumbar spine** ($\mathfrak{L}^{\mathfrak{f}}$): spinous process (1), vertebral body (2), intervertebral disc (3), transverse process (4), pedicle (5), posterior interapophyseal joints (6), sacrum (7).
- **Lateral lumbar spine** ($\mathfrak{L}^{\mathfrak{l}}$): spinous process (1), vertebral body (2), intervertebral disc (3), transverse process (4), pedicle (5), posterior interapophyseal joints (6), intervertebral foramina (7), lamina (8).
- **Pelvis** ($\mathfrak{P}$): sacrum (1), sacroiliac joint (2), cotyloid cavity (3), pubic symphysis (4), hip bones (5), hip joint (6), ischium (7), femur (8).
- **Knee** ($\mathfrak{k}$): patella (1), femoro-tibial joint (2), medial and lateral intercondylar tubercles (3), condyles (4), fibula (5), tibial plateau (6).

In total, 56 structures have been taken into account. The average votes for each structure evaluated over LNCE and EOSE are reported in Figure 3.23. Regional qualitative measures can be defined from these votes to derive summarized evaluations. An anatomical region is considered as perfectly visible as long as all the structures therein get a vote of 5/5. Then, a percentage is assigned to the anatomical ROI by comparing the total of the votes to the ideal optimum. The average scores and points gained by the proposed method are provided in Table 3.4.

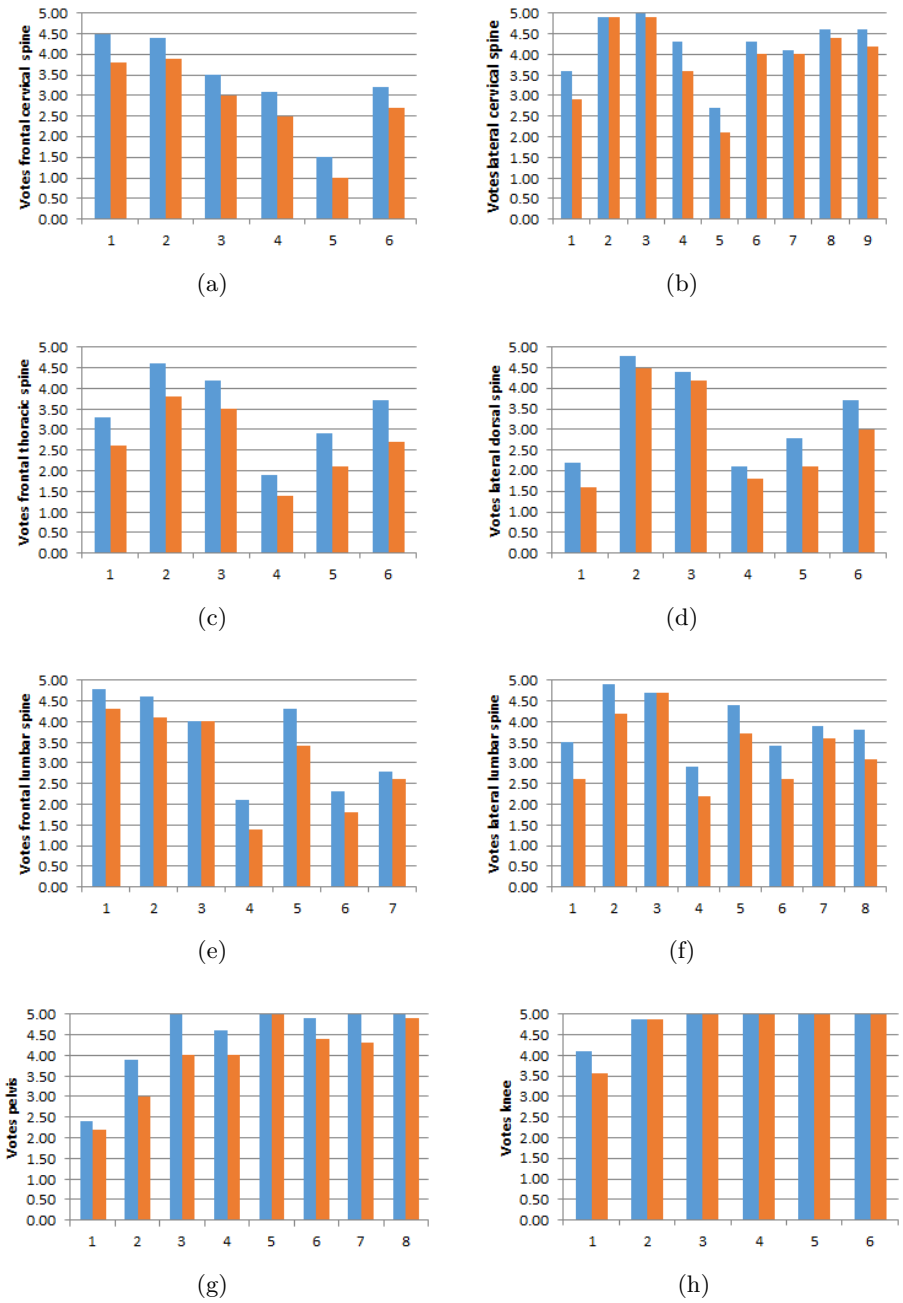


Figure 3.23: *Average votes assigned to the structures. Blue = LNCE and orange = EOSE: (a)* $\mathfrak{C}^{\mathfrak{f}}$*; (b)* $\mathfrak{C}^{\mathfrak{l}}$*; (c)* $\mathfrak{D}^{\mathfrak{f}}$*; (d)* $\mathfrak{D}^{\mathfrak{l}}$*; (e)* $\mathfrak{S}^{\mathfrak{f}}$*;(f)* $\mathfrak{S}^{\mathfrak{l}}$*; (g)* $\mathfrak{P}$*; (h)* $\mathfrak{K}$*.*

These results show that the proposed method enhances the visibility of osteo-articular structures. Note that the two algorithms have similar performances for structures that are surrounded by thin soft tissues, e.g. the knees.

The spine is a complex structure because of the significant tissue superposition and the projection of a rotated 3D volume on a 2D plane. Since in all the regions of the spine the votes of LNCE are higher than EOSE ones, the proposed method is

Table 3.4: *Regional qualitative measures for each evaluated anatomical ROI.*

	$\mathfrak{C}^{\mathfrak{f}}$	$\mathfrak{C}^{\mathfrak{l}}$	$\mathfrak{D}^{\mathfrak{f}}$,	$\mathfrak{D}^{\mathfrak{l}}$	$\mathfrak{S}^{\mathfrak{f}}$	$\mathfrak{S}^{\mathfrak{l}}$	$\mathfrak{P}$	$\mathfrak{K}$
LNCE	67%	85%	69%	67%	71%	79%	90%	97%
EOSE	56%	78%	54%	57%	62%	67%	80%	95%
Gain	**+11**	**+7**	**+15**	**+10**	**+9**	**+12**	**+10**	**+2**

more appropriate for facilitating the visibility in challenging cases. However, some objects remain difficult to see. For example, in most of the cases the transverse process represents a challenge ($\mathfrak{C}^{\mathfrak{f}}{}_5$, $\mathfrak{T}^{\mathfrak{f}}{}_4$, $\mathfrak{T}^{\mathfrak{l}}{}_4$ and $\mathfrak{S}^{\mathfrak{f}}{}_4$). Globally, about 10 points are gained with respect to EOSE in the spine. The image quality in the pelvis improves significantly with the proposed method: using EOSE the image is of good quality (80%), but LNCE provides almost a perfect image (90%). This is due to the fact that the structures in the pelvis are quite regular and, therefore, easy to extract with the proposed method, even if the signal is low. The only problematic object is the sacrum ($\mathfrak{K}_1$) because is covered by a lot of soft tissues. Finally, for what concerns the knee, the improvement is negligible (see Table 3.4) and only the patella ($\mathfrak{k}_1$) is easier to delimit with LNCE. Indeed, the score is almost perfect because the absorption is very low. Then, in this region, it would be more interesting to compare the two algorithms in pathological cases that, for example, concern the health of the bone tissue.

In conclusion, this diagnostic feedback gives further value to the proposed method as it indicates that LNCE could potentially be used in clinical routine.

3.6 Conclusion

In this chapter we have discussed the joint denoising and contrast enhancement of X-ray digital images. First, we have characterized the noise model. We have then proposed the XNL-means filter as an extension of the classical NL-means filter where the distance among patches is automatically adjusted according to the signal-dependent noise level estimated by means of the percentile method. This filter has then been used to defined a flexible framework for both denoising and contrast enhancement of X-ray images. Its flexibility has been demonstrated by quantifying the robustness to changes in the input signal strength, patient morphotype and features of the multiple tissues in the field of view. Moreover, we have presented the possibility of choosing different options in the framework in order to fulfill different clinical needs. Indeed, the denoised input image can be fully enhanced to obtain a quite regular output where the edges are well preserved (algorithm DE). As an alternative, by comparing the input image and the filtered one encoded in multiple scales, non-parametric noise containment maps can be estimated and, then, used to limit the impact of the noise when enhancing detail coefficients (algorithm LNCE).

An important number of tests on both phantom and clinical images have led to the conclusion that DE could be helpful for tasks such as segmentation or registration, but it is sub-optimal for visual diagnosis due to a resolution loss equal to 0.2 up to 0.4 lp/mm quantified on phantoms and visually assessed on clinical images. On the other hand, the noise containment oriented approach overcomes this drawback, and, furthermore, improves the image quality compared with a parametric

noise containment-based algorithm that is used in clinical routine. Finally, we have assessed on phantoms images that the proposed noise containment-based method provides good image quality even when the dose is reduced of a factor equal to 7.10 over pediatric morphotypes, which implies that our method is promising in the optic of reducing the dose employed in pediatric exams. The results indicate that the very same statement is not valid for adult morphotype, but the gain in image quality still subsists and, hence, we estimate that the dose could also be reduced in other exams than the pediatric ones, but a factor 6.62 is probably too high.

The feedback from the users are positive for a large number of structures. However, according to the results presented in Section 3.5.2, some structures remain very difficult to see. This is strictly related to the acquisition conditions: the amount of dose is not sufficient to fulfill a given medical purpose. Therefore, it could be of great help in the diagnosis to quantify this information by means of an image quality measure. Note that this could also be useful in the opposite situation, i.e. the indicator would highlight in this case an excessive amount of dose with respect to the minimum sufficient one. This measure is the exposure index (IEC 62494-1, 2008; Shepard et al., 2009) and is discussed in the following chapter.

Chapter 4

Exposure index: overview and a new landmark based approach

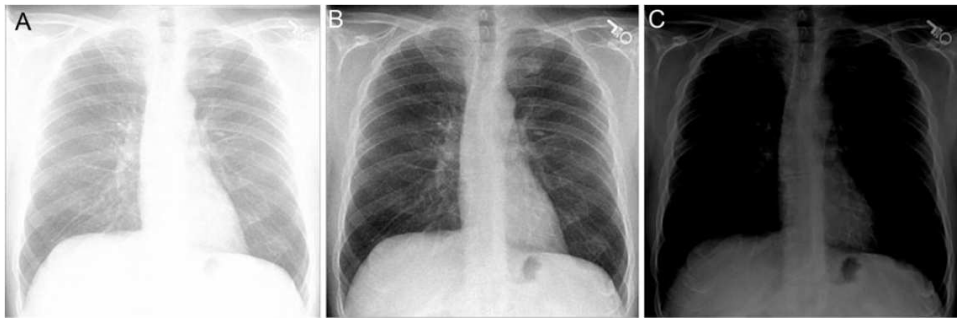
In the previous chapter we have shown that image processing techniques improve the image quality for a given amount of dose. Besides, tests on phantoms show that the proposed approach may allow for a sensible reduction of dose in pediatric exams while approximately preserving the same quality of a diagnostic EOS image post-processed with an algorithm used in clinical routine. However, the quality of the outcomes does not only depend on how the image is processed, but also on the chosen parameter setting. It would then be useful to have a measure capable of assessing if an image has been acquired at the correct amount of dose. According to the literature the exposure index (EI) answers this need. In the optic of defining an exposure management service for EOS images, this indicator of image quality is particular interesting. Therefore, in this chapter we give an overview of the EI for digital radiography and contextualize its role to the exposure management service. In particular, we stress the importance of defining significant ROIs to estimate EI values that are consistent across different exams.

Organization of the chapter - Section 4.1 overviews the concept of EI, the algorithm to estimate it and its role in an exposure management system. In particular, we introduce the measure of anatomical SNR as alternative to the definition of EI given in the norm and we clarify the link between the two measures. Section 4.2 presents the challenges related to EOS full-body images and applications for which the EI may be useful. Section 4.3 reviews state of the art methods for the definition of ROIs and analyzes their performances on full-body EOS images. In Section 4.4 we apply the localization of structures of interest to the estimation of EI values and, in particular, we propose a landmark-based method. Section 4.5 summarizes the main ideas and the open issues that are addressed in the following chapter.

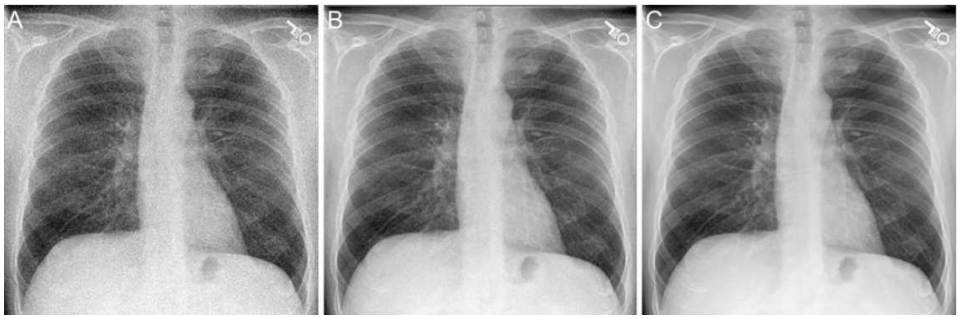
4.1 Measures of exposure level in digital radiography

4.1.1 Motivation

The EI has been proposed by the medical community because, as established in the last two decades (Seibert et al., 1996), digital X-ray imaging systems present the phenomenon of *exposure creep*. This is due to one of the advantages of digi-



(a)



(b)

Figure 4.1: *The exposure creep in digital radiography: (a) Screen-film imaging; (b) Digital imaging. The capital letters indicate: (A) underexposure; (B) correct exposure; (C) overexposure. The images are extracted from the paper by Seibert and Richard (2011).*

tal radiography over screen-film detectors, that is the capacity in producing well contrasted outcomes at wider dose ranges than analogical systems (Shepard et al., 2009). In detail, in screen-film imaging, the quality totally depends on acquisition conditions because the image is not post-processed. For example, Figure 4.1a shows that traditional screen-film systems give an immediate intuitive feedback on the image quality. On the contrary, the global brightness on digital images is practically not affected by changes in acquisition conditions (see Figure 4.1b), because it is adjusted in the post-processing. As a consequence, generally, underexposed images do not look too bright and overexposed images do not look too dark. However, the noise level changes according to acquisition conditions. While the underexposed image has much worse image quality in terms of noise than the correctly exposed one, the overexposed case does not present any particular issue and it even entails an improvement of the SNR. There is then a risk that users get used to the better quality of overexposed images by, hence, increasing the average radiation exposure in radiography examinations (Shepard et al., 2009). For this reason, it is useful to have an indicator that can help controlling the image quality and acquiring images at the ALARA amount of dose. The EI expresses the amount of signal level that reaches the detector and, hence, it must not be confused with patient radiation dose. Nevertheless, since it is proportional to the squared SNR (Seibert and Richard, 2011), it can be used to define the lower limit of radiation exposure depending on the medical purpose of the exam and the maximum acceptable amount of noise for clinicians.

The EI has been standardized by means of a joint initiative of the International Electrotechnical Commission (IEC 62494-1, 2008) and of the American Association of Physicists in Medicine (Shepard et al., 2009) with the aim of comparing different digital radiography systems or techniques between various institutions.

Shepard et al. (2009) have also pointed out that the EI could be used as an indicator for signal level in a automatic exposure control oriented system. This aspect mainly motivates out interest in the EI measure, as one of the objectives of this thesis consists in studying the feasibility of introducing, in the acquisition workflow of an EOS image, an automatic system capable of delivering the same amount of signal despite variations in radiological thickness. Finally, in our work we focus on estimating image quality on EOS images and on embedding this indicator in an exposure management system, as discussed in Section 4.2.2. On the contrary, the comparison of EOS with other digital imaging systems based on the EI is out of the scope of this thesis.

4.1.2 Exposure index algorithm

Figure 4.2 shows the flowchart of the EI algorithm. The pre-processing allows obtaining the image $\mathbf{u}_r$ that coincides with the raw image defined in Section 3.1.1 and from which the EI value is estimated. The pre-processing steps allow neglecting imperfections of the detector and, hence, consist in gain, offset and dead pixels correction. The geometric distortions are also corrected. Nevertheless, other kinds of image processing such as noise reduction and contrast enhancement have to be avoided because these operations significantly influence the image quality and the EI has to quantify the amount of exposure from raw images (IEC 62494-1, 2008).

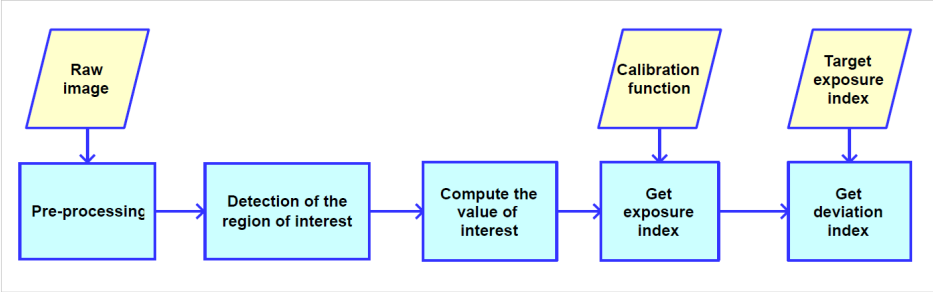


Figure 4.2: *Sequence of steps that allows computing the exposure index and the deviation index.*

Then, a subset Ψ of the whole pixel space Ω is defined to associate the EI with X-rays that are attenuated by meaningful structures according to the medical purpose of the undergoing exam. In the literature, this step is called ROI selection and Section 4.3 is mainly dedicated to present state of the art methods for ROI definition and to evaluate their adaptability to EOS images. The methods that are given as examples in the norm IEC 62494-1 (2008) and by Shepard et al. (2009) are based on gray level histogram thresholding, positioning of rectangular or circular shapes at the center of the envelop of the patient or a combination of these two techniques. The ROI selection may be problematic because of the presence of metallic objects, unexpected positions of body parts and changes in the morphotype of the patient.

Given the mentioned issues, the techniques used to define the ROIs are just given as examples and are not imposed by the norm IEC 62494-1 (2008). Moreover, the manufacturer is invited to give users the possibility of validating and, eventually, of modifying the automatically selected ROI. Therefore, according to the original formulation, the EI should rely on user-interaction to validate the measures, which can be seen as a remedy to the aforementioned issues related to the ROI definition. Nevertheless, according to our analysis on the use that is made of the EI in clinical routine, this is not actually true: the users collect the EI and the corresponding deviation index values (Mothiram et al., 2013), whereas the control on the automatically computed ROIs is, to the best of our knowledge, neglected. As a consequence, it is fundamental to be able to estimate these ROIs in an automatic and robust way in the optic of providing consistent EI values in clinical routine.

From the distribution of the gray levels of the pixels $x_i \in \Psi$, a value of interest (VOI) v is extracted. Since v has to represent the central tendency of the distribution, it is equal to the median of $\mathbf{q}(x_i),\ x_i \in \Psi$ (Shepard et al., 2009). The EI value is then computed as follows:

$$EI = c_0 g(v) \tag{4.1}$$

where c_0 is a constant fixed at $100 \mu Gy^{-1}$ according to the norm IEC 62494-1 (2008) and $g(.)$ is a calibration function. The value returned by $g(v)$ is the Kerma in the air at the receptor associated with the ROI Ψ and it is expressed in μGy. The function $g(.)$ depends on the X-ray system and must be defined in the X-ray standard beam geometry and calibration conditions specified in the norm IEC 62494-1 (2008). The definition of this function for EOS system is not addressed in this thesis as it is beyond our original scope.

The EI value on its own is not really significant as what really matters is to assess if the amount of noise in a given image is tolerable according to the anatomical region of interest and the medical purposes of the exam. The manufacturers are in charge of defining target exposure (EI_t) index values according to ideal ALARA dose conditions. However, the users must be allowed to eventually update these target values. The deviation index (DI) from the correct exposure EI_t is then computed as follows:

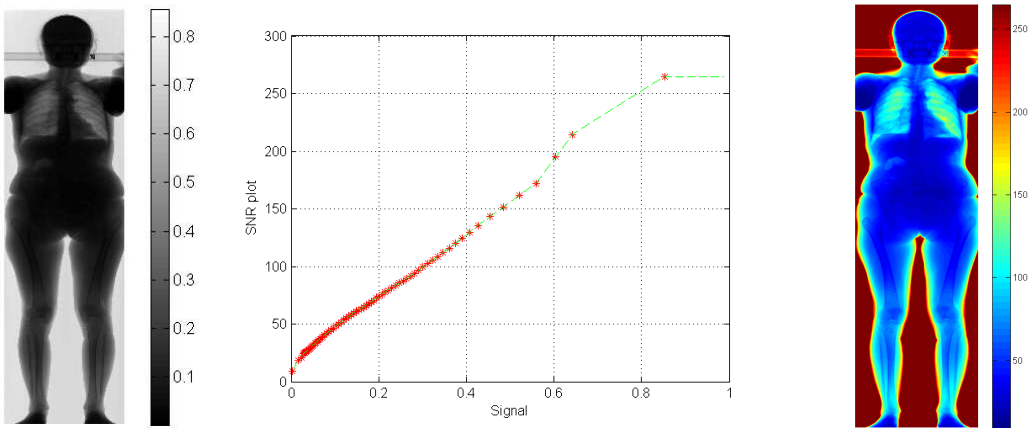
$$DI = 10 \log_{10} \left(\frac{EI}{EI_t} \right) \tag{4.2}$$

where $DI < 0$ and $DI > 0$ respectively indicate underexposure and overexposure. In practice, the acquisition setting would need to be changed only if $|DI| > 1$, i.e. if the estimated EI value changes of $+25\%$ and -20% with respect to the target EI value.

4.1.3 Proposed measure based on SNR

We propose to compute another image quality measure that is indirectly connected to the exposure level in X-ray images. The calibration function block (see flowchart in Figure 4.2) is replaced by a SNR function that maps intensity levels to SNR values.

For this purpose, we exploit our work on the estimation of noise levels from clinical images with the percentile method (see Section 3.1.3). Since each bin t of the estimated noise curve is associated with a signal level μ_t, a SNR curve $s(.)$ is



(a) (b) (c)

Figure 4.3: *The SNR on clinical images as alternative to the exposure index. From the image in (a) we define a look-up-table between the signal values and the SNR values, that is the SNR function in (b). The resulting SNR map is reported in (c).*

computed by exploiting the empirical SNR definition (see Section 2.3.1). Formally, the anatomical SNR associated with a bin t is defined as follows:

$$s_t = \frac{\mu_t}{\sigma_t} \tag{4.3}$$

where σ_t is the noise standard deviation at the bin t that is estimated by using the percentile method. Finally, the resulting samples s_t are interpolated to obtain the curve $s(.)$. It is worth noting that the consistency of this measure is straightly justified by the correct estimation of the noise curve $\sigma(.)$ associated with the image, which has been demonstrated in Section 3.1.4.

Figure 4.3b shows the $s(.)$ function estimated from the input image in Figure 4.3a. Figure 4.3c contains the SNR map where the value at a pixel x_i represents the SNR in a window of size 15×15 centered at x_i. Note that the estimated SNR values may not be consistent in the background and in very thin soft tissues due to detector saturation and proximity to the borders of the patient envelop, respectively. However, this does not constitute an issue for the EI application because anatomical structures of interest are located where the estimation of the SNR is meaningful. Indeed, the highest SNR value in an anatomical ROI is registered in the lungs and it is approximately equal to 120 (see the colorbar in Figure 4.3c). Up to this value, the function $s(.)$ follows a logical trend when compared to the curve in Figure 3.2c. The estimated curve looks almost linear because the majority of the intensity values associated with the bins are higher than 0.05. The root-square like trend would be more clearly visible at lower signal levels as, however, it can be observed by looking at the leftmost side of the curve in Figure 4.3b.

Finally, the SNR-based EI is computed as follows:

$$\tilde{EI} = s(v). \tag{4.4}$$

Compared to the definition given in Equation 4.1, the calibration process is replaced by SNR estimation directly from the image and, hence, there is no need to model

the behavior of the detector or to simulate the effects of the pre-processing steps that are used to get the image $\mathbf{u}_r$.

4.1.4 Link between EI and SNR based measures

The definition given in Equation 4.4 does not respect the norm IEC 62494-1 (2008), but it has many points in common with it. First, the final purpose of the EI is respected as the SNR is clearly an indicator of the amount of noise and, thus, of the exposure level in the raw image. Secondly, it is true that the DI value changes by using the definition given in Equation 4.4, however, since the EI is proportional to the squared SNR (Seibert and Richard, 2011), i.e. $EI \propto SNR^2$, by assuming that the EI and EI_t values have approximately the same proportionality factor with respect to the associated SNR values, the following relation is valid:

$$DI \approx 10 \log_{10} \left(\frac{SNR}{SNR_t} \right)^2 = 2DI_{SNR}. \tag{4.5}$$

As a consequence, the DI value computed as in the norm IEC 62494-1 (2008) can be approximated with $DI = 2DI_{SNR}$, by using the definition given in Equation 4.4. Therefore, as long as the indicator quantifies the deviation relative to ideal conditions, the two measures are equivalent. Note that this assessment is also valid if the indicator of image quality is embedded in an exposure management framework to drive the acquisition parameter tuning.

Finally, while the definition of the measure has been deeply studied, the technique to select the ROI Ψ, which is fundamental to estimate robust EI values on clinical images, needs to be specifically addressed. In this case, both EI and $\tilde{EI}$ quantify the image quality and, hence, we will use the definition given in Equation 4.4. Without loss of generality, we will consider anatomical SNR and EI values as synonyms in our discussion.

4.1.5 Exposure management

The automatic exposure control (AEC) is widely used for exposure management of X-rays images. A device is placed between the patient under exam and the detector, and it measures the amount of signal. The emission of X-rays is stopped when the signal level reaches a predefined threshold that corresponds to the ideal acquisition conditions. In digital radiography, the AEC is used to preserve the ideal EI value at varying thickness, but this raises several issues (Shepard et al., 2009). Indeed, the EI could be computed in a region that is different in position and size from the one associated with the AEC. Besides, the positioning of the AEC device is not trivial in some applications, e.g. in intensive care units (Foos et al., 2012). For these reasons, some researchers have investigated the possibility of controlling the amount of delivered dose without an AEC device. This is possible by reading images that are acquired at considerably inferior amounts of dose than those used in diagnostic exams. An operator can use this preview, also called trial shot, to interpret the image quality in a given ROI, and, then, exploit this information to decide which is the most convenient parameter setting for the diagnosis. Nevertheless, this demands to know very well the system and significantly slows down the workflow. Therefore,

the optimal image quality measure is rather automatically predicted from the preview and the default acquisition parameters are properly changed if they mismatch the forecast. Foos et al. (2012) have proposed to use the CNR that is computed as a function of average intensities in two ROIs that are placed according to the anatomical region under exam, and of the noise model. It is worth noting that any image quality indicator can be used for this purpose, such as EI and $\tilde{EI}$ defined in Equations 4.1 and 4.4, respectively.

Over the past years, the AEC has also been extensively used to reduce the dose in computed tomography (CT). Most of the CT exams present multiple ROIs with different attenuation values. Then, a fixed tube current entails a sub-optimal distribution of dose. For this reason, AEC devices are used to modulate the tube current in three dimensions depending on anatomical regions and size of the patient. A consistent amount of work proves the advantage in terms of dose reduction (McCollough, 2005) and the exploitation in clinical routine has been studied. For example, Söderberg and Gunnarsson (2010) have compared the performances of different manufactured AEC systems and Papadakis et al. (2014) have studied the robustness to changes in patient size and anatomical regions. Nevertheless, there are no publications that clearly explain how AEC systems define image quality levels. An online document[1] suggests that the Siemens Care Dose 4D CT uses a low dose trial shot for predicting tube current profiles.

Finally, image quality indicators, such as the EI, can also be used to predict the ALARA optimal conditions from previews. This is a relevant application by considering the objectives of this thesis and, thus, is studied on EOS images. While the importance in the clinical workflow of the EI and AEC have been widely addressed in the literature, in the following sections we focus on a less studied aspect, that is the definition of a significant ROI Ψ. We contextualize the problem to EOS images by both evaluating state of the art methods and, then, proposing an alternative and a more robust approach based on landmark detection.

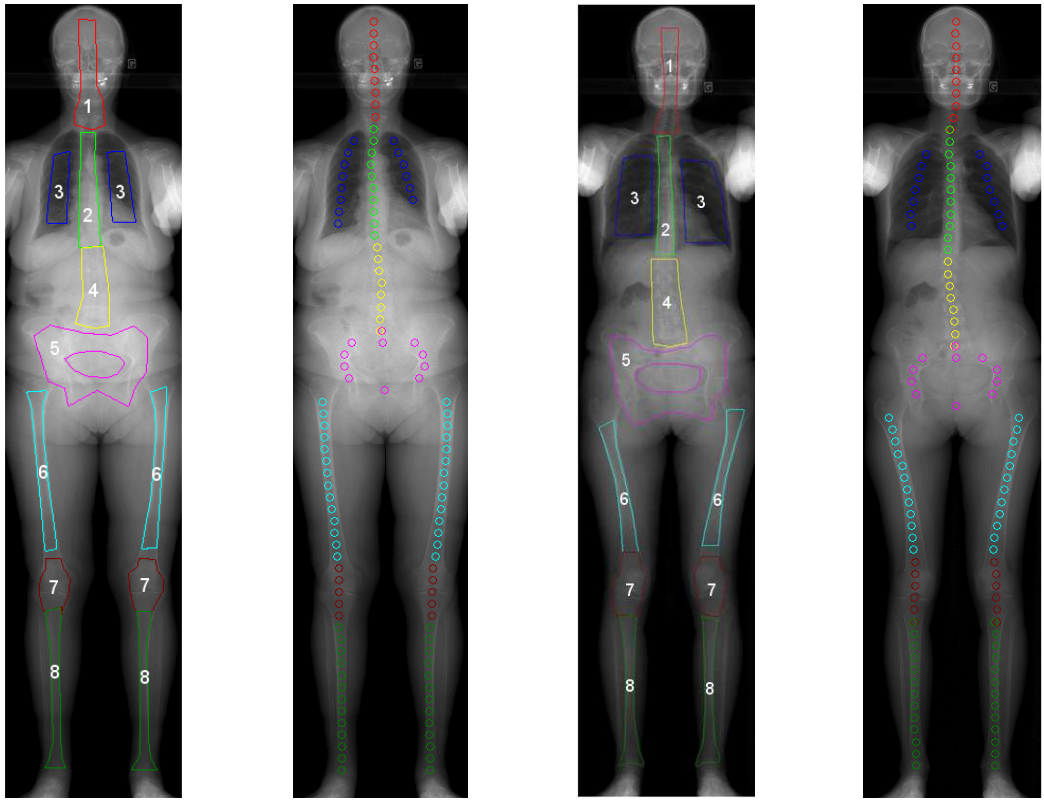
4.2 Set of exposure index values on EOS images

4.2.1 Challenges related to full-body X-ray images

A full-body EOS image shows different anatomical structures of interest in the field of view. Therefore, a unique measure of image quality does not contain a sufficient amount of information to give clinicians a direct feedback on the quality of an EOS acquisition. This assessment is intuitive but can be empirically confirmed. The most complex case in terms of amount of anatomical structures is analyzed, i.e. a full-body exam. Eight regions of interest are taken into account: head and cervical spine ($\mathcal{A}_1$), thoracic spine ($\mathcal{A}_2$), lungs ($\mathcal{A}_3$), lumbar spine ($\mathcal{A}_4$), pelvis ($\mathcal{A}_5$), femur ($\mathcal{A}_6$), knee ($\mathcal{A}_7$) and tibia ($\mathcal{A}_8$). These ROIs (see Figure 4.4) are manually segmented and, then, the corresponding values of interest v_r are computed as the median values of the distribution of intensity levels at the pixels belonging to the ROIs, that is:

$$v_r = \text{median}\{\mathbf{u}_r(x_i) \mid x_i \in \mathcal{A}_r\}. \tag{4.6}$$

[1] http://www.mayo.edu/research/documents/care-dose-4d-ct-automatic-exposure-control-system/DOC-20086815



(a) (b) (c) (d)

Figure 4.4: *The EI value is computed from ROI where the X-rays are attenuated by an anatomical structure of interest. Therefore, in order to define ground truth EI values, the ROIs relative to* $\mathcal{A}_r$, $r = 1, \ldots, 8$ *are manually defined. Here two examples for patients of different size are shown: (a) big adult; (c) adult. As discussed in Section 4.4.2, each anatomical structure can be well described by a cluster of points that are reported in Figures (b) and (d), corresponding to (a) and (c), respectively.*

Two full-body EOS exams are sufficient to present the challenges related to full-body EOS images. The parameter setting for both exams are the same (90 kV , 200 mA and 4 C), the patients have the same gender, but different sizes and ages (70 and 54 years old). The EI values are obtained by using the formula in Equation 4.4 from the values of interest v_r associated with the ROIs $\mathcal{A}_r$ (Equation 4.6). The results, which are reported in Figure 4.5, show that the image quality can be quantified by a set of EI values, each one corresponding to an anatomical ROI. The EI values significantly change according to the corresponding anatomical regions, which indirectly shows that a unique measure is not sufficient to evaluate the quality of an EOS image. The number of elements in the set depends on the number of ROIs that are in the field of view. For example, for a diagnosis of lower limbs, there would be only four EI values that correspond to $\mathcal{A}_5$, $\mathcal{A}_6$, $\mathcal{A}_7$ and $\mathcal{A}_8$. The set of EI values follows an intuitively expected trend and can be considered as a signature that describes the amount of attenuation in anatomical regions for a given patient. For

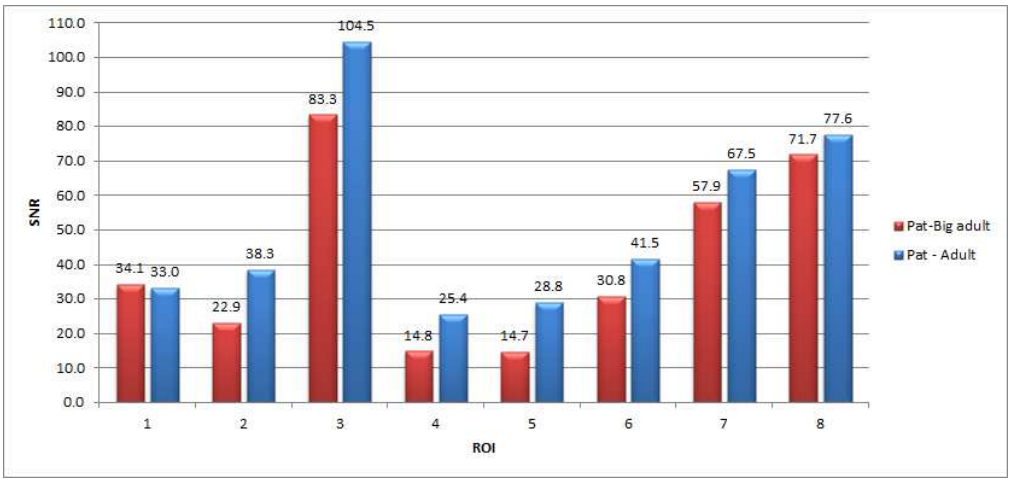


Figure 4.5: *Sets of EI values computed over patients of morphotypes: big adult (red); adult (blue). These sets of values well describe the difference between the sizes of the two patients and indirectly quantify the amount of X-ray attenuation in anatomical regions.*

example, the pelvis and the lumbar spine exhibit the lowest EI values because, for a frontal acquisition, they are the thickest regions to pass through. Similarly, given the significant presence of air, the highest EI value is not surprisingly associated with the lungs.

The sets of values also capture the difference in size between patients that have been imaged in the same conditions. Indeed, the EI values computed from the patient in Figure 4.4a are lower than those computed from the patient in Figure 4.4c. However, it is worth noting that the relation between the elements that belong to the set of EI values cannot be simply modeled. Indeed, for the patient in Figure 4.4a the EI value in $\mathcal{A}_5$ is 2.10 times higher than in $\mathcal{A}_4$, but for the other patient it increases by $\times 1.44$. The sets of values in Figure 4.5 do not only reflect the thicker tissues to pass, but also the distribution of bone, fat and muscular tissues, i.e. the radiological thickness. As a consequence, it is necessary to compute each element of the set rather than inferring them from a single one.

4.2.2 Possible applications

Given a full-body image, multiple EI values are necessary to have a complete information, but, at the same time, it is not intuitive to explain them as, classically, only one EI value is assigned to a X-ray examination. To address this issue, an EI value could be obtained as a weighted mean of those associated with the aforementioned anatomical structures that appear in the field of view. These weights should be predefined depending on the objective of the examination. For example, if the user wants to perform measures of lower limb alignment, higher importance should be given to the EI values computed at the anatomical ROIs $\mathcal{A}_r$, $r = \{5, 6, 7, 8\}$, i.e. the structures in the legs plus the pelvis. On the other hand, higher weights should be assigned to the contributions from the anatomical ROIs $\mathcal{A}_r$, $r = \{1, 2, 4\}$, if the spine is diagnosed. This principle can be also applied to acquisitions that do

not cover the whole body. For example, in a chest examination, both the anatomical ROIs $\mathcal{A}_2$ and $\mathcal{A}_3$ appear in the field of view, but only the lungs should be considered.

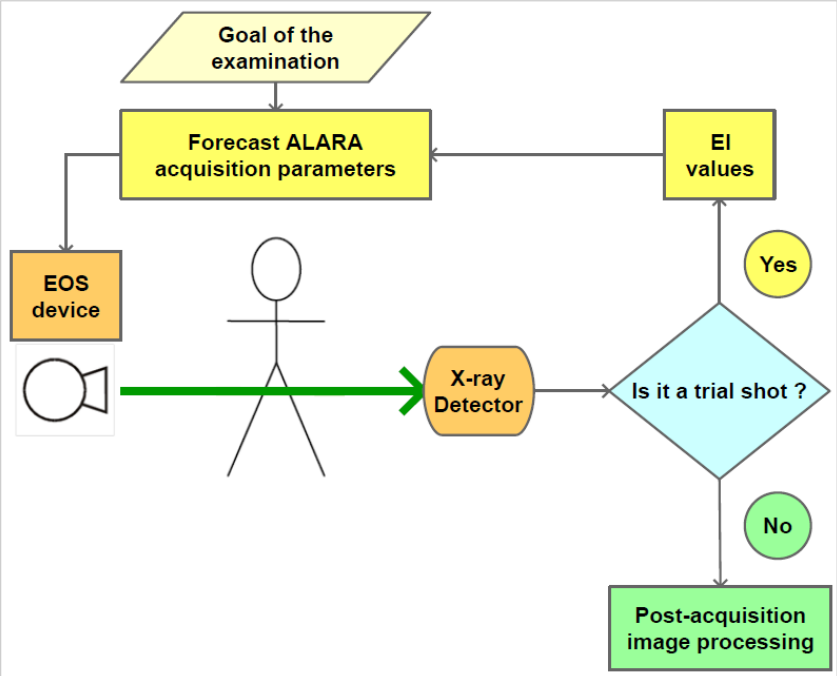


Figure 4.6: *Sequence of steps that represents the previewed AEC system with EOS device.*

Figure 4.6 shows a sequence of steps that suggests how the set of EI values could also be used to implement an AEC system on EOS device. Given the presence of multiple anatomical structures and the acquisition technique based on the linear scanning of the C-arm on which the detector is embedded, we suggest to use a trial shot to drive the acquisition parameter setting. First, a preview image is acquired and the EI set of values are computed from it. These are then combined with an indication provided by the user that specifies the parameters of the examination, i.e. the anatomical region of interest and the aim of the study. A unique EI is computed over preview images as if it was a normal EOS images, i.e. with a weighted average of the values associated with the anatomical structures in the field of view. Afterwards, a module takes care of forecasting the optimal acquisition parameters according to the automatically computed input and coherently with ALARA guidelines. A second scan is performed, this time at a radiation exposure that allows achieving the predefined clinical goal, while limiting as much as possible the dose absorbed by the patient. The process implies that the estimation needs to be provided almost in real time, i.e. the interval that passes between the end of the preview acquisition and the re-positioning of the C-arm for the second scan. The data obtained from the latter are then given as input to the image processing chain.

The set of EI values extracted from a preview image could also be used to drive changes in the tube current during the C-arm scan. Indeed, similarly to CT imaging, at a fixed tube current, the distribution of dose is sub-optimal because the radiological thickness of the anatomical structures changes in different regions of the

body.

Finally, these applications justify the interest in the EI, which, as for any other 2D planar X-ray system, can be computed on EOS images. Nevertheless, the presence of multiple types of structures in the field of view, as in CT scans, requires to compute a set of EI values rather than a single one. The suggested applications require to establish how to define the aforementioned weights to get a single EI from multiple ones or, as for the AEC, how to formally predict good acquisition parameters. We are, however, going to focus on another non-trivial problem that is common to these applications, i.e. the fast and robust identification of ROIs to be used for estimating the EI values. We study this matter by starting from the critical overview of state of the art methods of ROI selection re-adapted to EOS images.

4.3 State of the art approaches applied to EOS images

4.3.1 Overview of state of the art approaches

At first, we may simply adapt the methods for the selection of ROIs that have been suggested in the norm IEC 62494-1 (2008). Therefore, the principal ideas are summarized here. These methods are quite simple as they essentially consist in histogram thresholding or in using shapes with a low amount of degrees of freedom, e.g. squares or circles, placed at the center of the image.

A common step prior to the ROI selection consists in segmenting the envelop of the patient body. Indeed, the part of the image where the X-rays have not been attenuated shall be neglected in order to avoid introducing a positive bias in the measure. The envelop of the patient can be extracted by simply applying a threshold on the gray levels as the intensities values are extremely different depending on whether there is attenuation or not. However, on EOS images a fixed threshold may not be robust enough because of the non linearity of the detector (Despres et al., 2005) and, hence, we propose a slightly more elaborated approach. The pixels of the image after calibration are classified according to their intensity values by using the k-means method (MacQueen, 1967) with four clusters. Then, an initialization of the mask of the envelop is defined as the cluster with the highest centroid gray level. Afterwards, some morphological operations that are often used to segment brain tissues in magnetic resonance (MR) imaging (Mangin et al., 1998) are applied to extract the final mask. In detail, the initial mask is eroded with a circular structuring element of radius equal to 0.2 cm and, then, only the biggest connected component is preserved because external objects are not relevant for computing the EI. Finally, the resulting mask is dilated with the same structuring element. The envelop so computed well respects the contours of the patient. Some marginal soft tissues may not be preserved, but this does not represent a problem for the EI application because the structures of interest are not located at the borders of the envelop. Some examples of segmented body envelops are given in Appendix A. Once a subset of anatomical pixel $\Phi \subset \Omega$ is defined, the following methods can be used to define the ROI $\Psi \subseteq \Phi$.

4.3.1.1 Histogram threshold

Two thresholds $\tau = [\tau_l; \tau_h]$ are applied to the gray levels histogram of $\mathbf{u}(x_i)$ where $x_i \in \Phi$. These thresholds are defined as a function of two percentile values of the histogram. The percentile values are defined according to prior assumptions on the link between intensity values and corresponding physical meaning. For example, the leftmost part of the histogram is associated with strong absorption that is intuitively higher in bone tissues, i.e. the information of interest. Nevertheless, changes in patient morphotypes, acquisition conditions, presence of metallic objects and so on, influence the gray level distribution and, hence, this way of defining the ROI may not be consistent across different exams. We will refer to the histogram-based method with the abbreviation *hROI* and Ψ_τ will denote the corresponding ROI.

4.3.1.2 Center of the image

The ROI can be placed at the center of the image by assuming that the most significant information is centered. The ROI is then, for example, a square or a circle that covers at least 20% of Φ IEC 62494-1 (2008). In our experiments, we use a square Ψ_s that covers 50% of Φ. Potential problems may occur if the anatomical information of interest is not in proximity of the central axis of the image, e.g. spine in bending exams. Note that the shape is not put at the center of the envelop but rather at the center of the image because the patient could be decentralized on purpose, for example to examine only a lung or a shoulder. However, in case of full-body EOS images we will make the assumption that this ambiguity is not present, which is the case if the patient is correctly placed. We will refer to the centered square-based method with the abbreviation *sROI* and Ψ_s will denote the corresponding ROI.

4.3.2 Evaluation of the state of the art methods on EOS images

These methods are studied on a dataset of six patients that share similar characteristics and are imaged by using the same parameter setting. The morphotype is normal/big adult, i.e. body mass index (BMI) indicatively between 20 and 30. In this dataset, there are no strong deformations of anatomical structures, such as idiopathic scoliosis. Moreover, the position of the patients is well centered with respect to the field of view and eventual metallic objects are neglected. Therefore, in a first moment, we consider quite simple cases to limit assessing whether methods from the literature adapt well to EOS images or not. As for the parameters of the *hROI* method, two configurations of the threshold-based method are tested, $[0.01; 0.25]$ and $[0.35; 0.60]$, which means that the anatomical information of interest is relative to, respectively, strong and medium X-ray absorption regions. As for *sROI* method, a square of side equal to half the width of the image is defined and the final ROI Ψ_s is obtained by intersecting it with the envelop of the patient.

We assume that EI values computed at the manually defined ROIs $\mathcal{A}_r$, $r = 1, \dots 8$, correspond to the ground truth (see Table 4.1[2]). The consistency of state of the art methods is evaluated in terms of the DI between the EI values computed

[2]The images in Figures 4.4a and 4.4c are part of this dataset and correspond, respectively, to the patients number 4 and 6.

Table 4.1: *EI values computed from manually selected ROI. The images in this set of patients are used to evaluate automatic state of the art methods for ROI selection. A visual representation of these values for the patients number 4 and 6 is given in Figure 4.5.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6
$\mathcal{A}_1$	36.0	26.6	30.1	34.1	45.7	33.0
$\mathcal{A}_2$	24.1	31.3	21.0	22.9	25.0	38.3
$\mathcal{A}_3$	72.0	68.4	70.2	83.3	70.0	104.5
$\mathcal{A}_4$	14.1	17.0	13.0	14.8	15.6	25.4
$\mathcal{A}_5$	14.3	15.8	11.5	14.7	16.3	28.8
$\mathcal{A}_6$	33.7	25.7	35.0	30.8	35.5	41.5
$\mathcal{A}_7$	66.6	59.2	74.4	57.9	72.2	67.5
$\mathcal{A}_8$	78.2	69.3	70.8	71.7	75.5	77.6

at manually and automatically defined ROIs. Similarly to Equation 4.2, a DI value from the ground truth is computed as follows:

$$DI = 10 \log_{10} \left(\frac{EI}{EI_{gt}} \right) \tag{4.7}$$

where EI_{gt} is the ground truth value associated with a given anatomical ROI. The ideal result is a DI equal to zero, but a margin of error is accepted and different degrees of errors are, thus, considered: $|DI| \in [0, 0.25)$, $|DI| \in [0.25, 0.5)$, $|DI| \in [0.5, 0.75)$, $|DI| \in [0.75, 1)$ and $|DI| \in [1, +\inf)$ mean, respectively, negligible, low, medium, high and extreme errors. In order to quantify the performances in terms of precision of the EI value associated with a anatomical ROI $\mathcal{A}_r$, the root mean square error (RMSE) is computed as follows:

$$RMSE_{\mathcal{A}_r} = \sqrt{\frac{\sum_{z=1}^{N_p} (DI_{r,z})^2}{N_p}} \tag{4.8}$$

where N_p is the number of patients in a given database and $DI_{r,z}$ is the DI value obtained for the patient z and the region $\mathcal{A}_r$. Similarly, the accuracy of the EI estimation for a patient z is quantified with the following measure:

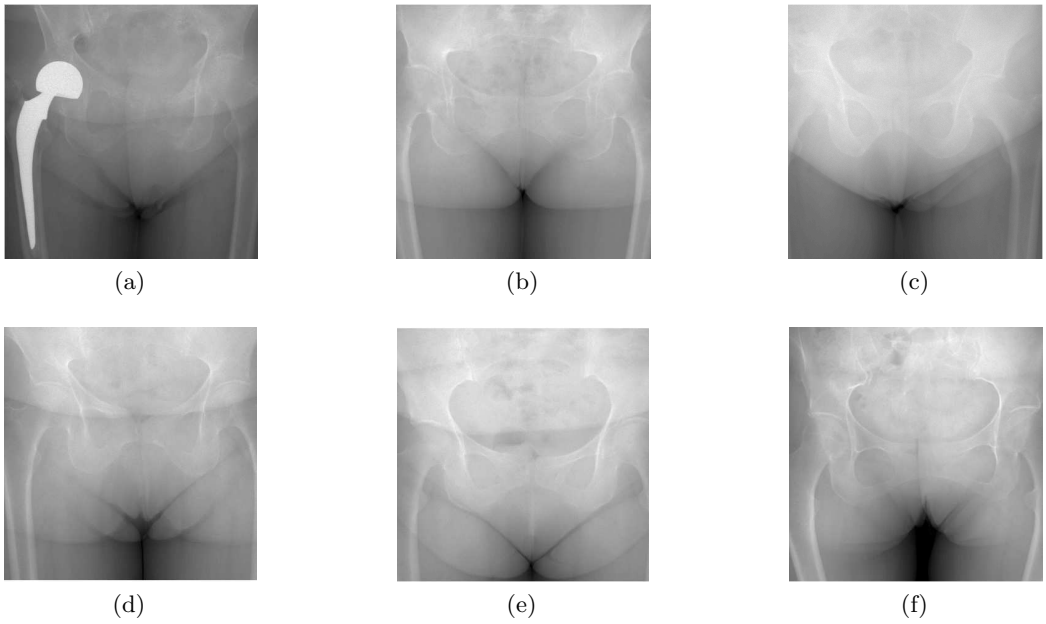
$$RMSE_{p_z} = \sqrt{\frac{\sum_{r=1}^{N_\mathcal{A}} (DI_{r,z})^2}{N_\mathcal{A}}} \tag{4.9}$$

where $N_\mathcal{A}$ is the number of anatomical ROI $\mathcal{A}_r$ in the field of view of patient z image. Moreover, we compute $\sigma(DI_r)$ that indicates if a given method provides a stable measure in a given anatomical region over various exams.

4.3.2.1 One EI value from a full-body image

In Section 4.2.1 we have pointed out that it is necessary to describe a full-body image with a set of EI values rather than a unique one. In this section we conduct an

experience that objectively justifies this assessment. Indeed, it could be discussed that a unique indicator, if consistent, could also be sufficient. In other words, if a unique value allows certifying that a given full-body image is acquired according to ALARA guidelines, the final purpose of the EI is satisfied as well. In this optic, it is observed that in a frontal acquisition, the X-rays are absorbed in the pelvic region above all. Similarly, the isotropic center of mass, which coincides with the centroid of the envelop mask, falls in the pelvic region. So, the histogram based approach with $\tau = [0.01; 0.25]$ or the centered square one could be used to estimate an indicator of full-body image quality according to the EI in the pelvis. Note that this approach does not take into account the image quality in structures that are diagnostically relevant for the full-body, but this does not prevent from obtaining an image quality indication. By using the method *hROI*, the following DI values with respect to ground truth EI values associated with the ROI $\mathcal{A}_5$ (Table 4.1) are obtained: 0.49, 0.32, 0.60, 0.66, 0.36 and -0.15. On the other hand, the ROIs Ψ_s provide the following DI values: 1.97, 0.94, 1.43, 1.61, 0.62 and 0.55. The method *sROI* gives much worse results than the histogram-based approach. This is due to slight changes in patient size or position, which make the position of the centroid uncertain with respect to the pelvis.



(a) (b) (c)

(d) (e) (f)

Figure 4.7: *ROIs Ψ_s extracted from full-body images (e.g. Figures 4.4a and 4.4c): (a) patient #1; (b) patient #2; (c) patient #3; (d) patient #4; (e) patient #5; (f) patient #6.*

This can be observed in Figure 4.7 that shows the central windows used to automatically compute the EI for the images of the database of Table 4.1. In detail, the position of the window is quite well placed for patients 5 and 6 (respectively, Figures 4.7e and 4.7f), which is coherent with the indication given by DI values. On the contrary, this does not occur in the other cases where the windows are too low and, so, the presence of soft tissues that surround the groin significantly biases the

measure.

The results improve if *hROI* is chosen, but still, the quality of the measure depends on the patient: one negligible, three low and two medium estimation errors are obtained. The approach *hROI*, with the chosen thresholds, associates the pelvis with the ROI Ψ_τ that corresponds to 24% of the most attenuated X-rays, excluding the 1% most absorbing one. This is not unconditionally true and depends on size and tissue density, which vary according to the patient. Indeed, by comparing the values in Table 4.1 and the output DI values, the hypothesized relation between strong gray levels and pelvis holds only for the patient with the lowest BMI, i.e. number 6, but is less and less true as the BMI increases. This happens because of, for patients of big size, low absorption values correspond to soft tissues in thick regions too. Therefore, a positive bias is introduced and the pelvis is judged more exposed than it actually is.

Finally, even considering quite simple cases, one single quality measure for full-body images is not robust to small variations in patient size and distribution of tissues. This is a further element that adds to the analysis conducted in Section 4.2.1 and, hence, confirms the need of computing a set of EI values on images that cover more than one anatomical structure.

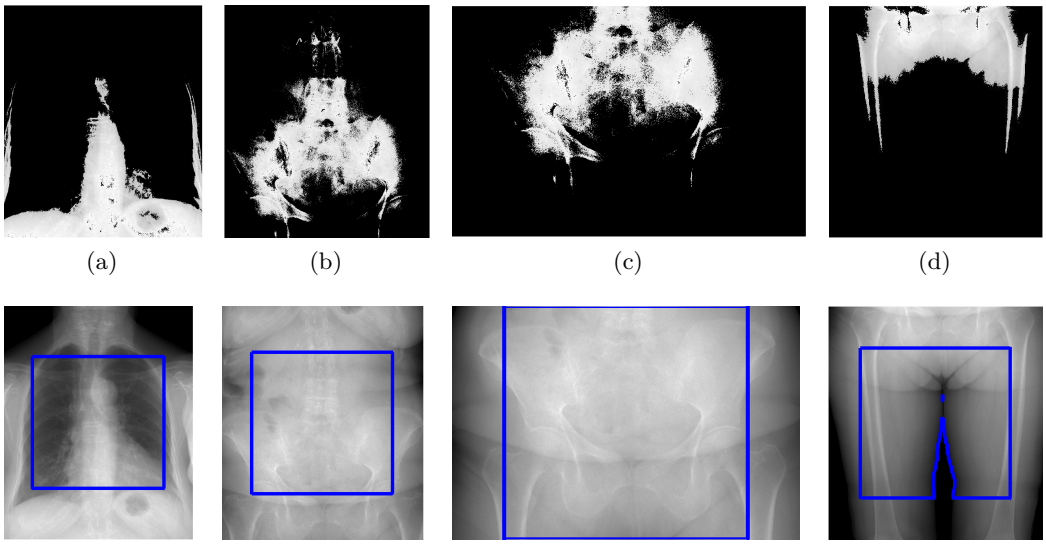


Figure 4.8: *Automatic ROI segmentation. First row (a-d): ROIs* Ψ_τ *with* $\tau = [0.01; 0.25]$*; second row (e-h): ROIs* Ψ_s*. Columns from left to right: (a,e)* $\mathcal{A}_2$*; (b,f)* $\mathcal{A}_4$*; (c,g)* $\mathcal{A}_5$*; (d,h)* $\mathcal{A}_6$*.*

4.3.2.2 Multiple exposure index in manually defined sub-windows

A possible approach to remedy to the presence of multiple structures in the field of view consists in dividing a full-body image in sub-windows where only one anatomical ROI needs to be defined. On EOS system, when the aim of the study is, for example, the lumbar spine, the upper and lower limits of the C-arm scan are defined to irradiate only the concerned region. Similarly, we can estimate from

a full-body image a set of sub-images associated with the regions $\mathcal{A}_r$. Thus, for each region $\mathcal{A}_r$ two values $[{c_r}^u; {c_r}^l]$ indicate the upper and lower values. In the analysis that follows the values $[{c_r}^u; {c_r}^l]$ have been manually defined for simplicity, but this could also be automatically achieved. For example, rough crop limits could be computed by rigidly registering an atlas of the human body on a target full-body image. Some examples of sub-images for the regions, $\mathcal{A}_2$, $\mathcal{A}_4$, $\mathcal{A}_5$ and $\mathcal{A}_6$ are given in Figure 4.8. The state of the art methods for ROI selection are tested on these sub-images. Note that the vertical limits $[{c_r}^u; {c_r}^l]$ are the same for the regions $\mathcal{A}_2$ and $\mathcal{A}_3$. In the analysis led in this section the lungs are not considered, but the parameters of the tested methods could be adapted to capture the lungs rather than the thoracic spine. For example, the thresholds should be adjusted to extract the less absorbing regions, or geometrical shapes should be placed to the side of the central axis.

Table 4.2: *DI values between the ground truth EI values (Table 4.1) and the automatically computed ones by using hROI with* $\tau = [0.01; 0.25]$*, and corresponding* $RMSE_{\mathcal{A}}$*,* $RMSE_p$ *and* $\sigma(DI)$ *measures.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6	$\text{RMSE}_{\mathcal{A}}$	σ_{DI}
$\mathcal{A}_1$	−0.57	−0.69	−0.25	−0.67	−1.11	−0.55	**0.69**	**0.26**
$\mathcal{A}_2$	−1.29	−2.40	−1.67	−1.07	−1.26	−1.09	**1.54**	**0.46**
$\mathcal{A}_4$	−0.19	−0.63	−0.73	−0.18	−0.23	−0.14	**0.42**	**0.24**
$\mathcal{A}_5$	−0.25	−0.37	−0.19	−0.21	−0.48	−0.63	**0.39**	**0.16**
$\mathcal{A}_6$	−2.92	−1.45	−3.48	−1.88	−2.49	−1.00	**2.36**	**1.14**
$\mathcal{A}_7$	−0.28	−0.49	−0.35	−0.38	−0.42	−0.27	**0.37**	**0.08**
$\mathcal{A}_8$	−0.56	−0.44	0.12	−0.85	−0.40	−0.91	**0.61**	**0.34**
RMSE_p	**1.25**	**1.15**	**1.50**	**0.93**	**1.17**	**0.74**		

Table 4.3: *DI values between the ground truth EI values (Table 4.1) and the automatically computed ones by using hROI with* $\tau = [0.35; 0.60]$*, and corresponding* $RMSE_{\mathcal{A}}$*,* $RMSE_p$ *and* $\sigma(DI)$ *measures.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6	$\text{RMSE}_{\mathcal{A}}$	σ_{DI}
$\mathcal{A}_1$	0.82	0.58	1.30	0.30	0.33	0.88	**0.78**	**0.35**
$\mathcal{A}_2$	2.36	−0.27	1.36	1.78	2.11	2.26	**1.84**	**0.90**
$\mathcal{A}_4$	0.09	0.13	−0.21	0.60	0.60	0.59	**0.43**	**0.32**
$\mathcal{A}_5$	0.88	0.73	0.50	0.90	0.50	0.67	**0.72**	**0.16**
$\mathcal{A}_6$	1.63	1.86	1.38	1.64	1.44	1.57	**1.59**	**0.16**
$\mathcal{A}_7$	1.04	0.88	0.96	0.96	0.99	0.66	**0.92**	**0.12**
$\mathcal{A}_8$	0.85	1.18	1.65	0.82	1.18	0.49	**1.09**	**0.37**
RMSE_p	**1.28**	**0.97**	**1.16**	**1.12**	**1.17**	**1.19**		

The DI values from the ground truth for the *hROI* method with $\tau = [0.01; 0.25]$ and $\tau = [0.35; 0.65]$ and for the *sROI* method are reported in Tables 4.2, 4.3 and 4.4, respectively. Moreover the performances of the methods on the whole dataset are summarized by the $\text{RMSE}_{\mathcal{A}}$, RMSE_p and $\sigma(DI)$ measures. The study of the performances is completed by a qualitative evaluation of the resulting ROIs that are reported in Figure 4.8.

Table 4.4: *DI values between the ground truth EI values (Table 4.1) and the automatically computed ones by using sROI, and corresponding* $RMSE_{\mathcal{A}}$, $RMSE_p$ *and* $\sigma(DI)$ *measures.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6	$RMSE_{\mathcal{A}}$	σ_{DI}
$\mathcal{A}_1$	0.58	0.33	0.63	−0.14	−0.16	0.22	**0.39**	**0.31**
$\mathcal{A}_2$	3.10	1.12	2.59	4.20	3.12	3.43	**3.07**	**0.94**
$\mathcal{A}_4$	0.06	−0.13	−0.21	0.03	−0.03	−0.02	**0.10**	**0.09**
$\mathcal{A}_5$	0.29	0.39	0.33	0.50	0.08	0.16	**0.32**	**0.14**
$\mathcal{A}_6$	1.45	1.86	1.67	1.87	1.71	1.62	**1.70**	**0.14**
$\mathcal{A}_7$	0.84	0.70	0.66	0.69	0.76	0.48	**0.70**	**0.11**
$\mathcal{A}_8$	0.80	1.13	1.37	0.96	1.24	0.54	**1.05**	**0.28**
$RMSE_p$	**1.39**	**0.98**	**1.33**	**1.80**	**1.46**	**1.46**		

The method hROI with thresholds $\tau = [0.01, 0.25]$ returns results that present strong negative biases with respect to the ground truth and, so, are not satisfying. Indeed, the anatomical ROIs $\{\mathcal{A}_4, \mathcal{A}_5, \mathcal{A}_7\}$, $\{\mathcal{A}_1, \mathcal{A}_8\}$ and $\{\mathcal{A}_2, \mathcal{A}_6\}$ low, medium and severe RMSE are obtained, respectively. These performances are intuitively due to the disagreement between the selected ROI and the target anatomical structures, which can be observed in the first row of images in Figure 4.8. For example, as for the region $\mathcal{A}_2$, Ψ_τ includes only the lower part of the thoracic spine and the soft tissues in the upper part of the lumbar region, whereas the upper part of the thoracic spine is not considered (Figure 4.8a). A similar effect occurs in $\mathcal{A}_6$, in which the lower part of the pelvis is retained in place of the femurs (see Figure 4.8d). The results are slightly better in the regions $\mathcal{A}_4$ and $\mathcal{A}_5$ because the signal is quite homogeneous in these regions. However, in the lumbar region, only the vertebrae L4, L5 and the sacrum are taken into account (see Figure 4.8b) and, in the pelvis, the ischium and the cotyloid cavity are excluded from the ROI Ψ_τ (see Figure 4.8c). In general, as indicated by the bottom row in Table 4.2, the EI values are not well estimated for any patient in the database. While $RMSE_{\mathcal{A}_1}$ indicates a medium error, the correspondent $\sigma(DI_1)$ value is low which means that *hROI* gives another information than the EI value extracted from the manually defined ROI, but that is consistent across different exams. On the contrary, this is not true for $\mathcal{A}_2$ and $\mathcal{A}_6$ that, hence, remain problematic regions.

A different choice of thresholds, $\tau = [0.35; 0.65]$, does not improve the results. The estimation this time tends to indicate higher exposure than the real one. An option to obtain better solutions could consist in adapting the thresholds to each region $\mathcal{A}_r$. Nevertheless, the values in Tables 4.2 and 4.3 discard this option. For instance, by analyzing the results in $\mathcal{A}_4$ for the patient number 2, higher thresholds give better results (−0.63 versus 0.13), while for the patient number 4 it is the contrary (−0.18 versus 0.60). However, as shown in Table 4.1, the ground truth values for these patients are close to each other (17.0 and 15.6). Finally, the method *hROI* is not quantitatively robust and it is impossible to solve this issue by simply changing parameters. It would be necessary to use a completely adaptable method capable of setting threshold parameters according to both patient and anatomical region of interest, which, in practice, is extremely difficult. However, with this parameter setting *hROI* returns very low $\sigma(DI_r)$ values, except for $r = 2$, and is

then preferred. In any case, this option is only sub-optimal as the values $\text{RMSE}_{\mathcal{A}_r}$ and RMSE_{p_z} indicate strong errors and the region of the thoracic spine is not well described at all.

The method *sROI* (see Table 4.4) provides good results in $\mathcal{A}_4$. The ROI covers a wider region than the lumbar spine (see Figure 4.8f), but the indicator is good because the bone tissues are present in higher amount and, thus, the median tendency of the ROI matches well with the ground truth EI. On the contrary, in $\mathcal{A}_2$ the ROI contains both pulmonary tissues and bones of thoracic spine (see Figure 4.8e), by hence causing severe errors. In the pelvis, the ROI presents two significant portions of different tissues too (see Figure 4.8g), but the errors are lower compared to $\mathcal{A}_2$ because the difference between associated signals is lower. In these two last cases, better results could be achieved by using other shapes than the square, but it is not trivial to define which ones given the different sizes and shapes of the structures among patients. Finally, the most severe errors occur in the legs because the information of interest is far from the central axis (see Figure 4.8h). Nevertheless, the $\sigma(DI_r)$ values, with $r = 6, 7, 8$, are very low, i.e. the associated EI values are consistent. On the other hand, this is not the case in region $\mathcal{A}_2$ and, hence, *sROI* is sub-optimal. Moreover, we expect that the performances of this method would decrease in presence of deformed structures (e.g. idiopathic scoliosis) that are not considered in this dataset.

In conclusion, even considering quite simple cases, the methods suggested in the norm IEC 62494-1 (2008) do not provide robust EI values. The EI values should be computed at anatomically meaningful ROIs. However, the method *hROI* fails in respecting this idea because the gray level is not a strong feature for describing anatomical structures in EOS and, in general, digital radiography images. Indeed, the superposition of different tissues causes a change of intensity levels inside the same anatomical structure. As for the method *sROI*, it is assumed that the discriminant information relative to exposure is placed around the centroid of the image. Nevertheless, this hypothesis is not always true as it depends on the type of anatomical region. Simple prior geometrical shapes are not adapted to follow the variability in shape and size of the structures.

Given theses issues, we think that unexpected conclusions from the analysis of the EI values in clinical routine may be due to the technique used to define the ROI. For example, Mothiram et al. (2013) have observed that the EI is affected by variations in gender or time of exposure, which makes it difficult to interpret EI values. The experiences conducted in this section confirm this issue.

As a consequence, in the following section we propose an alternative way to address the problem with the objective of providing consistent EI estimations despite the changes that normally occur in clinical examinations.

4.4 Anatomical structure detection applied to the exposure index

The low accuracy on EI value estimated at ROIs defined with state of the art methods are mainly due to the inconsistency between the selected regions and the actual position of anatomical structures. Therefore, the positions of these structures on a full-body image should be detected to solve the issue.

4.4.1 Bounding box based approach

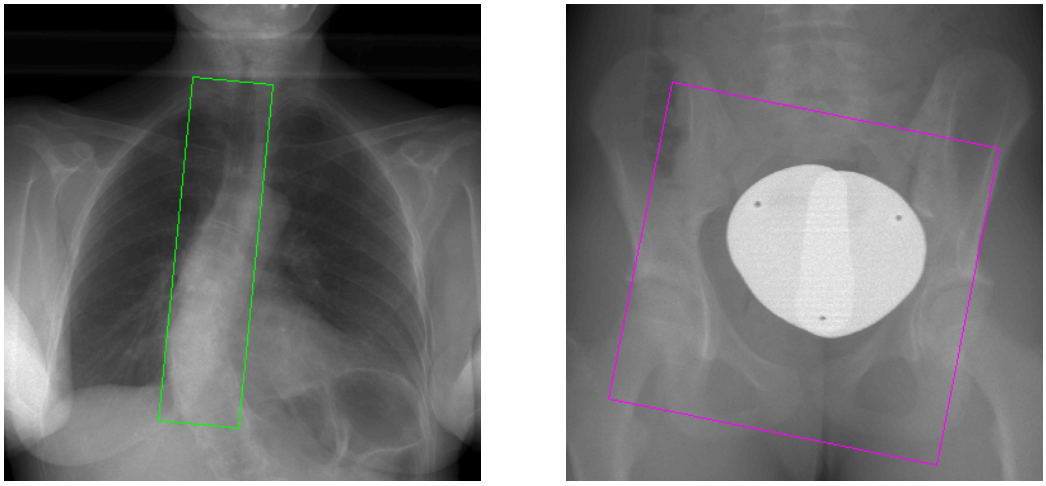
Anatomical structures are detected in medical images by placing bounding boxes around the target objects. Intuitively, these bounding boxes could then be used as ROIs for computing EI values. This aspect has been widely addressed in the recent literature on medical applications and a synthetic overview of these works is given in Section 5.1. First, we limit ourselves to evaluate if this detection technique can be exploited for the EI application. When 2D images are processed, the bounding box is an oriented rectangle that depends on five parameters: the coordinates of the upper left corner, the width, the height and the orientation. Given a manually segmented structure (e.g. Figure 4.4a), the corresponding ideal bounding box is defined according to the set of parameters that allows completely containing the structure while minimizing the area of the rectangle. We will refer to this method with the acronym *bbDet* as based on the detection of bounding boxes.

Table 4.5: *DI values between the ground truth EI values (Table 4.1) and the automatically computed ones by using bbDet, and corresponding* $RMSE_{\mathcal{A}}$*,* $RMSE_p$ *and* $\sigma(DI)$ *measures.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6	$RMSE_{\mathcal{A}}$	σ_{DI}
$\mathcal{A}_1$	−0.08	−0.32	0.21	−0.20	−0.36	−0.15	**0.24**	**0.19**
$\mathcal{A}_2$	0.22	0.12	0.79	0.40	0.85	0.36	**0.53**	**0.27**
$\mathcal{A}_3$	−0.05	−0.41	−0.15	−0.21	−0.16	−0.17	**0.22**	**0.11**
$\mathcal{A}_4$	0.06	0.08	0.04	0.09	0.03	0.03	**0.06**	**0.02**
$\mathcal{A}_5$	0.08	0.08	0.11	0.20	0.05	−0.11	**0.12**	**0.09**
$\mathcal{A}_6$	0.11	0.33	0.29	0.23	0.13	0.17	**0.22**	**0.08**
$\mathcal{A}_7$	0.27	0.25	0.08	0.19	0.12	0.16	**0.19**	**0.07**
$\mathcal{A}_8$	0.53	0.69	0.48	0.84	0.54	0.54	**0.62**	**0.12**
$RMSE_p$	**0.23**	**0.34**	**0.36**	**0.37**	**0.39**	**0.26**		

The adaptability to the application of interest is evaluated according to the similarity between EI values computed in these rectangles and the ground truth ones (Table 4.1). The resulting DI values are reported in Table 4.5 and, in general, are quite low, which shows that the detection of anatomical structures can be used in the EI algorithm. Indeed, the errors in the anatomical regions are all negligible, except for the thoracic spine ($\mathcal{A}_2$) and for the tibia ($\mathcal{A}_8$), where the errors are medium. In the thoracic spine, the rectangle takes into account some of the surrounding pulmonary tissues, by hence increasing the EI value. In region $\mathcal{A}_8$, the signal is not very homogeneous, and, hence, the imperfect correspondence between the rectangle and the tibia entails a slight increase of the EI value. Furthermore, it is worth noting that all the values $\sigma(DI_r)$, $r = 1, \ldots, 8$ are very low, which indicates that EI values estimated with *bbDet* are repeatable.

Finally, the EI values computed at oriented rectangles that localize the structures of interest give the same information as manually segmented regions and are robust across different exams. Nevertheless, intuitively, since there are only 5 degrees of freedom, the oriented rectangles could be not adapted to certain cases. For example, it could be difficult to follow the deformation of the spine due to idiopathic scoliosis (see Figure 4.9a). In an another example, the bounding box could be sufficient to capture a given anatomical structure, but some gray levels at pixels belonging to



(a) (b)

Figure 4.9: *Possible problematic cases for ROI selection by using bounding boxes: (a) scoliosis in the thoracic spine; (b) Metallic object. The case (a) entails a significant increase in EI value, as the ROI fails in following the deformation of the structure and, thus, takes into account some low density tissues. The case (b) does not practically cause an error in EI estimation, but it would be better to avoid taking into account the gray levels associated with the metallic object.*

the ROI could be outliers as associated with a metallic object (see Figure 4.9b). Quantitatively, for the cases presented in Figure 4.9, the ground truth EI values are, respectively, equal to 24.9 and 22.1, and those computed in the oriented rectangles are equal to 31.0 and 22.5, which implies that the DI values are equal to 0.97 and -0.07. Therefore, because of extreme bending of the spine, non relevant portions of tissues are included in the ROI and this introduces a bias in the EI measure. In the other case, the metallic object covers a lower percentage of the ROI than the pelvis tissues that are then represented by the central tendency of the gray level distribution. However, it would be preferable to avoid to take into account pixels that correspond to metallic objects because even slight errors on the position or the size of the box could cause a reduction of the associated EI values.

4.4.2 Proposed landmark based approach

Given the highlighted potential issues of using oriented rectangles as ROIs, a new formulation of the problem is proposed in this section. The objective is to better follow the shape of anatomical structures while keeping low the computational load and limiting the impact of outliers. We suggest to associate with an anatomical ROI a cluster of points that belong to the structure and, then, estimate the EI from measures computed in local patches P centered at these points. Therefore, each anatomical ROI is described by a cluster of points located inside the structure. Two examples of suitable solutions are shown in Figures 4.4b and 4.4d. Note that the number of degrees of freedom for describing a structure is equal to the number of landmarks that compose the cluster. As a consequence, the shapes of anatomical structures can be more precisely described. Here, we only focus on the definition

of EI values by means of the proposed landmark based approach and on showing its adaptability to the problem. On the other hand, the proposed method for the detection and recognition of the landmarks is discussed in Chapter 5.

A local EI measure can be computed considering the distribution of gray levels at the pixels $x_i \in P_i$. We use circular patches of radius equal to 128 pixels, i.e. about 23 mm, that on an adult patient approximately cover the area of a vertebra. The circle has to be large enough to avoid the measure to be excessively affected by noise and, at the same time, small enough to guarantee the gray level distribution to be approximately mono-modal. In this way the central tendency of a distribution will be a more representative descriptor of the actual amount of signal and, hence, of the EI value in a region. Therefore, local measures are much less biased by presence of outliers than global or regional ones.

Formally, given a set of landmarks $l_j \in \mathcal{L}_r$, where $\mathcal{L}_r$ is the cluster associated with the ROI $\mathcal{A}_r$, the corresponding values of interest $v(l_j)$ and, then, the local EI values $e(l_j)$ are computed. The EI value at the ROI $\mathcal{A}_r$ is, finally, computed by using the following weighted sum of local EI values:

$$EI_r = \frac{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j) e(l_j)}{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j)} \tag{4.10}$$

where the weights $\omega(l_j)$ assess the accuracy of the measure provided by the landmark $l_j \in \mathcal{L}_r$ by giving higher importance to $e(l_j)$ values that are computed from homogeneous gray level distributions because the corresponding VOI $v(l_j)$, computed as the median, is a good descriptor of the signal level. It is worth noting that a landmark that is at the center of a metallic object has a high weight, as the signal is constant, but its local EI value is an outlier. In this case by analyzing the set of local EI estimates, it is possible to easily recognize these cases and reject them by setting their weights to zero, as we will discuss in Section 5.7.3. Formally, we use the entropy to define the weights $\omega(l_j)$. The entropy of a random variable X is defined as follows:

$$H(X) = \sum_{x \in S_X} -P(x) log(P(x))$$

where S_X is the set of elements of the discrete distribution that are taken into account to compute the entropy and, here, corresponds to the gray levels at the pixels $x_i \in P_j$ where P_j is the patch centered at the landmark l_j. Then, the weights are defined by using the following exponential function:

$$\omega(l_j) = \exp\left(-\frac{H(l_j)}{\alpha_H}\right) \tag{4.11}$$

where $H(l_j)$ is the entropy computed at the landmark $l_j \in \mathcal{L}_\nabla$ and α_H is a constant smoothing parameter set to 2 in our tests. We will refer to this method with the acronym *lDet* as based on the detection of landmarks.

The DI values between the ground truth EI in Table 4.1 and the EI computed with the proposed landmark-based method are reported in Table 4.6. The results are comparable to those obtained with ideal oriented rectangles (Table 4.5). However, there is an improvement in region $\mathcal{A}_2$ that is justified by a better capacity in following

Table 4.6: *DI values between the ground truth EI values (Table 4.1) and the automatically computed ones by using lDet, and corresponding $RMSE_{\mathcal{A}}$, $RMSE_p$ and $\sigma(DI)$ measures.*

	Pat_1	Pat_2	Pat_3	Pat_4	Pat_5	Pat_6	$RMSE_{\mathcal{A}}$	σ_{DI}
$\mathcal{A}_1$	−0.09	−0.10	−0.06	−0.18	−0.47	0.18	**0.23**	**0.19**
$\mathcal{A}_2$	−0.14	0.35	0.63	0.33	0.09	0.30	**0.35**	**0.24**
$\mathcal{A}_3$	0.05	−0.07	0.09	−0.04	0.17	0.25	**0.13**	**0.11**
$\mathcal{A}_4$	−0.01	0.03	0.01	0.09	0.06	−0.16	**0.08**	**0.08**
$\mathcal{A}_5$	−0.02	−0.28	−0.08	0.03	−0.25	−0.49	**0.26**	**0.18**
$\mathcal{A}_6$	0.80	0.46	0.57	0.53	0.42	0.41	**0.55**	**0.13**
$\mathcal{A}_7$	0.19	0.04	−0.11	−0.02	−0.22	−0.05	**0.13**	**0.13**
$\mathcal{A}_8$	0.25	0.53	0.57	0.45	0.33	0.23	**0.42**	**0.13**
$RMSE_p$	**0.31**	**0.30**	**0.37**	**0.28**	**0.29**	**0.29**		

the bending of the spine. This is confirmed by repeating the experience previously conducted on the image that shows a case of scoliosis (Figure 4.9a), but this time with the proposed method. The estimated EI is 27.32 and, hence, a DI equal to 0.42, which is still not optimal but coherent with the results in Table 4.6. We note a reduction of the quality of estimation in region $\mathcal{A}_6$ that occurs because some soft tissues on the side of the femur are considered in the local measures too. However, as for the method *bbDet*, all the values $\sigma(DI_r)$, $r = 1, \ldots, 8$, are very low which indicates that *lDet* is a robust method as well.

4.4.3 Robustness to localization errors

We have proved that the detection-based methods *bbDet* and *lDet* perform better than the ROI-based methods *hROI* and *sROI* in terms of congruence with manually segmented regions and robustness of the estimated EI values. The landmark-based approach is more robust to strong bending of anatomical structures, but this is not sufficient to opt for this approach. An intuitive advantage of computing the EI from local measures is represented by the redundancy of this technique. Indeed, landmarks that belong to the same cluster tend most likely to provide similar EI values and this limits the impact of outliers.

So far, only ideal bounding boxes and landmarks have been considered and it remains to quantify how eventual errors on the bounding boxes or landmarks positions influence the EI estimates. Therefore, we measure how the EI values change according to simulated localization errors of 5, 10, 15, 20, 25 and 30 mm. For each simulation, the bounding boxes or the landmarks are misplaced with respect to their ground truth positions. The errors for bounding boxes could be induced by wrong estimations of the position, orientation and size parameters. At first, only errors on the position are taken into account, but changes in orientation or size could probably cause EI estimation errors as well. As for the landmarks, we evaluate how the accuracy decreases as the positions of all or part of the landmarks is incorrectly estimated.

For each degree of error in distance, 100 simulations are executed, and the mean and standard deviation of the EI values are computed. Then, the DI values are computed with respect to the ideal EI measures, i.e. those computed from not displaced

Table 4.7: *Effect on EI measure induced by simulated errors on the localization of the bounding boxes. The errors are measured as DI of the mean EI with respect to the ground truth and the standard deviation EI value over 100 simulations for each degree of localization error. The results in this table concern the anatomical regions of the image in Figure 4.4a.*

Error	Errors EI	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$
5 mm	mean DI	0.09	0.25	-0.27	0.02	0.09	0.17	0.15	0.10
	std EI	0.09	0.32	0.37	0.01	0.05	0.07	0.23	0.07
10 mm	mean DI	0.18	0.51	-0.54	0.04	0.17	0.35	0.29	0.21
	std EI	0.15	0.48	0.61	0.03	0.14	0.56	0.52	0.22
15 mm	mean DI	0.27	0.76	-0.89	0.07	0.26	0.54	0.43	0.32
	std EI	0.25	0.47	1.55	0.02	0.19	0.10	1.27	0.18
20 mm	mean DI	0.38	1.01	-1.25	0.09	0.37	0.71	0.61	0.45
	std EI	0.25	0.98	1.55	0.11	0.13	0.29	0.63	1.14
25 mm	mean DI	0.50	1.23	-1.68	0.14	0.45	0.83	0.77	0.58
	std EI	0.58	1.64	1.30	0.04	0.29	0.33	0.78	1.97
30 mm	mean DI	0.67	1.60	-2.09	0.16	0.54	1.00	0.98	0.83
	std EI	0.73	2.05	0.90	0.11	0.34	0.47	0.76	2.01

landmarks or bounding boxes. The results from these experiences are reported in Tables 4.7, 4.8, 4.9 and 4.10 for bounding boxes and for 100%, 50% and 25% misplaced landmarks, respectively. The values in these tables are computed from the patient in Figure 4.4a, but the considerations that follow are valid for all the patients. As expected, the $|DI|$ value increases with higher errors in mm. However, the trends are different according to the method and anatomical ROI too. We summarize the results reported in the tables by indicating for each method and $\mathcal{A}_r$ at which distance in millimeters the DI values get higher than 0.25, i.e. up to which distance the errors can be considered negligible:

- Bounding boxes (Table 4.7): $\mathcal{A}_1 = 15$ mm; $\mathcal{A}_2 = 5$ mm; $\mathcal{A}_3 = 5$ mm; $\mathcal{A}_4 > 30$ mm; $\mathcal{A}_5 = 15$ mm; $\mathcal{A}_6 = 10$ mm; $\mathcal{A}_7 = 10$ mm; $\mathcal{A}_8 = 15$ mm.
- 100% misplaced landmarks (Table 4.8): $\mathcal{A}_1 > 30$ mm; $\mathcal{A}_2 = 15$ mm; $\mathcal{A}_3 = 15$ mm; $\mathcal{A}_4 > 30$ mm; $\mathcal{A}_5 > 30$ mm; $\mathcal{A}_6 = 20$ mm; $\mathcal{A}_7 = 20$ mm; $\mathcal{A}_8 = 15$ mm.
- 50% misplaced landmarks (Table 4.9): $\mathcal{A}_1 > 30$ mm; $\mathcal{A}_2 = 25$ mm; $\mathcal{A}_3 = 20$ mm; $\mathcal{A}_4 > 30$ mm; $\mathcal{A}_5 > 30$ mm; $\mathcal{A}_6 > 30$ mm; $\mathcal{A}_7 > 30$ mm; $\mathcal{A}_8 = 20$ mm.
- 25% misplaced landmarks (Table 4.10): $\mathcal{A}_1 - \mathcal{A}_2 > 30$ mm, $\mathcal{A}_3 = 25$, $\mathcal{A}_4 - \mathcal{A}_8 > 30$ mm.

In general, the approach *lDet* is more robust than *bbDet*. This is due to the redundancy that is introduced in the proposed landmark-based method: if the whole region is misplaced, the central tendency of the gray level distribution will probably be biased as well, whereas measures from local patches only partially count in the final estimation. It is worth noting that if all the landmarks are wrong the average DI values with respect to the EI values computed from correctly placed landmarks

Table 4.8: *Effect on EI measure induced by simulated errors on the localization of all the landmarks in the cluster. The errors are measured as DI of the mean EI with respect to the ground truth and the standard deviation EI value over 100 simulations for each degree of localization error. The results in this table concern the anatomical regions of the image in Figure 4.4a.*

Error	Errors EI	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$
5 mm	mean DI	0.01	0.07	-0.04	-0.01	0.00	0.00	0.02	0.00
	std EI	0.05	1.20	0.79	0.06	0.07	0.23	0.38	0.20
10 mm	mean DI	0.01	0.17	-0.13	0.01	0.01	0.04	0.07	0.05
	std EI	0.14	2.30	1.77	0.10	0.15	0.52	0.72	1.42
15 mm	mean DI	0.02	0.34	-0.29	0.03	0.03	0.15	0.11	0.28
	std EI	0.35	3.37	3.28	0.14	0.25	1.05	1.36	3.96
20 mm	mean DI	0.02	0.73	-0.63	0.04	0.07	0.26	0.28	0.48
	std EI	0.57	6.05	5.62	0.19	0.33	1.43	2.63	5.96
25 mm	mean DI	0.06	1.20	-1.07	0.06	0.11	0.27	0.54	0.67
	std EI	0.71	9.06	8.17	0.24	0.40	1.52	4.24	8.07
30 mm	mean DI	0.13	1.78	-1.47	0.08	0.18	0.30	0.78	0.90
	std EI	1.10	11.90	10.34	0.30	0.45	1.78	7.66	11.67

is lower than for bounding boxes, while the standard deviation of EI values obtained through all the simulations is higher. Therefore, in some cases, the measures could be significantly wrong. However, if part of the landmarks is correctly located, the EI estimates will be exact within a wide range of spatial error. Quantitatively, all the EI values are correctly estimated if, for example, at least 75% (50%) of landmarks are correctly located within 20 (15) mm, whereas to obtain the same results with *bbDet* the spatial error has to be inferior to 5 mm. Given these results, we consider unnecessary to simulate errors on the orientation or the size of the bounding boxes that probably would lead to similar conclusions.

The estimation of EI values is more affected by localization errors in some regions than in others, which is coherent with the results presented in Section 4.4.1 and 4.4.2. The EI values are quite correct in regions $\mathcal{A}_{4-5}$ regardless the detection method and the degree of localization error and, therefore, it will be easy to provide a correct exposure here. On the other hand, in the chest $\mathcal{A}_{2-3}$ the estimation is less stable given the proximity of structures at low and high density (i.e. the lungs and the thoracic spine) and, hence, the proposed landmark-based method is even more interesting.

As for the computational load, *bbDet* requires to detect 12 rectangles on a full-body image, i.e. one for each not connected regions (see for example Figure 4.4a), and thus 60 parameters. The approach *lDet* demands to estimate the coordinates of 116 points[3], i.e. 232 parameters. However, the y-coordinates of the landmarks are constrained to be equal to a multiple of 128. In this way the number of parameters to be estimated is reduced to the number of landmarks, while fully representing the structures in the full-body (see Figure 4.4b). Moreover, note that the search for landmarks can be highly optimized because a lot of them are aligned and, so, some

[3]The number of landmarks could change according to the morphotype of the patient.

Table 4.9: *Effect on EI measure induced by simulated errors on the localization of half the landmarks in the cluster. The errors are measured as DI of the mean EI with respect to the ground truth and the standard deviation EI value over 100 simulations for each degree of localization error. The results in this table concern the anatomical regions of the image in Figure 4.4a.*

Error	Errors EI	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$
5 mm	mean DI	0.00	0.00	-0.03	-0.01	0.00	0.00	0.00	0.00
	std EI	0.04	0.53	0.35	0.03	0.04	0.13	0.15	0.06
10 mm	mean DI	-0.01	0.01	-0.07	0.00	0.01	0.01	0.02	0.03
	std EI	0.07	0.86	0.72	0.06	0.09	0.34	0.31	0.60
15 mm	mean DI	-0.02	0.11	-0.14	0.00	0.02	0.07	0.03	0.11
	std EI	0.23	1.23	1.13	0.08	0.14	0.58	0.50	1.70
20 mm	mean DI	-0.04	0.15	-0.28	0.02	0.04	0.11	0.08	0.22
	std EI	0.29	1.83	2.01	0.08	0.16	0.71	0.88	2.39
25 mm	mean DI	-0.04	0.26	-0.41	0.02	0.06	0.12	0.12	0.29
	std EI	0.45	2.32	3.51	0.11	0.17	0.85	1.26	3.02
30 mm	mean DI	-0.02	0.57	-0.67	0.04	0.09	0.14	0.21	0.42
	std EI	0.58	2.24	5.09	0.13	0.21	0.97	1.52	3.82

positions can be inferred by interpolating nearby ones. For example, this is clearly the case for the landmarks in the spine and in the legs.

Finally, the proposed method offers a robust solution to obtain EI measures by aggregating local estimates at points that belong to clusters associated with anatomical ROIs. This method can be easily embedded into a semi-automatic algorithm for computing the EI values after an acquisition: the radiographer would have to trace one or a few lines roughly in correspondence with the anatomical ROI and, then, points sampled on the lines would be used as input to the landmark-based method. Note that this would require to store both the image used to compute the EI (see Figure 4.2) and the output display image on the PACS. Nevertheless, a semi-automatic algorithm limits the solution to the EI as defined in the norm IEC 62494-1 (2008) and, as explained in Section 4.1.2, it should not rely on user-interaction if we want it to be used in clinical routine. Moreover, a semi-automatic approach does not apply to an exposure management system that, according to the discussion in Section 4.2.2, needs to be managed in a completely automatic and fast way, otherwise the workflow would be excessively slowed down. In order to address these questions, we need to define an algorithm for automatic detection and recognition of the clusters of landmarks associated with the ROIs. Given the complexity of the task, this aspect is exhaustively described in a dedicated chapter, that is Chapter 5.

4.5 Conclusion

In this chapter, the concept of exposure index and its potential benefits have been presented. We have suggested to use the anatomical SNR that relies on the noise level estimation with the percentile method as an alternative to the EI as defined in the norm IEC 62494-1 (2008). This allows us to obtain the same information

Table 4.10: *Effect on EI measure induced by simulated errors on the localization of 25% landmarks in the cluster. The errors are measured as DI of the mean EI with respect to the ground truth and the standard deviation EI value over 100 simulations for each degree of localization error. The results in this table concern the anatomical regions of the image in Figure 4.4a.*

Error	Errors EI	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$
5 mm	mean DI	0.00	0.01	-0.02	-0.01	0.00	0.00	0.01	0.00
	std EI	0.13	0.35	0.22	0.03	0.02	0.08	0.06	0.04
10 mm	mean DI	0.00	0.03	-0.05	-0.01	0.00	0.01	0.02	0.02
	std EI	0.23	0.54	0.39	0.05	0.04	0.25	0.17	0.25
15 mm	mean DI	-0.02	0.04	-0.09	-0.01	0.01	0.04	0.03	0.06
	std EI	0.24	0.77	0.58	0.07	0.06	0.39	0.27	0.91
20 mm	mean DI	-0.04	0.08	-0.16	-0.01	0.01	0.06	0.07	0.12
	std EI	0.29	0.95	0.93	0.09	0.07	0.51	0.52	1.30
25 mm	mean DI	-0.03	0.13	-0.26	-0.01	0.02	0.07	0.10	0.15
	std EI	0.27	1.04	1.46	0.08	0.08	0.58	0.70	1.36
30 mm	mean DI	-0.03	0.19	-0.38	0.00	0.03	0.05	0.13	0.21
	std EI	0.33	1.38	1.84	0.10	0.10	0.73	0.78	1.78

as the one provided by the EI value while not requiring to model the detector and, besides, to consider the actual noise level in the image $\mathbf{u}_r$ from which the EI is estimated. A study of the algorithm has allowed us to point out that the classical methods used to define the ROI do not return stable EI values, as intensity levels vary a lot because of tissue superposition and the eventual inconsistency between the positions of structures of interest and the center of the image. Then, we have shown that more robust and repeatable EI measures can be obtained by relying on anatomical structure detection. Besides, this allows managing the presence of different anatomical structures in the field of view while being efficient. Moreover, we have proposed a landmark-based approach that, by exploiting redundancy of local estimates, is more robust to detection errors than the bounding box-based method.

Our method could be straightly embedded in a semi-automatic algorithm. Nevertheless, this does not empirically apply to clinical routine. Besides, one of the goal of this thesis is to define an exposure management service, which can rely on the proposed technique for computing EI values, but requires to automatically define the clusters of landmarks. This is addressed in the following chapter.

Chapter 5

Automatic landmark detection to estimate exposure indices

According to the discussion on EI value estimation on clinical images introduced in Chapter 4, we propose in this chapter a new method for the automatic detection and recognition of the landmark clusters on a full-body EOS image. The method is evaluated according to the accuracy of the EI values computed at the automatically detected landmarks in different anatomical ROIs, and for various clinical cases.

Organization of the chapter - Section 5.1 gives a synthetic overview of the state of the art methods that address the problem of anatomical structure detection. Section 5.2 presents the notations used in this chapter. Section 5.3 introduces the spatial relations between some control points that are associated with the anatomical clusters and are used to initialize the method. Section 5.4 presents how to define salient points that allow for a sparse search of the solution. Sections 5.5 and 5.6 specify how to detect and recognize the landmarks on frontal and lateral view acquisitions, respectively. Section 5.7 evaluates the proposed approach. Section 5.8 summarizes the contributions of this chapter.

5.1 State of the art methods for detection of anatomical structures

We introduce the subject of this chapter by giving a rapid overview of the methods that address object detection in medical applications. In the literature, other medical imaging modalities than digital radiography are considered, i.e. mainly CT and MR scans. Our method is not a simple extension to our images of existent approaches, but rather an original one. Nevertheless, this survey allows highlighting the main and common ideas of anatomical structure localization techniques.

5.1.1 Atlas registration-based techniques

An atlas is a model that embeds statistical information on anatomical structures. An atlas is computed from one or more annotated samples and the resulting template can then be registered on a new sample to segment anatomical structures. In details, the atlas is geometrically transformed to maximize the correspondence between the source and the target. A registration that uses only linear geomet-

rical transformations is called rigid, whereas, if a spatially varying model is used, the technique is called deformable registration. The non linear registration requires higher computational load than the linear one, but it also increases the quality of the registration. Please refer to Sotiras et al. (2013) and references therein for a complete overview on deformable registration methods applied to medical imaging.

An atlas associated with manual annotations on a unique sample does not provide a robust model as it does not capture singular features on the target image. For example, consider the problem of registering an atlas computed from a healthy patient on a pathological observed patient. Therefore, it is preferable to estimate a probabilistic atlas from multiple annotated images. The resulting model is a probability map that takes into account all the possible deformations of a structure that are represented in the manually annotated dataset. Probabilistic atlases have been widely used in the literature. For example, Mazziotta et al. (1995) have used an atlas to get a probability map associated with brain structures and Fenchel et al. (2008) to label anatomical structures in MR full-body images. However, the price to pay for a more general model is a lower precision and, generally it is difficult to achieve robust inter-patient registration by relying only on probabilistic atlases (Criminisi et al., 2013).

Methods based on multi-atlases increase the performances. In this type of approach, the target is registered on all the atlases in the database and the best match is retained as solution, i.e. the model that allows maximizing the correlation with the smallest amount of deformations. For example, multi-atlases have allowed a considerable improvement in quality of brain images segmentation compared to single-atlas method (Heckemann et al., 2006). Recently, Wolz et al. (2013) have exploited it to generate specific-patient atlases to simultaneously segment multiple organs in abdomen CT scans. Nevertheless, the main drawback of multi-atlases based registration method is the high computational load. Besides, pathology-related cases such as missing organs, presence of a tumor or strongly deformed structures, significantly reduce the accuracy of the registration. Therefore, other techniques are analyzed.

5.1.2 Classification

Anatomical structures can be detected by relying on trained binary classifiers: each pixel/voxel is a possible position associated with the object to detect and, according to the exploited features and the learned model, a true or false label is assigned to the pixel depending on whether it matches with the anatomical structure or not. This procedure requires to test every pixel in the image, which is inefficient, especially if multiple objects have to be detected. Therefore, extensive efforts have been dedicated to optimize this search by means of smart hierarchic algorithms. Zhan et al. (2008) have proposed to hierarchically drive the multi-organ search by selecting the following step that maximizes the expected information gain, which has a double advantage. First, the most informative landmark is probably also the easiest one to detect. Secondly, considering the pairwise spatial relations between the objects, the procedure maximally reduces the search space for the other organs. Liu et al. (2010) have used a sub-modular approach to define the optimal sequence order of landmarks detection by using a Greedy procedure that reduces the overall cost. Basically, at each step the landmark with minimal associated search space is selected according to the already detected objects. The effectiveness of these

methods is strictly related to the application context: the anatomical structures reciprocally constraint the search space in a relatively easy way to predict because their number of instances is known and the spatial relations are stable.

Random forests applied to medical imaging (Criminisi et al., 2009) directly learn the global pairwise relations between objects rather than modeling them at algorithmic level. This approach is probabilistic and naturally multi-class oriented, i.e. it is not necessary to learn a classifier for each organ. However, it is not trivial to compare the performances between hierarchic local classifiers and multi-instance ones: Liu et al. (2010) claim lower errors and execution time than Criminisi et al. (2009), but this comparison is not entirely fair because the dataset is not the same.

In a recent work, Gao et al. (2014) have proposed to address the problem of the extensive search by encoding the image in a sparse representation. In details, a set of salient points is detected by using differences of Gaussian functions, like in the SIFT algorithm by Lowe (2004), and geometric descriptors are then associated just with these points to train a random forest classifier that is used to detect the wrist in CT scan images.

Alternatively, in some works pixels are mapped to the organ locations, i.e. the classifier estimates parameters of oriented rectangles that surround the organ of interest. In this case, in addition to the position, the size and orientation of the organ are computed and salient landmarks can be extracted as a by-product (Criminisi et al., 2013). A very popular approach is the Marginal Space Learning (MSL) algorithm (Zheng et al., 2008, 2009). The estimation of the position, orientation and scale parameters is conducted in a low dimensional space by using in cascade three classifiers trained to learn overall position, then position and orientation, and finally position, orientation and scale. Basically, the parameter setting is limited by the solution at the previous step, which allows reducing the computational load of about an order of magnitude and getting a more precise estimation. An additional optimization is achieved by considering that solutions are constrained because a given organ has a limited number of possible scales, orientations and positions (Zheng et al., 2009). Nevertheless, as pointed out by Criminisi et al. (2013), this approach presents some scalability issues as for each organ three classifiers need to be trained. Therefore, regression-based approaches have recently emerged as a preferred approach to address the organ detection and volume estimation.

5.1.3 Regression

A regressor estimates a continuous value rather than a discrete one. Each pixel/voxel votes for its relative position with respect to a target landmark or bounding box and the pairwise spatial relations between target objects are automatically modeled by merging the multiple votes. The estimates are associated with a sparse set of pixels, which avoids exhaustively searching for the solution and, hence, makes these methods efficient (Zhou et al., 2007). Nevertheless, the regressor has to associate a continuous value or vector to each sample rather than a discrete one and, then, it is more challenging to learn a robust model. In practice, this complexity can be managed only if the geometrical relations between objects in the scene are coherent in many different images of the same scene. Since this is often the case in medical imaging, regressor learning has been proved to be particularly adapted to this context (Zhou, 2014). Essentially, the main difference between existing ap-

proaches in the literature relies in the way the regressor is formulated. Zhou (2010) has used a boosting regressor to detect and segment the left ventricle endocardial borders in ultrasound images. Criminisi et al. (2013) have identified the positions of multiple organs in CT scans by exploiting the regression forests, i.e. the formulation of the popular random forest classifier for regression. A comparison between boosting, forests and cascade is out of the scope of the works presented in this thesis. Please refer to Yin et al. (2007) for further information about this aspect.

The visual features used to train the regressor critically influence the final performances. Criminisi et al. (2013) have used mean gray level intensities over cubical regions of the volume. Haar-like features (Viola and Jones, 2001) have been used in other works (Zhou, 2010; Lindner et al., 2012, 2013) because they can be efficiently computed by using integral images and have been proved to be effective in many applications. However, these features basically compute differences of gray levels, which, depending on the medical modality, may not be reliable descriptors. For example, in MR images the intensity levels only give relative values and suffer from field inhomogeneity. Therefore, Pauly et al. (2011) have used 3D local binary patterns (Ojala et al., 1996) along with gray levels to train a regression forest in order to complete the model with textural information. The definition of more informative and fast-to-compute features remains a challenging research matter far from being solved (Zhou, 2014).

The regressor can be used in coarse to fine approaches, i.e. to first roughly capture the envelop of a structure of interest in images that cover wide areas and, then, to refine the shape of the object according to an adapted model. For example, Glocker et al. (2012) have relied on this idea to extract vertebrae in CT scans, Gauriau et al. (2013) to segment the liver in CT scan, Lindner et al. (2013) to segment bones in planar 2D radiography images, and so on.

Given the complexity of the learning task, the solution returned by regression-based approaches could present some outliers (Zhou, 2014). Therefore, some recent approaches present a hybrid regression-classification formulation: the regressor is used to globally initialize the detection task and, then, a classifier refines the solution in a narrowed search space. This idea has been presented by Lay et al. (2013) that have combined a randomized k-NN regression function with classifiers trained with locally discriminant information. Compared to a formulation that rely on a hierarchical structure, the approach proposed by Lay et al. (2013) is significantly more efficient. Gauriau et al. (2014) extend the idea of cascade of regressors by modeling local vote distributions depending on shape priors of the organ to detect. In details, to each organ is associated a *confidence map*, i.e. a probability distribution defined according to a probabilistic atlas, which allows conditioning the vote distribution and, so, refining the bounding box position and limits.

As for the definition of landmark clusters introduced in Chapter 4, both classification and regression approaches are interesting. From the literature it emerges that regression is more efficient than classification in addressing multi-object detection, but regression-based approaches have not been applied yet to full-body digital radiography images. As pointed out by Lindner et al. (2013) multiple aspects make anatomical structures detection challenging on planar 2D radiography images: the image quality significantly changes from an exam to another, there are rotational issues due to the projection of the 3D volume on a 2D plane and the intensity values inside the same structure are not homogeneous given tissue super-

position. Lindner et al. (2012) have shown that it is possible to localize the femoral head on pelvis radiography images, but the area of investigation is quite limited and, so, the model to learn is not very complex. As for EOS images, the learning procedure on a full-body image would be much more complex. So, it remains to establish if it is an achievable goal and, mostly, which are the most discriminant features to use to model the problem. Moreover, all these methods require many manually annotated data, which is always an issue in a medical imaging context. For this reason, the employment of unsupervised and semi-supervised methods is one of the most interesting perspectives in this field (Zhou, 2014).

Finally, from this description we retain that global to local and sparse search approaches are the best options to detect anatomical structures on full-body images.

5.1.4 Sequential models

The sequential models can be used in hierarchical approaches for recognition and segmentation of structures in an image. This family of methods is particularly interesting when the pairwise spatial relations between the objects to detect are stable while intensity and shape are not. Therefore, these methods have been applied to the detection and recognition of internal anatomical structures in brain MR images (Bloch et al., 2003; Colliot et al., 2006) that exhibit these properties. In detail, the procedure is driven by spatial relations that are encoded in a graph of the scene, where each node represents an object of interest. Distances and relative positions, like *close to*, *left of* and so on, are some examples of spatial relations between objects used to define a model capable of describing the scene. Please refer to Bloch (2005) for a survey on the definition of spatial relations.

Once the model is defined, each node of the graph provides fuzzy representations of the spatial domain associated with the other objects to detect. The fuzziness allows taking into account in a continuous way the natural changes in shape and relative positions of the structures. The spatial relations are described by parametric membership functions that depend on the domain and need to be trained. The learning procedure (Atif et al., 2007) consists in setting the parameters of the structuring element that describes a spatial relation, so that all the targeted structures are included in this support according to the manually annotated samples of a database. The object is localized by combining all its associated fuzzy sets. Finally, the object detection and recognition is carried out in a sequential framework that decomposes the problem into several sub-problems. The solution of one sub-problem contributes to facilitate the solution of the following ones. However, this demands to pre-define the order of the recognition sequence, which could be sub-optimal. As a consequence, Fouquier et al. (2012) have proposed an extension that allows avoiding this constraint. In details, visual salience (Treisman, 1985; Itti and Koch, 2001) is combined with spatial relations in order to define the sequence order that progressively simplifies the most the detection task. Similarly Zhan et al. (2008) have relied on mutual information to define the order of the hierarchical classification. Moreover, spatial knowledge is used to verify the consistency of the results and, eventually, backtrack the segmentation order.

Sequential models and regression-based approaches return comparable results because they both associate a spatial domain to target structures. However, the domains estimated by means of a sequential model and a regressor are, respectively,

probabilistic or fuzzy and deterministic[1]. Besides, the extensive regressor training procedure is replaced in sequential models by prior knowledge given by structuring elements related to spatial relations. So, the learning process is much easier at the price of higher algorithmic complexity. Finally, one should evaluate if it is easier to define consistent features for training a regressor or, instead, to explicitly model the spatial relations.

Similarly to structural models, a cascade of classifiers exploits pairwise spatial relations to sequentially narrow the solution space. However, the global context is used to define the order of the sequence and not to actually detect the object position which is a task solved by the classifier.

On EOS images this method could be useful to localize, for example, the starting and ending points of a cluster related to an anatomical structure and, hence, to initialize a local search for ideal landmark positions. However, the spatial consistency needs to be verified and the relations between structures to be defined.

5.1.5 Specificity of the landmark detection on EOS images

Some common aspects have been pointed out in the presented overview. First, it is important to exploit the consistency of spatial relations between the objects in the scene. Secondly, the technique of search for the solution should be sparsely conducted for the sake of both efficiency and accuracy. Finally, a global to local strategy should be preferred to only global or local ones because it gives higher performances in terms of both precision and computational load. The most recent works rely on extensive learning procedures and tend to prefer regression over classification, or to combine the two of them. However, EOS images combine the challenges of planar 2D radiography images and the wide search space. Therefore, the definition of consistent features for training a regressor is a highly non trivial task that needs to be further studied. The landmark detection is conducted on the image $\mathbf{u}$ that is not post-processed with the contrast enhancement and, thus, gradient- or texture-based features would not give a reliable indication. The method proposed in Chapter 3 could significantly simplify the definition of good descriptors, but it is more convenient to avoid this step as the EI values have to be efficiently estimated. Moreover, EOS images can be acquired at very different radiation exposure levels and, since the noise strength changes as well, it is difficult to define visual features that are robust in every case.

Sequential models seem to be more convenient to address our problem because they define descriptors that encode the spatial relations that intuitively are stable on EOS images and demand few labeled samples. This can help to both reduce the search space and recognize to which cluster the landmarks belong, and we use it to initialize our approach as presented in Section 5.3. However, a local analysis should be used to further improve detection and recognition. In particular, we propose to regularize sparse sets of salient points, introduced in Section 5.4, to define the anatomical clusters of landmarks on both frontal (Section 5.5) and lateral (Section 5.6) views.

[1]The regression forest allows associating a measure of robustness with the bounding box, but the domain itself is fixed.

5.2 Main notations

The notations used in the proposed landmarks detection method are listed here to facilitate the reading of the chapter.

The images and their sizes in pixels are denoted with bold and uppercase letters, respectively, e.g. $\mathbf{u}$ of size $R \times C$. On the other hand, 1D signals are indicated with the notation $\hat{u}$.

Points and sets of elements will be denoted as follows:

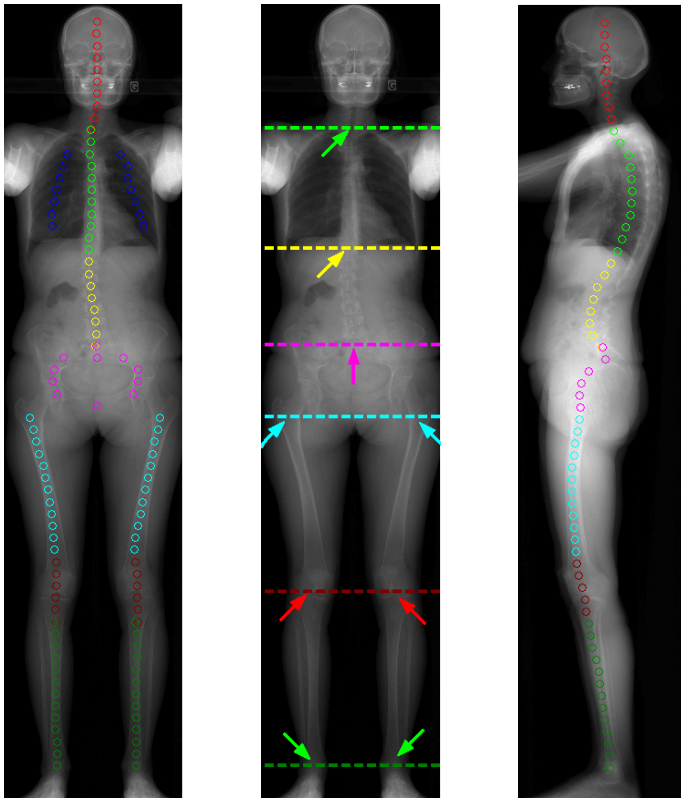
- A point: lowercase letter, e.g. p;
- Long version of the notation for a point: $p(x, y)$;
- x-coordinate of a point p_j: x_j;
- y-coordinate of a point p_j: y_j;
- Set of points: calligraphy letters, e.g. $\mathcal{P}$;
- Number of elements in a set $\mathcal{P}$: $|\mathcal{P}|$;
- Projection of $\mathcal{P}$ on x-coordinate axis: $\mathcal{P}_x = \{x_j \mid p_j \in \mathcal{P}\}$;
- Projection of $\mathcal{P}$ on y-coordinate axis: $\mathcal{P}_y = \{y_j \mid p_j \in \mathcal{P}\}$;
- Given a horizontal line whose points have all the same y-coordinate y_f, the points $p \in \mathcal{P}$ that lie on this line are defined as follows: $p(x_j, y_f)$, $j = 1, \ldots, |\mathcal{P}_{y_f}|$, where $|\mathcal{P}_{y_f}|$ is the number of points. This notation is used to describe operations that concern just a line.

Let us remember some notations that have been introduced in the previous chapter and will be used here as well. The clusters of anatomical landmarks l_j associated with an anatomical ROI $\mathcal{A}_r$ will be denoted by $\mathcal{L}_r$, i.e. $l_j \in \mathcal{L}_r$. The considered anatomical ROIs and corresponding color codes are the following ones: head ($\mathcal{L}_1$ - red), thoracic spine ($\mathcal{L}_2$ - green), lungs ($\mathcal{L}_3$ - blue), lumbar spine ($\mathcal{L}_4$ - yellow), pelvis ($\mathcal{L}_5$ - magenta), femurs ($\mathcal{L}_6$ - cyan), knees ($\mathcal{L}_7$ - dark red) and tibiae ($\mathcal{L}_8$ - dark green).

Finally, the local EI estimate associated with an anatomical landmark l_j will be denoted by $e(l_j)$. The EI corresponding to an anatomical region $\mathcal{A}_r$ will be then computed as follows:

$$EI_r = \frac{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j) e(l_j)}{\sum\limits_{l_j \in \mathcal{L}_r} \omega(l_j)} \tag{5.1}$$

where $\omega(l_j)$ is the weight associated with the local EI estimate at l_j (see Section 4.4.2).



(a) (b) (c)

Figure 5.1: *According to the manually annotated landmarks on the frontal view acquisition (a), 6 horizontal lines (b) can be identified to initialize the method (from top to bottom): T1 (green); T12 (yellow); L5 (magenta); proximal femur (cyan); knees (red); ankles (green). This initialization is also valid on the corresponding lateral view acquisition (c).*

5.3 Initialization of the method from spatial relations

The problem of detection and recognition of anatomical clusters of $\mathcal{L}_r$ is initialized by roughly separating the field of view in disjoint sub-windows according to the considered anatomical ROIs $\mathcal{A}_r$. The coarse identification of these areas helps to both narrow the search space for the landmarks $l_j \in \mathcal{L}_r$ and to verify if an estimated solution is coherent with the spatial relations between clusters. The manually annotated landmarks on frontal view acquisitions (Figure 5.1a) indirectly provide the ideal division in sub-windows (Figure 5.1b). For example, the thorax is identified as the region between the upper and lower landmarks of the cluster $\mathcal{L}_2$. In order to initialize our method, it is then necessary to identify 6 horizontal lines that pass through the following control points c_i: the vertebra T1, the vertebra T12[2], the vertebra L5, the proximal femur, the distal femur and the ankle. The problem is initialized by simply estimating the y-coordinates y_j of the control points $c_j \in \mathcal{C}$,

[2]We do not strictly need to associate a control point with an exact vertebra but rather to distinguish the chest from the abdomen. Therefore, the control point T12 may correspond to, for example, T10.

i.e. the projection $\mathcal{C}_y$ on y-axis where $\mathcal{C}$ is a set composed by the 6 control points.

Since the image is divided along the vertical direction, which is the common axis between the frontal and lateral acquisitions, the same initialization can be used on the two views. As justified in Section 5.5.1, the majority of the control points is estimated from frontal acquisitions.

Figure 5.1b shows that the control points can be described by simple spatial relations such as relative position (e.g. the vertebra T1 is above the vertebra T12) and distance. The relative positions of the control points is fixed, but, in order to encode the distances, a subset of 9 full-body EOS exams is used for learning. These samples well represent the variability of a bigger database composed by 82 patients, and the 9 patients have different ages, genders and morphotypes. Two approaches are studied to define the distances between control points, i.e. the absolute distances in cm from a given control point and the relative distances with respect to two already detected control points.

5.3.1 Absolute distance from a given control point

Table 5.1: *Rough estimation of $y_j \in \mathcal{C}_y$ for the ankle (y_6), proximal femur (y_4), L5 (y_3) and T12 (y_2) from T1 (y_1), and of the distal femur (y_5) from the ankle by using distance measures (κ_j) in cm computed from a set of 9 patients. According to statistical measures related to the distance (see Equation 5.2), a control point c_j is associated with an interval $[\alpha_j, \beta_j]$ that constrains the position of the control point on a new test image.*

	y_6 from y_1	y_4 from y_1	y_3 from y_1	y_2 from y_1	y_5 from y_6
$\bar{\kappa}_j$	127.29 *cm*	54.59 *cm*	39.28 *cm*	21.43 *cm*	35.97 *cm*
$\sigma(\kappa_j)$	9.95 *cm*	5.11 *cm*	3.13 *cm*	1.99 *cm*	3.20 *cm*
$\min(\kappa_j)$	112.50 *cm*	45.92 *cm*	34.44 *cm*	18.37 *cm*	32.14 *cm*
$\max(\kappa_j)$	142.34 *cm*	61.99 *cm*	43.62 *cm*	25.25 *cm*	41.33 *cm*
α_j	102.55 *cm*	40.81 *cm*	31.31 *cm*	16.38 *cm*	28.94 *cm*
β_j	152.29 *cm*	67.10 *cm*	46.75 *cm*	27.24 *cm*	44.53 *cm*
$\|\beta_j - \alpha_j\|$	49.74 *cm*	26.28 *cm*	15.44 *cm*	10.86 *cm*	15.58 *cm*

Table 5.1 reports the average ($\bar{\kappa}_j$), standard deviation ($\sigma(\kappa_j)$), minimal ($\min(\kappa_j)$) and maximal ($\min(\kappa_j)$) distances between a given horizontal line $y_j \in \mathcal{C}_y$, that is assumed to be already detected (e.g. T1), and another one, computed over the training database. Then, each control point c_j is associated with an interval $[\alpha_j, \beta_j]$ that constrains the position of c_j on a new test image according to the manually annotated data of the training set. Formally, the limits of the interval are defined as follows:

$$\alpha_j = \min(\kappa_j) - \sigma(\kappa_j) \qquad \beta_j = \max(\kappa_j) + \sigma(\kappa_j). \tag{5.2}$$

The last row of Table 5.1 reports the width of the interval and is used to quantify the incertitude on the initialization of the position of a control point c_j. The results indicate that the initialization is not very precise as, for example, the distance between the proximal femur and T1 could be between 40.81 *cm* and 67.10 *cm*, which is a quite wide range. This lack of precision is due to the heterogeneity of the data in the training set. The accuracy could be improved by considering a bigger

training set divided into different age groups. The size of the anatomical structures taken into account could also be obtained from the literature on the human anatomy. However, we are rather interested in the relative distances between control points and, thus, another approach is introduced in the following section.

5.3.2 Proportions

Table 5.2: *Rough estimation of $y_j \in \mathcal{C}_y$ for the proximal femur (y_4), L5 (y_3), T12 (y_2) and the distal femur (y_5) by considering proportions ϱ_j from neighbors (see Equation 5.3). The values for T1 (y_1) and the ankle (y_6) are assumed to be known. The same measures as the ones presented in Table 5.1 are reported here.*

	$y_4 \mid \{y_1, y_6\}$	$y_3 \mid \{y_1, y_4\}$	$y_2 \mid \{y_1, y_3\}$	$y_5 \mid \{y_4, y_6\}$
$\bar{\varrho_j}$	0.43	0.72	0.55	0.51
$\sigma(\varrho_j)$	0.02	0.02	0.02	0.01
$\min(\varrho_j)$	0.41	0.70	0.53	0.49
$\max(\varrho_j)$	0.45	0.75	0.58	0.53
α_j	0.39	0.67	0.51	0.47
β_j	0.47	0.77	0.60	0.55
$\|\beta_j - \alpha_j\|$	11.39 *cm*	4.96 *cm*	3.49 *cm*	6.43 *cm*

The position of a target control point will be intuitively more stable to changes in the data by exploiting the fact that the human body is approximately proportional. Formally, a horizontal line at y_j can be described according the proportion between the distances from two other horizontal lines of known positions. For example, let us assume that y_1 and y_6 have already been identified, and that y_4 has to be found. The following proportionality factor can then be associated with y_4:

$$\varrho_4 = \frac{y_4 - y_1}{y_6 - y_1}. \tag{5.3}$$

Table 5.2 reports the results of the training if the proportion between structures is considered rather than absolute distance. In particular, the width of the ranges is computed as follows:

$$\|\beta_4 - \alpha_4\| = \varrho_4 \max_i (y_6^i - y_1^i) \tag{5.4}$$

where y_j^i is the y-coordinate of the control point c_j on the patient i of the training set. The extension of the interval refers then to the worst case scenario, i.e. the maximal distance between the two referential horizontal lines over the training set. These values indicate that by exploiting the proportionality principle the initialization of the problem is more precise, i.e. a local analysis to define the exact position of the control points will be limited to a smaller region, increasing then both efficiency and robustness. Quantitatively, by using κ_j the search for the proximal femur, L5, T12 and distal femur will be conducted in areas that cover 26.28 *cm*, 15.44 *cm*, 15.58 *cm* and 10.86 *cm*. On the other hand, by relying on the proportions ϱ_j, the same control points will be searched in areas that cover 11.33 *cm*, 4.96 *cm*, 6.43 *cm* and 3.49 *cm*. The problem can then be initialized by using the learned proportionality factors given the upper and lower control points, i.e. the vertebra T1 and the ankle. Section 5.5.1 specifies how to detect these two initial horizontal points.

Some spatial relations in the horizontal direction can also be used to facilitate the localization of some landmarks $l_j \in \mathcal{L}_r$. For example, the clusters in the left and right lungs constrain the position of the landmarks $l_j \in \mathcal{L}_2$ (thoracic spine). Similarly, the clusters $\mathcal{L}_4$ (lumbar spine) and $\mathcal{L}_6$ (femurs) help defining the landmarks in the pelvic region as explained in Section 5.5.5.

The initialization from the frontal view is straightly used on the corresponding lateral view. However, it would be also useful to get the rough position of the most prominent points of the kyphotic and lordotic curves in the spine (see the clusters $\mathcal{L}_2$ and $\mathcal{L}_4$ in the image in Figure 5.1c). By computing the proportionality factor in this case, the interval associated with the most prominent point in the lumbar spine is equal to $[-0.01, 0.71]$ which implies that the initialization is highly uncertain. This is due to the posture of the patient that influences the position of the most prominent point. Similarly, the changes in the posture do not allow for a significant initialization on the horizontal direction and, thus, the lateral acquisition can only rely on information extracted from the frontal view to initialize the positions of control points.

5.4 Salient points

The initialization presented in the previous section imposes constraints on the position of the clusters and allows verifying the consistency of the estimation, but it does not provide information to locate the points that belong to a given cluster. The manual annotations in Figures 5.1a and 5.1c highlight that the landmarks to detect are sparse, i.e. they occupy only a small portion of the anatomical set of pixels Φ. In this section we address the definition of sets of salient points that are exploited in the local analysis, i.e. within a given cluster, to detect the landmarks. We formalize then the concept of saliency on EOS images and how to exploit it to obtain the sparse sets of salient points.

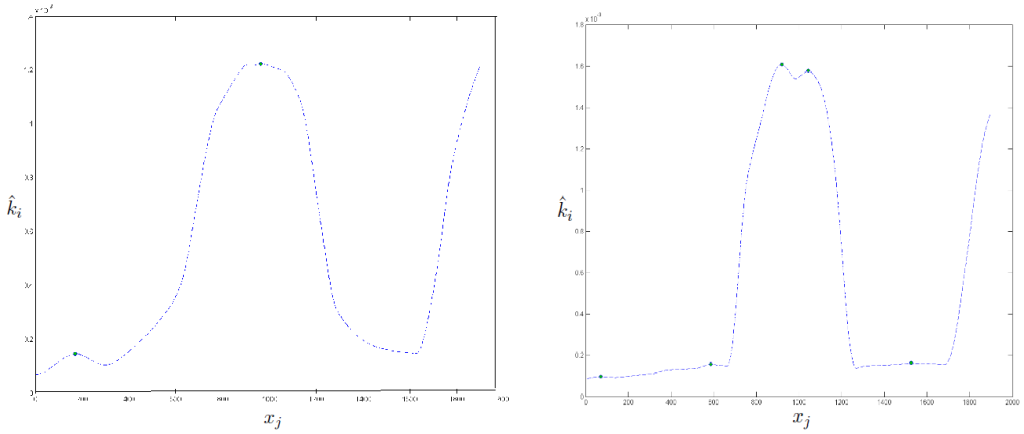
5.4.1 Saliency as change in density

From a very abstract point of view, the information in X-ray images can be associated with changes in the intensity levels as related to tissues of different densities. Therefore, when we look at an EOS image, the attention is essentially captured by these variations of signal. For example, in the abdomen, we first look at the lumbar vertebrae and then at the surrounding tissues because the bones absorb more signal and have irregular shapes. As a consequence, the relevant information can be associated with differences of intensity levels or with the presence of non-smooth image features. As discussed in Section 5.1.5, the search for the landmark positions is conducted on the image $\mathbf{u}$ without applying the contrast enhancement for the sake of efficiency. Therefore, only intensity level differences are exploited to define the salient points. Moreover, the information is scale-dependent and, hence, the search for salient points should be led at a scale that contains the features of interest. For example, a coarse scale could be sufficient to capture the envelop of the spine, whereas an analysis at finer scales would be required if the goal is to capture the internal structures of the vertebrae.

In practice, given the observed image $\mathbf{u}$ of size $R \times C$, non-overlapping sub-images $\mathbf{k}_i$ of size $S \times C$ are considered, where $S \ll R$, extracted from the vertical

positions $y_i \in \{1, \ldots, R\}$. A sub-image $\mathbf{k}_i$ is then projected to a 1D signal $\hat{k}_i$ where each element corresponds to the column-wise average. Afterwards, the signal is smoothed by using a linear average kernel of size S. This sequence of operations encodes the information related to X-ray absorption as a function of the sub-window height S that gives the information of scale. In the degenerate case $S = 1$, a signal $\hat{k}_i$ simply coincides with the horizontal profile of the image $\mathbf{u}$ at the line y_i.

The saliency in terms of intensity level variation is captured by local maxima or minima of the function $\hat{k}_i$ that are associated with peaks of signal at the detector and of absorption, respectively. The salient information of each sub-image $\mathbf{k}_i$ will be then linked with a set of point $\mathcal{P}_i$, where the x-coordinates x_j are the positions of salient points on $\hat{k}_i$ and $y_j = 0.5(y_i + y_{i+1}), \forall j$, i.e. the y-coordinates are simply the centers of the sub-images $\mathbf{k}_i$.



(a) (b)

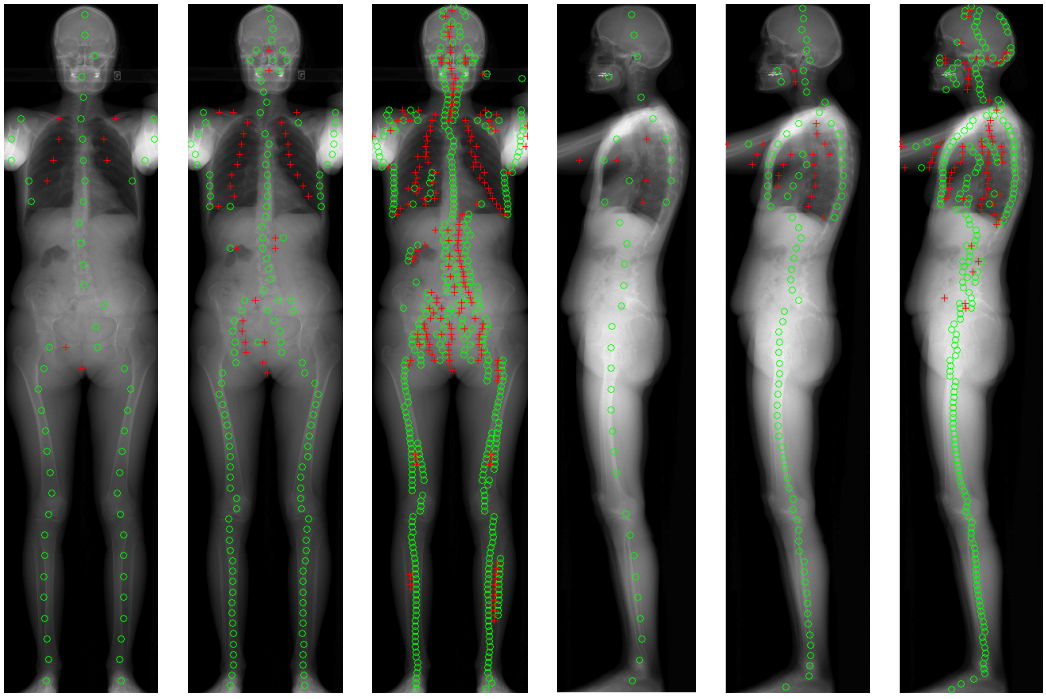
Figure 5.2: *A signal $\hat{k}_i$ that encodes the neck on a frontal view image at different scales: (a) $S = 128$; (b) $S = 64$. The scale has an impact on the amount of captured information and outliers.*

Figures 5.2a and 5.2b show two examples of functions $\hat{k}_i$ that capture the same region but at difference scales, i.e. $S = 128$ and $S = 64$, respectively. Since the structure in the region is the neck, the relevant information is located at the center (i.e. $x_j \simeq 1000$) while some outliers associated with the shoulders appear in the leftmost and rightmost sides of the signals $\hat{k}_i$. We consider as outliers the salient points that are associated with regions that are not covered by any of the anatomical clusters $\mathcal{L}_r$ taken into account. By setting $S = 128$, only one value is detected at the center of the cervical spine, whereas, by using $S = 64$, two points inside the vertebra are found. Therefore, coarser scales have the advantage of being sparser and, hence, present less outliers, but may fail in capturing some relevant details. On the other hand, the finer the scale the more precise the analysis is, but the number of outliers also increases. As a consequence, the choice of the value S should depend on the estimated minimal degree of precision that is required to well describe a structure of interest. In the given example, $S = 128$ is the preferable setting because there is no need to associate two points with the vertebra, for applications to EI estimation.

The scale values associated with the anatomical structures to capture have been

empirically defined in the method that is described in the following sections. Nevertheless, in a future version of the algorithm, we could automatically set the values of S depending on the number of detected points in a region and on their approximate relative distances.

5.4.2 Applications



(a) (b) (c) (d) (e) (f)

Figure 5.3: *Sets of salient points $\mathcal{P}_{(l,s)}$ and $\mathcal{P}_{(h,s)}$ on a frontal view acquisitions at (a) $S = 256$; (b) $S = 128$; (c) $S = 64$ and corresponding lateral view at (d) $S = 256$; (e) $S = 128$; (f) $S = 64$.*

Figure 5.3 shows some examples of salient points on the frontal and lateral view acquisitions for a given exam at different scales S. The red stars indicate salient points $p_j \in \mathcal{P}_{(l,s)}$ associated with peaks of signal at the detector (index l) and computed at the scale S (index s), whereas the green circles are the salient points $p_j \in \mathcal{P}_{(h,s)}$ that correspond to strong absorption regions (index h) at the scale S (index s). Note that the salient points are computed over the raw image $\mathbf{u}$, but in Figure 5.3 we display the points on the output image. For the latter, the gray levels look up table is inverted with respect to $\mathbf{u}$, i.e. high and low intensity values correspond to strong and low absorptions, respectively.

On frontal acquisitions, the points $p_j \in \mathcal{P}_{(l,s)}$ are mainly located in the lungs with a limited amount of outliers in the region that surrounds the groin. Note that a quite good approximation of the anatomical cluster $\mathcal{L}_3$ is obtained at scale $S = 128$. On the other hand, $\mathcal{P}_{(l,256)}$ does not give a precise enough description and many points $p_j \in \mathcal{P}_{(l,64)}$ are located in the spine and the pelvis, but peaks of signal in these regions are not robust as they mainly depend on the morphotype.

The points $p_j \in \mathcal{P}_{(h,s)}$ cover the whole field of view as peaks of absorption are mostly due to the presence of bone tissues. The majority of the points $p_j \in \mathcal{P}_{(h,s)}$ well fits the spine and the bones in the legs both with $S = 256$ and $S = 128$, which implies that the clusters $\mathcal{L}_r$ with $r \neq 3, 5$ (i.e. all the ROIs except the lungs and the pelvis) could be well estimated by regularizing these salient points. On the other hand, it is more complicated to get a proper initialization in the pelvic region because in strong absorption regions the signal tends to be homogeneous. In this case, it is preferable to rely on an over-complete representation such as the one obtained by setting S to 64 .

On lateral acquisitions, the salient points $p_j \in \mathcal{P}_{(l,s)}$ are concentrated in the thoracic region but they do not provide useful information considering that the cluster $\mathcal{L}_3$ (lungs) is neglected on lateral views. On the contrary, all the landmarks $l_j \in \mathcal{L}_r$, with $r \neq 2$ (i.e. except for the thoracic spine) are well initialized by using the salient points $p_j \in \mathcal{P}_{(h,128)}$, that offer the best compromise between completeness of the representation and reduction of outliers. Besides, note that the salient points $p_j \in \mathcal{P}_{(h,128)}$ in the chest well respond to the thoracic cage, which helps defining the cluster $\mathcal{L}_2$ as described in Section 5.6.2.

These remarks are valid for all the 82 patients in our database and, furthermore, no noticeable change of trend has been observed on patients with high body mass index or on preview acquisitions. Such good results are due to the properties of the high dynamic range detector that is used to acquire EOS images: tissues at different density levels can be distinguished even at very low signal levels.

The definition of the salient points completes the initialization of the method introduced in Section 5.3. Indeed, according to the positions of the control points c_j, the salient points can be assigned to the classes associated with the ROIs $\mathcal{A}_r$. Besides, the outliers can be removed by both exploiting the consistency of spatial relations provided by the model and by locally regularizing the salient points according to dedicated descriptors and spatial relations between intra-cluster landmarks. Similarly, the control points c_j (Section 5.3) can be refined by exploiting the information provided by the salient points. In the following sections, we define a method that applies these ideas to automatically detect and recognize the landmarks necessary to automatically estimate EI values.

5.5 Landmarks detection and recognition on frontal view acquisitions

Figure 5.4 summarizes the algorithm of detection and recognition of the anatomical clusters $\mathcal{L}_r$ on a frontal view image that is the first to be processed. The inputs to the algorithm are the learned spatial relations, the image $\mathbf{u}_r$ and the mask that separates the set $\Phi \subset \Omega$ of pixels associated with X-ray absorption from the background. These data are used at all the steps of the method. The salient points are not computed over the whole pixel space Φ but rather on sub-windows defined during the detection procedure and at a scale chosen according to the useful information to encode. Once the control points for ankle and T1 are detected, those for the proximal femur and L5 are sequentially located. Afterwards, the problem is divided into two independent sub-problems, i.e. the localization of landmarks in the upper part of the body (spine and lungs) and in the legs. The landmarks in the leg

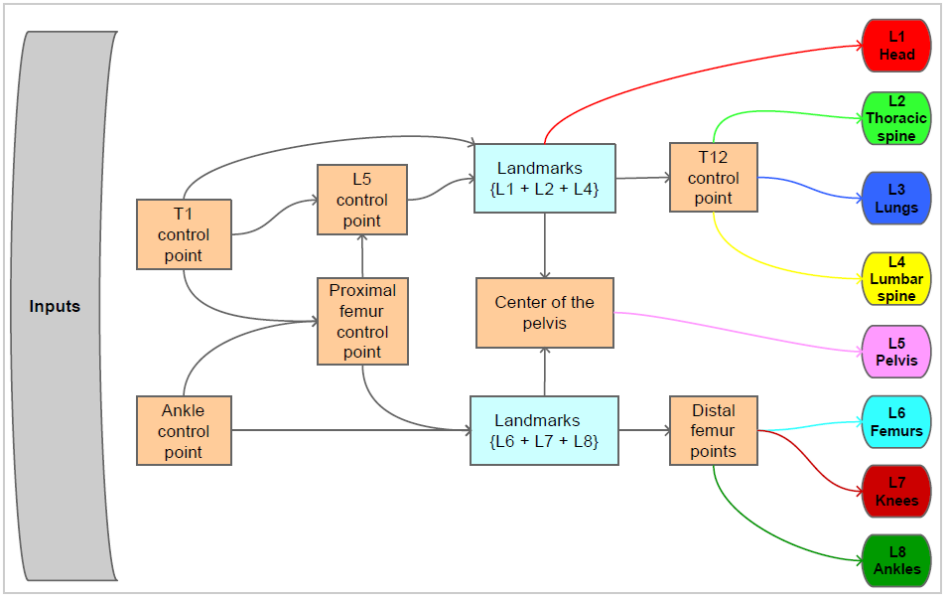


Figure 5.4: *The control points are sequentially detected according to the learned spatial relations and local analysis of image features. The landmarks in the upper part of the body and in the legs are first detected and then the definition of the control points for T12 and distal femur allows separating them into* $\mathcal{L}_1$-$\mathcal{L}_2$-$\mathcal{L}_3$-$\mathcal{L}_4$ *and* $\mathcal{L}_6$-$\mathcal{L}_7$-$\mathcal{L}_8$*, respectively. The cluster* $\mathcal{L}_5$ *at the last step because it is the most difficult region where to define consistent salient points.*

$l_j \in (\mathcal{L}_6 \cup \mathcal{L}_7 \cup \mathcal{L}_8)$ are located as described in Section 5.5.2 and the division in the three sets of interest is directly derived from the control point for distal femur. In the spine, the landmarks $l_j \in (\mathcal{L}_2 \cup \mathcal{L}_4)$ are obtained as described in Section 5.5.3 and the landmarks $l_j \in \mathcal{L}_1$ are easily placed according to the position of T1 and to the mask Φ. In order to get $\mathcal{L}_2$ and $\mathcal{L}_4$, the control point for T12 needs to be defined. The landmarks $l_j \in \mathcal{L}_3$ are inferred from $\mathcal{L}_2$ and by exploiting the salient points $\mathcal{P}_{(l,s)}$ (see Section 5.5.4). Finally, the cluster $\mathcal{L}_5$ is obtained with local analysis and by exploiting the already detected landmarks $l_j \in \mathcal{L}_4$ and $l_j \in \mathcal{L}_6$ as discussed in Section 5.5.5.

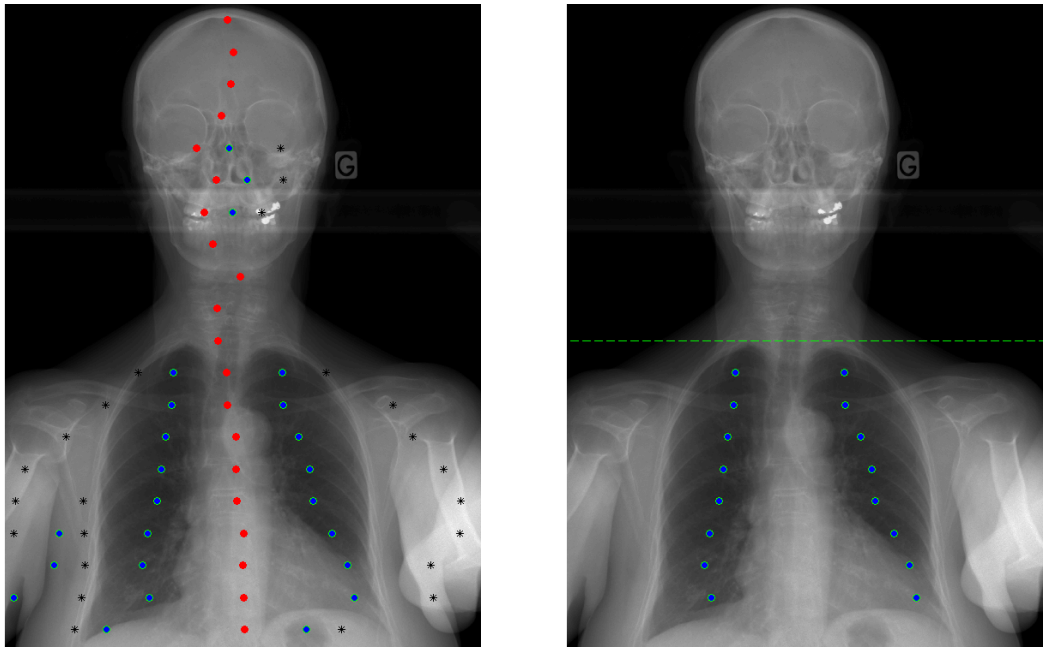
In the following sections, we first specify how to detect the control points on a given frontal view test image and we then discuss how to refine the landmark positions in each anatomical ROI.

5.5.1 Detection of control points

5.5.1.1 T1

The control points $c_j \in \mathcal{C}$, or rather their y-coordinates $y_j \in \mathcal{C}_y$, are sequentially defined. The procedure starts by identifying two main control points that are the vertebra T1 and the ankle. These points are chosen among the others not only because they are located at the upper and bottom extrema of the image, but mostly because they are the easiest ones to detect in a completely unsupervised way. While the estimation of y_6 (ankle) is discussed in Section 5.5.1.2, we present here how to

define y_1 (T1).



(a) (b)

Figure 5.5: *Detection of the control point for T1. Figure (a) shows the process of detection that relies on the position of the points* $p_j \in \mathcal{P}_{(h,128)}$ *(blue),* $p_j \in \mathcal{P}_{(h,128)}$ *(black) and* $t_j \in \mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ *(red) that are associated with the landmarks near to the central vertical axis. Figure (b) displays the retained points* $t_j \in \mathcal{T}_1 \subset \mathcal{P}_{(l,128)}$ *that roughly correspond to the landmarks* $l_j \in \mathcal{L}_3$ *and the horizontal line that passes through the control point T1.*

The vertebra T1 is above the lungs and centered with respect to the landmarks $l_j \in \mathcal{L}_3$ associated with the left lung from one side and with the right lung from the other. This relation can be exploited in order to localize the control point T1. Figure 5.3b shows that the salient points $p_j \in \mathcal{P}_{(l,128)}$ quite well capture the thorax and, hence, we can analyze their distribution in order to infer the y-coordinate of c_1. In practice, given an image $\mathbf{u}$ of size $R \times C$ a sub-window $\mathbf{k}$ of size $V \times C$ is extracted from the first line of the image and by setting $V = 0.30R$, as, empirically, $y = y_1$ coincides with one of the first V lines of a full-body image. Figure 5.5a shows an example of image $\mathbf{k}$ and of the salient points $\mathcal{P}_{(l,128)}$ (blue circles) and $\mathcal{P}_{(h,128)}$ (black stars) computed therein.

A vertical line of equation $x = \gamma_x$ is considered, where γ_x is computed from the points $p_j \in \mathcal{P}_{(h,128)}$ as follows:

$$\gamma_x = \sum_{p_j \in \mathcal{P}} \frac{x_j}{|\mathcal{P}|} \tag{5.5}$$

where the indices of the notation $\mathcal{P}_{(h,128)}$ are omitted for simplicity. We then consider a subset $\mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ that is composed by points $t(x_f, y_f)$, one for each line $y = y_f$,

$f = 1, \ldots, |\mathcal{P}_y|$, where the x-coordinates x_f are defined as follows:

$$x_f = \arg\min_{x_j} \left(d(p(x_j, y_f), x = \gamma_x)\right) \tag{5.6}$$

where $d(p(x_j, y_f), x = \gamma_x)$ is the shortest Euclidean distance between a point $p(x_j, y_f)$ and the vertical line $x = \gamma_x$. Figure 5.5a shows that the points $t_j \in \mathcal{T}_0$ (red points) roughly matches with the thoracic spine in the chest region and, by exploiting the symmetry between the landmarks in the right and left lungs, some outliers can be removed from $\mathcal{P}_{(l,128)}$. For each horizontal line $y = y_f$, $f = 1, \ldots, |\mathcal{P}_y|$, the point $t(x_f, y_f)$ is fixed. The nearest points to the left and to right of $t(x_f, y_f)$ are then retained as possible candidates of left and right lungs landmarks, respectively, to build a subset $\mathcal{T}_1 \subset \mathcal{P}_{(l,128)}$. Formally, the x-coordinates of the left (x_l) and right (x_r) points on a given line $y = y_f$ are defined as follows:

$$\begin{cases} x_l = \arg\min\limits_{x_j < \gamma_x} \left(d(t(x_f, y_f), p(x_j, y_f))\right) \\ x_r = \arg\min\limits_{x_j > \gamma_x} \left(d(t(x_f, y_f), p(x_j, y_f))\right) \end{cases} \tag{5.7}$$

where $d(t(x_f, y_f), p(x_j, y_f))$ is the Euclidean distance between the points $t(x_f, y_f)$ and $p(x_j, y_f)$. The limits of lungs are then defined according to the distances between these points. In particular, the smallest and highest distances are retained as the upper and lower limits of the lungs, respectively. Figure 5.5b shows the subset $\mathcal{T}_1 \subset \mathcal{P}_{(l,128)}$ and the horizontal line that passes through the control point c_1 whose y-coordinate is defined as follows:

$$y_1 = \arg\min_{y_j \in \mathcal{T}_y} \left(d(t(x_l, y_j), t(x_r, y_j))\right) - S, \tag{5.8}$$

where S is the scale fixed in this case to 128. According to the tests on our database composed by 82 patients, this method has shown good properties, even if for some cases the position was associated with T2 rather than T1. However, this does not significantly influence the EI estimates. These good performances are due to the consistency of the distributions of salient points $\mathcal{P}_{(l,128)}$ and $\mathcal{P}_{(h,128)}$ in the chest.

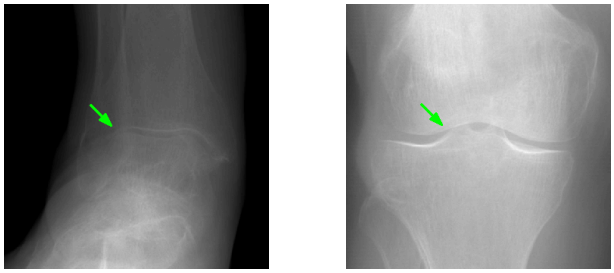
5.5.1.2 Control points in the legs

The control points in the legs are detected in the following order: ankle (c_6), proximal femur (c_4) and distal femur (c_5).

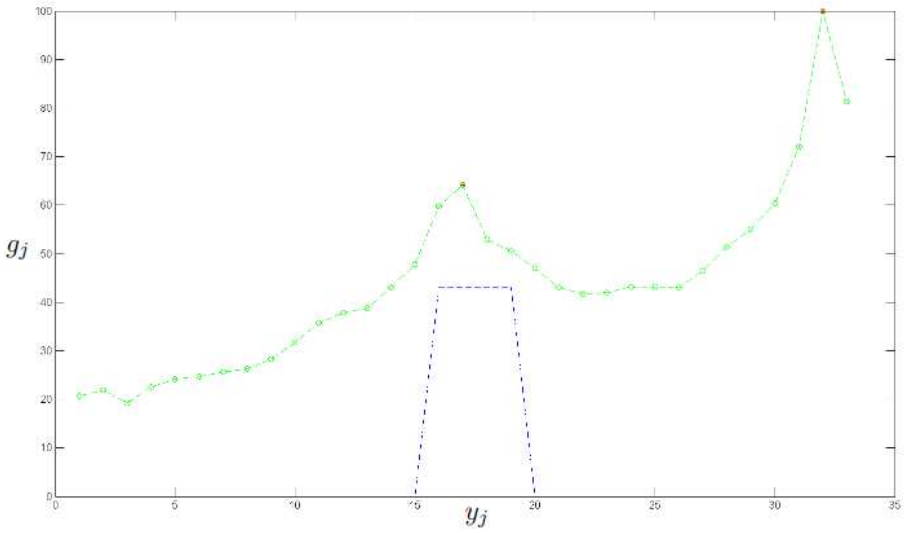
Figure 5.6a shows the lower bound of the tibia that precedes the ankle cavity. As observed in the image, the junction between the two bones causes a strong horizontal discontinuity which can be well captured by a gradient-based descriptor. In practice, we consider a sub-window $\mathbf{k}$ of size $V \times C$, $V = 0.30R$ that extracts the bottom part of a given test image $\mathbf{u}$. The salient points $\mathcal{P}_{(h,128)}$ are then computed over the image $\mathbf{k}$ and the following gradient based measure is associated with each point $p_j \in \mathcal{P}_{(h,128)}$:

$$g_j = \sum_{(x_i, y_i) \in P_j} \frac{\|\Delta_h(x_i, y_i)\|}{|P_j|} \tag{5.9}$$

where $\Delta_h(x_i, y_i)$ is the horizontal gradient value at the pixel of coordinates (x_i, y_i) that belongs to a circular patch P_j of radius equal to $S/2$ centered at the point p_j.



(a) (b)



(c)

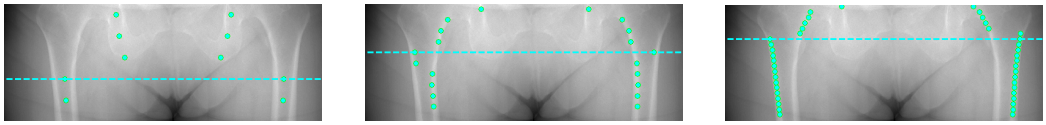
Figure 5.6: *Detection of control points for distal femur and ankle. Both the ankle (a) and the knee (b) present strong horizontal discontinuities that can be captured by using a gradient-based descriptor. The control points for distal femur and ankle are defined as local maxima of the function g_j (c) by imposing constraints on their possible positions according to the initialization presented in Section 5.3.*

The size of the patch is chosen such that the measure associated with a given point is independent from the ones computed at nearby points.

Considering the width of the human leg, by setting $S = 128$ only one landmark on each S horizontal line is obtained and, then, the measure g_j computed at the landmark l_j can be expressed as a function of its y-coordinate y_j. In other words, the measures g_j computed at the anatomical landmarks $l_j \in (\mathcal{L}_6 \cup \mathcal{L}_7 \cup \mathcal{L}_8)$ that belong to one leg can be represented as a mono-dimensional function of the y-coordinate y_j (Figure 5.6c). The control point for the ankle is located at the position of the last peak of the gradient-based measures. The left- and right-hand sides correspond to the upper femur and ankle regions, respectively. Therefore, the ankle is associated with the peak of the graph in the right-most side of the graph (red start in Figure 5.6c).

The same measure is used to define the control point for distal femur because, as shown in Figure 5.6b, this structure also presents a strong horizontal discontinuity.

The control point c_5 is then defined as the position of a peak of g_j by constraining the space of the possible solutions in the region (see blue line in Figure 5.6c) defined as a function of the proportion ratio (Equation 5.3) between the distances $c_5 - c_4$ and $c_6 - c_5$ learned from the manually annotated data (see Table 5.2).



(a) (b) (c)

Figure 5.7: *The control point for the proximal femur is detected by merging estimates from salient points computed at different scales: (a)* $\mathcal{P}_{(h,128)}$*; (b)* $\mathcal{P}_{(h,64)}$*; (c)* $\mathcal{P}_{(h,32)}$*.*

The proximal femur is the most difficult control point in the leg to detect given the proximity to another anatomical region, i.e. the pelvis, and for the high absorption due to tissue superposition compared to that in the knees and ankles. Figure 5.7 shows the sub-window **k** where to find the proximal femur which is defined according to the spatial relation model (Table 5.2) and, in details, Figures 5.7a, 5.7b and 5.7c display three sets of points $\mathcal{T}_0$ computed at the scales $S = 128$, $S = 64$ and $S = 32$, respectively. The set $\mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ is composed, for each line y_f, $f = 1, \ldots, |\mathcal{P}_y|$, by the leftmost and rightmost points of x-coordinates x_l and x_r, respectively. The extrema are retained because they logically are the most probable to correspond to the legs. Formally, x_l and x_r for a given y_f are defined as follows:

$$\begin{cases} x_l = \min\{x_j \mid p_j \in \mathcal{P}_{y_f}\} \\ x_r = \max\{x_j \mid p_j \in \mathcal{P}_{y_f}\} \end{cases} \tag{5.10}$$

where $\mathcal{P}_{y_f}$ is the set of points $p_j \in \mathcal{P}$ that lie on the line defined by $y = y_f$. The y-coordinate of the control point c_4 is then defined as follows:

$$y_4 = \underset{y_j \in \mathcal{P}_y}{\arg\max} \left(d\left(t(x_l, y_j), t(x_r, y_j)\right)\right), \tag{5.11}$$

which means that the proximal femur is on the line y_j where the points $t(x_l, y_j)$ and $t(x_r, y_j)$ are the most distant, which is coherent with the assumption made for the manual annotations. Nevertheless, the distribution of the points may change depending on the morphotype or the presence of metallic objects. Therefore, the position of the proximal femur is estimated multiple times, i.e. one time for each considered scale S. According to multiple tests conducted on part of the database, we remarked that either the estimates are similar regardless the chosen scale or two estimates are pretty close (Figures 5.7b and 5.7c) and one is wrong (Figure 5.7a). As a consequence, the final position is computed as the median of the estimates at the three considered scales, which is a pretty basic way of exploiting information at multiple scales, but good enough according to the tests and application of interest.

5.5.1.3 Control points in the spine

The control points in the spine, except for the vertebra T1, are the most difficult ones to precisely define on frontal view acquisitions and, hence, are reserved for

the last step. The difficulty comes from the complexity in distinguishing between a vertebra and another one, and consistent features are not trivial to define considering the rotational issues, the low contrast to noise ratio and the tissue superposition.

According to our tests, the definition of consistent features that allow capturing the vertebra L5 (c_3), despite changes in morphotype and entrance dose, remains an open question, but, for the EI application, this does not pose a significant problem for two reasons. First, according to the values in Table 5.2, the interval where to look for c_3 is relatively narrow, i.e. 4.96 cm in the worst case scenario. Secondly, as pointed out in Chapter 4, the EI values associated with the lumbar spine and pelvis are similar and, hence, eventual misplacement of c_1 would only slightly affect the exposure indicators. Finally, the position of L5 is just roughly inferred from the vertebra T1 and the proximal femur by using a proportionality factor equal to 0.72, i.e. at the center of the interval given by the model.

As for the control point T12 (c_2), it may be defined as a function of spatial relations with respect to the landmarks in the lungs as explained in Section 5.5.1.1. Nevertheless, this control point can be much more easily detected on the lateral view acquisition as presented in Section 5.6.4. Therefore, on the frontal view the control point p_2 is initialized according to the model by using the proportionality factor with respect to the vertebrae T1 and L5, and refined later on, after the analysis of the corresponding lateral view acquisition.

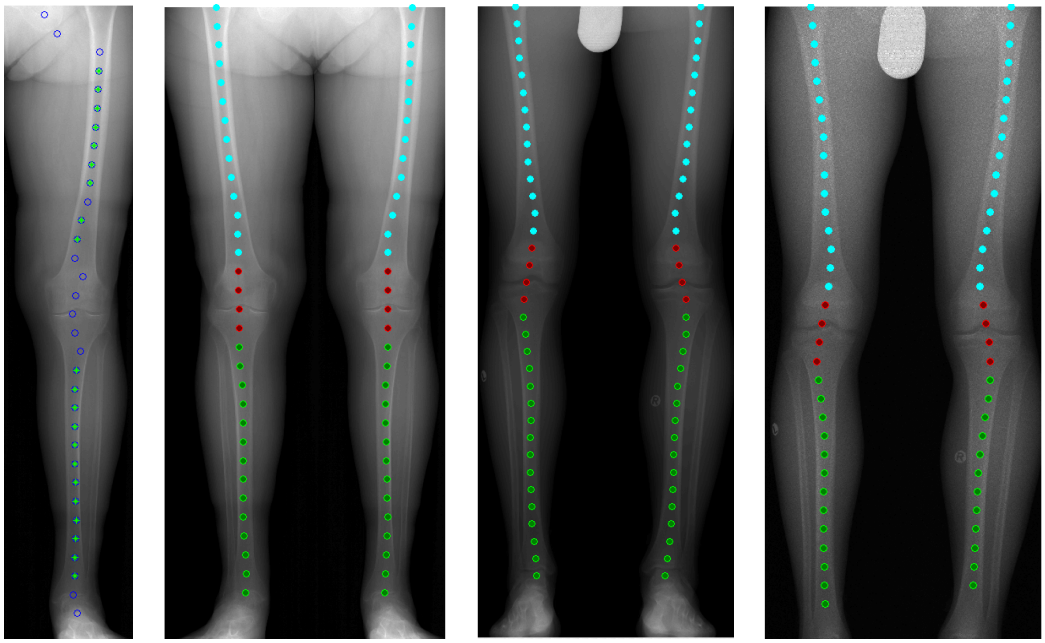
5.5.2 Detection of the landmarks in the legs

The landmarks $l_j \in (\mathcal{L}_6 \cup \mathcal{L}_7 \cup \mathcal{L}_8)$ are located over the image $\mathbf{k}$ that separates the lower limbs from the rest of the body according to the location of the proximal femur control point. The salient points $p_j \in \mathcal{P}_{(h,128)}$ and a subset $\mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ are defined as in Equation 5.10 to take into account only the landmarks that are probably associated with bone tissues.

The method processes separately each leg and, hence, considers a subset $\mathcal{T}_1 \subset \mathcal{T}_0$ that contains the salient points in the left or right leg. From the set $\mathcal{T}_1$, a cluster of significant points $\mathcal{T}_2$ is extracted to refine their positions. Figure 5.8a shows a leg and its associated sets $\mathcal{T}_1$ and $\mathcal{T}_2$. As revealed by the example, a point is considered significant if aligned with its two nearest neighbors. Formally, we consider the angles θ_u and θ_d between the vertical axis of the image and the vectors $\overrightarrow{t_j t_{j-1}}$ and $\overrightarrow{t_j t_{j+1}}$, respectively, $j = 2, \ldots, |\mathcal{T}_1|-1$. A weight υ_j is then associated with the point $t(x_j, y_j)$ as follows:

$$\upsilon_j = \exp\left(-\frac{|\theta_u - (\theta_d + \pi)|}{h}\right), \quad \theta_u, \theta_d \in [-\pi, \pi] \tag{5.12}$$

where h is the smoothing parameter that controls the decay of the exponential function and in our tests is fixed at 30. Therefore, the more the points $t(x_{j-1}, y_{j-1})$ and $t(x_{j+1}, y_{j+1})$ are symmetric with respect to $t(x_j, y_j)$ the higher the weight υ_j is. This principle is used to define the most meaningful points in the leg because the femur and the tibia are straight bones and, hence, nearby landmarks should be aligned. Then, a point $t_j \in \mathcal{T}_2$ is kept if the associated weight $\upsilon_j > \tau$, where τ is a threshold fixed at 0.75. As for the choice of the parameters introduced here, they have been selected depending on empirical evaluations during our tests. In practice, the definition of these parameters is not critical as it does not significantly influence the results. The cluster $\mathcal{T}_2$ neglects some outliers in the set $\mathcal{T}_1$ such as those near the



(a) (b) (c) (d)

Figure 5.8: *Detection process and examples of results for the clusters in the leg. A subset of relevant points $t_j \in \mathcal{T}_2 \subset \mathcal{T}_1$, where $\mathcal{T}_1 \subset \mathcal{P}_{(h,128)}$ corresponds to the landmarks associated with a single leg, is defined. Results: (b) adult morphotype; (c) young morphotype; (d) preview of the exam in (c). The cyan, red and green points correspond to $\mathcal{L}_6$, $\mathcal{L}_7$ and $\mathcal{L}_8$, respectively.*

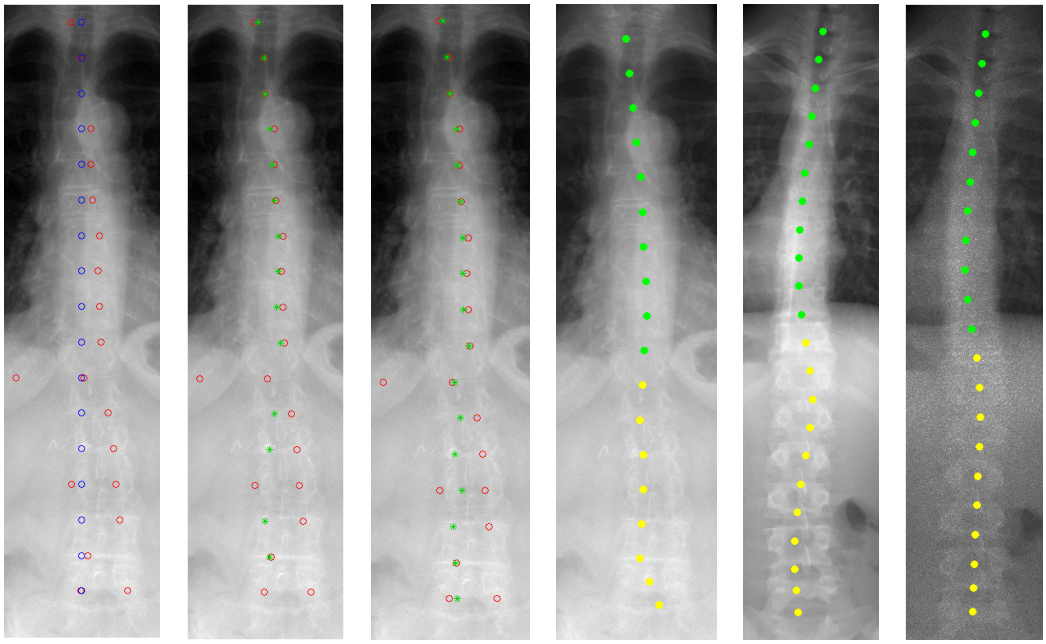
pelvis and off-center points in the knee. The set $(\mathcal{L}_6 \cup \mathcal{L}_7 \cup \mathcal{L}_8)$ is finally obtained by linearly interpolating the points $t_j \in \mathcal{T}_2$ and the corresponding anatomical clusters are separated by the control point distal femur.

Figures 5.8b, 5.8c and 5.8d show three examples of detection of landmarks in the legs for an adult patient, a young patient and relative preview acquisition, respectively. These examples well summarize the quality of results obtained on the whole dataset for the three mentioned categories and will be used again to complete our discussions on other anatomical regions or views. In particular, note that the entrance dose used in the diagnostic exam given as example is equal to 114.18 μGy, whereas in the corresponding preview it is equal to 1.43 μGy. Figures 5.8b, 5.8c and 5.8d prove that the clusters $\mathcal{L}_{6-8}$ are well defined regardless changes in the morphotype and the drastic reduction of dose. Nevertheless, by looking at the bottom of the image in Figure 5.8d, it can be noted that the ankle control point is not correctly located, but this is simply due to the fact that it is outside of the field of view.

5.5.3 Detection of the landmarks in the spine

The sub-image **k** that contains the thoracic and the lumbar spine is defined depending on the positions of the control points for T1 and L5. Besides, the horizontal field of view is narrowed as a function of the set $\mathcal{P}_{(l,128)}$ (see Section 5.5.1.1) that

roughly corresponds to the cluster associated with the lungs ($\mathcal{L}_3$). The positions of the landmarks $l_j \in \mathcal{L}_3$ constrain those associated with the spine ($l_j \in (\mathcal{L}_2 \cup \mathcal{L}_4)$) because the left and right lungs are symmetric with respect to the thoracic spine and, hence, to the lumbar spine as well. In practice the uppermost points $p_j \in \mathcal{P}_{(h,128)}$ (see Figure 5.5b) are considered to define the vertical crop limits. The field of view is enlarged by a padding equal to $S/2$ for each side.



(a) (b) (c) (d) (e) (f)

Figure 5.9: *Detection of the landmarks in the spine: (a) initialization with spatial prior positions (blue points) and $\mathcal{P}_{(h,128)}$ (red points); (b) iteration number 1; (c) iteration number 2. Results: (d) adult morphotype, (e) young morphotype and (f) preview of the exam in (e). The yellow and green points correspond to the landmarks $l_j \in \mathcal{L}_2$ and $l_j \in \mathcal{L}_4$, respectively.*

In order to solve the detection problem in the spine, we consider two sets $\mathcal{P}_{(h,128)}$ and $\mathcal{B}$, where $\mathcal{P}_{(h,128)}$ is the set of peaks of absorption computed over $\mathbf{k}$, and $\mathcal{B}$ initializes the problem by simply placing the landmarks on the vertical central axis of $\mathbf{k}$ (see Figure 5.9a). The solution is obtained by modifying these two sets in an iterative way. In details, for each line $y_f = 1, \ldots, |\mathcal{P}_y|$, the following energy function is minimized over t_j:

$$\varphi_b(t_j, b_f) + \lambda \varphi_p(t_j, p_f) \tag{5.13}$$

where λ weights the contributions from two terms $\varphi_b(.)$ and $\varphi_p(.)$ that are linked to the sets $\mathcal{B}$ and $\mathcal{P}_{(h,128)}$, respectively.

The first term is defined as follows:

$$\varphi_b(t_j, b_f) = \frac{(t_j - b_f)^2}{S^2} \tag{5.14}$$

where $t_j = t(x_j, y_f)$ is a candidate point that is tested for the minimization of the term in Equation 5.13, and $b_f = b(x_f, y_f)$ is the point that belongs to $\mathcal{B}$ and lies

on the horizontal line $y = y_f$. In practice, the minimization of the energy function is obtained by testing different values of j, i.e. by changing the x-coordinate of the test point. This term is called spatial term as it enforces prior spatial regularity by considering that the vertebrae are probably placed at the center of the image $\mathbf{k}$. Note that this will not be true for all the vertebrae of a patients affected by scoliosis, for example.

As a consequence, a second term $\varphi_p(t_j, p_j)$ is defined as follows

$$\varphi_p(t_j, p_f) = \left| 10 \log_{10} \left(\frac{e(t_j)}{e(p_f)} \right) \right|. \tag{5.15}$$

where $e(t_j)$ is the local EI estimate computed at the point t_j, which is a tested solution, and p_j is a point that belongs to $\mathcal{P}_{(h,128)}$ and lies on the horizontal line $y = y_f$. The superposition with tissues in strong absorption regions, e.g. the breast in Figure 5.9a, may cause outliers and, then, more than one point p_j on the horizontal line $y = y_f$. In a first step, we then avoid solving the problem if there are multiple salient points over the line $y = y_f$ (see Figure 5.9a). Equation 5.15 indicates that the local EI values computed at salient points associated with peaks of absorption are taken as references. These points give, indeed, a quite reliable initialization because, on images acquired from the frontal view, the spine is associated with peaks of absorption that distinguish it with respect to the surrounding tissues both in the thoracic and lumbar region. Therefore, some of the points $p_j \in \mathcal{P}_{h,128}$ will probably be quite close to the correct positions of the landmarks in the spine. Nevertheless, given the possible presence of outliers, the estimates from these points are not totally reliable and need, thus, to be regularized by enforcing prior spatial regularity with the first term in Equation 5.13.

The value of λ significantly influences the quality of the result: a too high value may lead to unstable positions because $\mathcal{P}_{(h,128)}$ is affected by outliers, e.g. the breast in the upper region of the abdomen in Figure 5.9a. On the other hand, a too low value may not allow following the spine because the patient is in bending position or the spine is distorted. Note that, with λ fixed to 0, the minimization of the energy function in Equation 5.13 will return the set $\mathcal{B}$ that does not take into account image features.

The process of minimization is repeated n times. Figure 5.9b shows the results after the first iteration ($\mathcal{T}_0$). As mentioned before, only lines for which there is only one point p_j are considered. The following step is initialized by replacing the set $\mathcal{B}$ with $\mathcal{T}_1$, which is the solution obtained at the previous step and linearly interpolated to fill potential holes. Figure 5.9c shows the result $\mathcal{T}_2$. Note that the value of λ is reduced by a factor equal to the number of the iteration, i.e. $\lambda/2$ at the second iteration, $\lambda/3$ at the third one and so on. This allows taking into account that $\mathcal{T}_i$, $i = 1, \ldots, n_i$, is nearer and nearer to the real solution and, hence, the importance of the initial EI estimates has to be progressively reduced.

The solution depends on two parameters, i.e. the number of iterations n and λ. In our experiments we have used $n = 4$ because, empirically, four iterations are sufficient to stabilize the estimation and no significant changes have been noted with a higher number. The initial value of λ has been fixed to 20 as a sort of compromise between spatial regularity and data-driven estimates, which gives good results on the whole tested images. Nevertheless, in order to optimize the method, it would be interesting to be able to adjust λ according to image features.

Figures 5.9d, 5.9e and 5.9f show that the localization of the clusters $\mathcal{L}_2$ and $\mathcal{L}_4$ is robust to changes in patient morphotype and amount of signal. There is only a slight problem for some landmarks $l_j \in \mathcal{L}_2$ in Figure 5.9d that are located almost at the border and not at the center of the vertebrae, but, according to the analysis provided in Section 4.4.3, this should not affect the estimation of the EI value. Finally, note that the estimated positions on the diagnostic (Figure 5.9e) and the corresponding preview (Figure 5.9f) images are very similar.

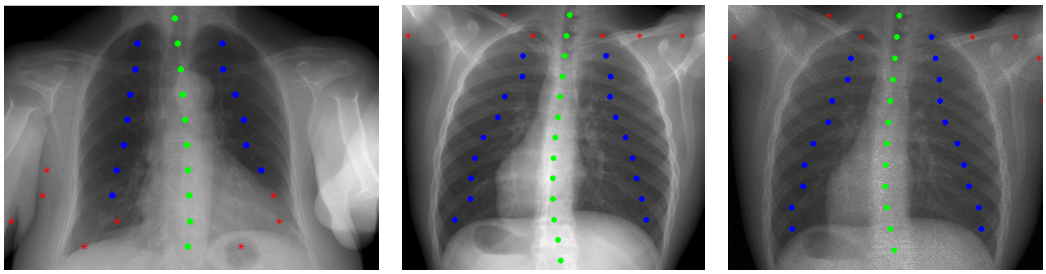
The landmarks in the anatomical ROI $\mathcal{A}_1$ do not need to be processed with the method proposed for the rest of the spine. Indeed, the landmarks $l_j \in \mathcal{L}_1$ are simply positioned on the vertical central axis extracted from the envelop of the head, i.e. over the horizontal line that passes across the vertebra T1. However, landmarks that are too distant from the control point T1 are neglected. Formally, only the landmarks l_j such that $\|x_j - x_{c1}\| < S$, x_{c1} being the x-coordinate of the control point for T1 are retained.

5.5.4 Detection of the landmarks in the lung

The definition of landmarks in the lungs is obtained from the sequence of steps presented in Section 5.5.1.1 that filters the points $p_j \in \mathcal{P}_{(l,128)}$ (Figure 5.5a) and provides a quite regular cluster (Figure 5.5b). However, it is preferable to avoid considering the landmarks associated with the heart or that are near the limits of the thoracic cage. As a consequence, a point p_j is taken into account only if the associated local EI estimate $e(p_j)$ respects the following condition:

$$10log_{10}\left(\frac{e(p_j)}{e_r}\right) + \tau > 0, \quad e_r = median\{e(p_j) \mid p_j \in \mathcal{P}\} \tag{5.16}$$

where τ is a threshold fixed at 1.5 that represents the acceptable (i.e. reduction of EI value) deviation with respect to the reference value.



(a) (b) (c)

Figure 5.10: *Results of the detection of lung cluster: (a) adult morphotype; (b) young morphotype;(c) preview of the exam in (b). The blue, red and green points correspond to $\mathcal{L}_3$, the points $p_j \in \mathcal{P}_{(h,128)}$ that are thresholded and $\mathcal{L}_2$, respectively.*

Figures 5.10a, 5.10b and 5.10c show good positioning of the landmarks $l_j \in \mathcal{L}_3$. We observe that some landmarks could be either missed, e.g. bottom left in Figure 5.10a, or considered while it should not be the case, e.g. top right in Figure 5.10c, which is due to the hard threshold. Nevertheless, given the robustness of the proposed method of estimation (Section 4.4.3), this problem is marginal.

5.5.5 Detection of the landmarks in the pelvis

The cluster $\mathcal{L}_5$ is modeled with a set of points regularly sampled on an ellipse of which the center position and the axis extensions need to be defined in order to solve the problem. The vertical axis is simply defined according to the positions of the control points for L5 and proximal femur and, hence, the clusters $\mathcal{L}_4$ and $\mathcal{L}_6$ help estimating the landmarks in the pelvis. On the other hand, the horizontal axis and the center are estimated from a local analysis of the content of the region. The coordinates of the center of the ellipse $o(x_c, y_c)$ are estimated as follows:

$$\begin{cases} x_c = \sum\limits_{p_j \in \mathcal{P}} x_j / |\mathcal{P}| \\ y_c = (y_3 + y_4)\,/2 \end{cases} \tag{5.17}$$

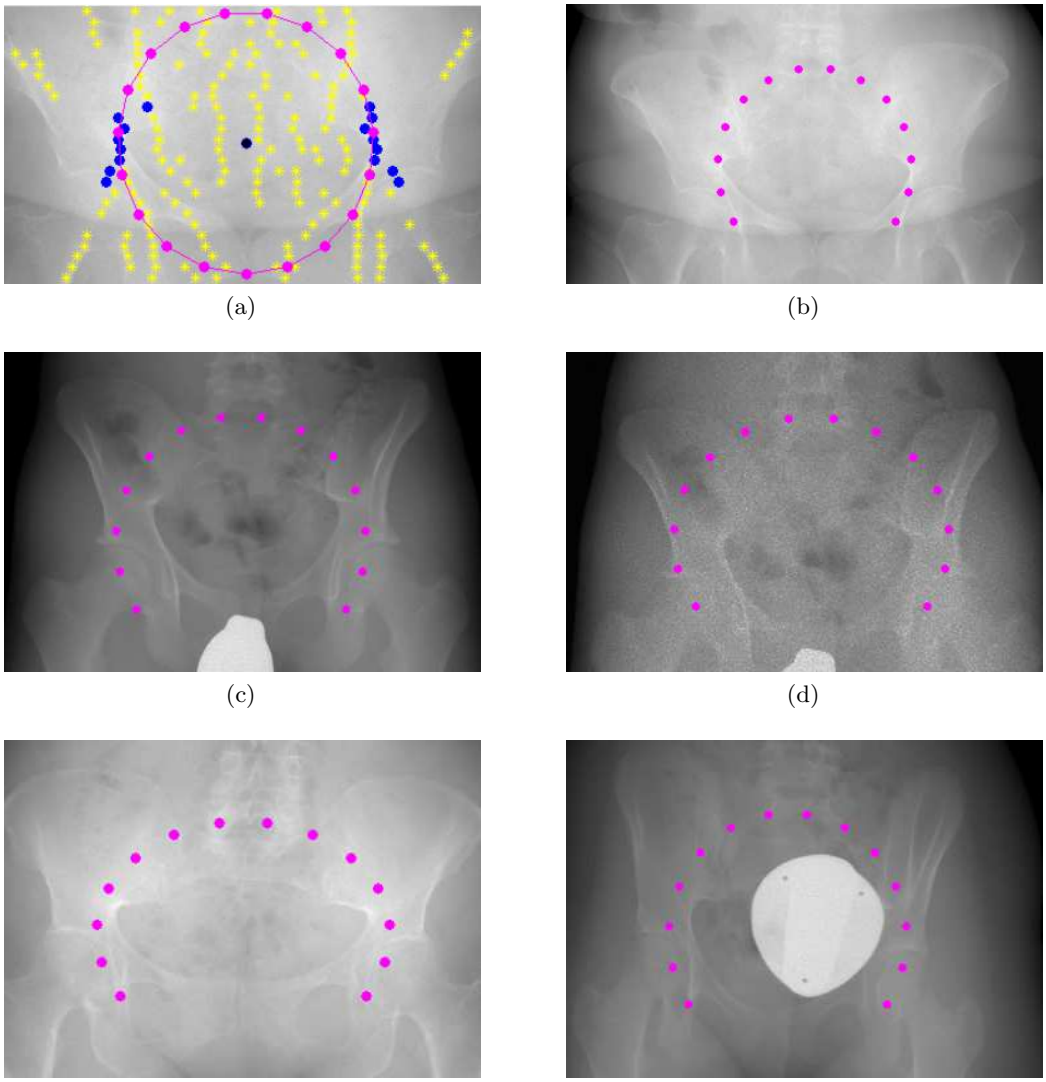
where the set $\mathcal{P}$ indicates the set of salient points $p_j \in \mathcal{P}_{(h,32)}$, and y_3 and y_4 are the y-coordinates of the control points for the vertebra L5 and the proximal femur, respectively.

Figure 5.11 shows the center of the ellipse o surrounded by the points $p_j \in \mathcal{P}_{(h,32)}$ that give an over-complete representation of the pelvis: some points are outliers, e.g. those inside the pelvis cavity, but many others correspond to bone structures. In particular, the leftmost and rightmost points near the center o match with the interior borders of the illium. In other exams, where the patient morphotype is smaller, the points may correspond to the exterior illium borders. The horizontal axis is assumed to be longer than the pelvis cavity while not going beyond the borders of the illium. From a subset of points that belong to $\mathcal{P}_{(h,32)}$ we can then estimate the extension of the horizontal axis of the ellipse. In practice, a subset $\mathcal{T}_0 \subset \mathcal{P}_{(h,32)}$ is defined by using Equation 5.10 and considering only a small area near the center $o(x_c, y_c)$, i.e. $p(x_j, y_f)$, $f = c-3, \dots c+3$. The length χ_x of the semi-horizontal axis of the ellipse is then defined as follows:

$$\chi_x = median\{d(o, t_j) \mid t_j \in \mathcal{T}_0\} \tag{5.18}$$

where $d(o, t_j)$ is the Euclidean distance of a point t_j from the center o of the ellipse. Once the parameters of the ellipse are estimated, a set of points $\mathcal{T}_1$ is defined by regularly sampling the ellipse. Figure 5.11 shows that some of the point $t_j \in \mathcal{T}_1$ in the lower part of the image correspond to soft rather than bone tissues. As a consequence, it is convenient to erase these landmarks. Formally, we consider the angles θ_l and θ_r between the vertical axis of the image and the vectors $\overrightarrow{ol(x_l, y_1)}$ and $\overrightarrow{ol(x_r, y_1)}$, respectively, where $l(x_l, y_1)$ and $l(x_r, y_1)$ are the uppermost landmarks belonging to the femur cluster. The angles θ_j between the vertical axis of the image and the vectors $\overrightarrow{ot_j}$, $t_j \in \mathcal{T}_1$, are also computed. The cluster $\mathcal{L}_5$ is then composed by the points $t_j \in \mathcal{T}_1$ that respect the condition $\theta_j \in [\theta_l, \theta_r]$.

Figure 5.11b gives an example of final result that well follows the shape of the pelvis on an adult exam. The landmarks detection is also good on smaller morphotypes, like for example the case in Figure 5.11c. Acceptable results are obtained on the corresponding preview too (Figure 5.11d). Nevertheless, the control point for the vertebra L5 is slightly higher than it should be and some outliers are then generated. In any case, these outliers only slightly influence the EI value estimation due to the homogeneity of the signal in this region, as discussed in Section 4.4.3. Finally,



(a) (b) (c) (d) (e) (f)

Figure 5.11: *An ellipse is fitted in the pelvis to find the cluster $\mathcal{L}_5$ by considering (a): the set $\mathcal{P}_{(h,32)}$ (yellow points), the center of the region $o(x_c, y_c)$ (black point), the set $\mathcal{T}_0 \subset \mathcal{P}_{(h,32)}$ retained to estimate the extension of the horizontal axis of the ellipse (blue points) and the resulting landmarks sampled on the ellipse (magenta points). Results: (b) adult morphotype; (c) young morphotype; (d) preview of the exam in (c); (d) ellipse such that $\chi_x > \chi_y$; (e) young morphotype with a metallic object in the field of view.*

note that the method adapts well to different shapes of the pelvis, e.g. length χ_x of the semi-horizontal axis of the ellipse bigger than the length χ_y of the semi-vertical one (Figure 5.11e), and is not affected by presence of metallic objects in the field of view (Figure 5.11f).

5.6 Landmarks detection and recognition on lateral view acquisitions

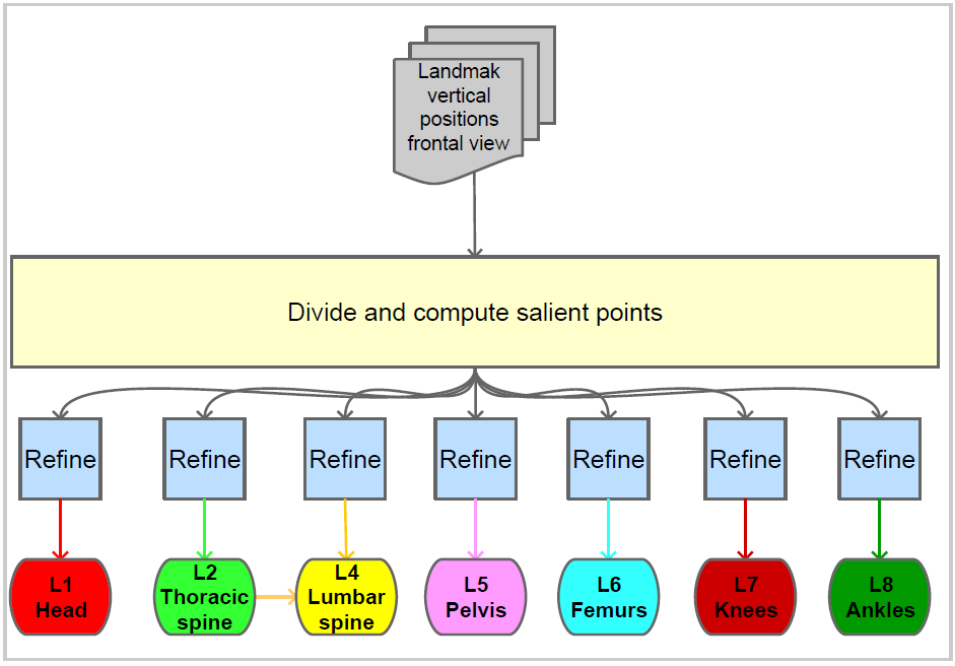
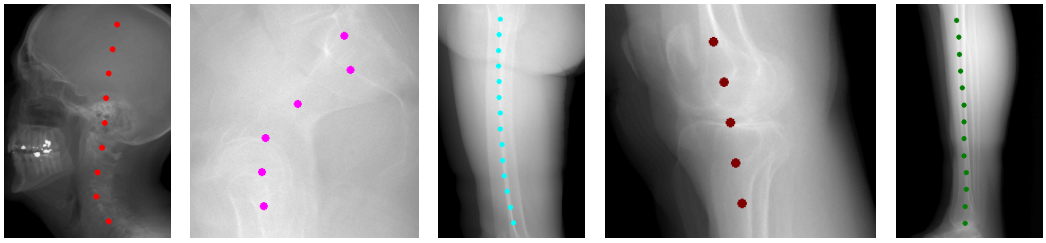


Figure 5.12: *Process of detection and recognition on the lateral view acquisitions. The results from the corresponding frontal view are used to separate the field of view into groups of anatomical regions. The landmarks $l_j \in \mathcal{L}_r$ are then refined with local analysis.*

The anatomical clusters of landmarks have also to be identified on the lateral view in order to fully estimate the image quality for a full-body examination. The considered classes are always the same. However, the EI in lungs is not evaluated because, from the lateral point of view, this region is not clearly distinguishable due to the superposition with other anatomical structures such as the heart and the thoracic vertebrae.

The method of detection and recognition on the lateral view images follows similar ideas as the ones used on the frontal views. Indeed, the approach is also based on the combination of global and local analysis and sequential narrowing of the solution space. Figure 5.12 shows that the control points defined on the frontal view acquisition are directly used on the lateral one as well, according to the discussion in Section 5.3. This drastically eases the process because the landmarks can be straightly labeled with the corresponding anatomical regions. As a consequence, the positions of the landmarks are independently estimated within each associated anatomical ROI $\mathcal{A}_r$, except for the cluster for the lumbar spine ($\mathcal{L}_4$) which exploits the landmarks in the thoracic spine ($\mathcal{L}_2$) too.

We discuss in the same section how the clusters $\mathcal{L}_1$ (Head) and $\mathcal{L}_{5-8}$ (pelvis and legs) are determined because their definitions rely on similar simple ideas. On the other hand, more detailed descriptions are dedicated to the landmarks $l_j \in \mathcal{L}_2$ and



(a) (b) (c) (d) (e)

Figure 5.13: *Detection of the landmarks l_j on lateral view images for the anatomical clusters: (a) $\mathcal{L}_1$; (b) $\mathcal{L}_5$; (c) $\mathcal{L}_6$; (d) $\mathcal{L}_7$; (e) $\mathcal{L}_8$.*

$l_j \in \mathcal{L}_4$ that are more challenging to detect.

5.6.1 Detection of the landmarks in the head, pelvis and legs

The landmarks $l_j \in (\mathcal{L}_1 \cup \mathcal{L}_5 \cup \mathcal{L}_6 \cup \mathcal{L}_7 \cup \mathcal{L}_8)$ are all well described by the salient points $\mathcal{P}_{(h,s)}$, and, hence, their final positions are estimated with a limited number of easy steps.

Head and cervical spine. Once the salient points $p_j \in \mathcal{P}_{(h,128)}$ are computed, the subset $\mathcal{T} \subset \mathcal{P}_{(h,128)}$ is defined by considering one point for each horizontal line $y = y_f,\ f = 1, \ldots, |\mathcal{P}_y|$. Formally, the x-coordinate of a point $t_e \in \mathcal{T}$ is defined as follows:

$$x_e = \begin{cases} \min\{x_j \mid p_j \in \mathcal{P}_{y_f}\} & \text{if orientation PA} \\ \max\{x_j \mid p_j \in \mathcal{P}_{y_f}\} & \text{if orientation AP} \end{cases} \tag{5.19}$$

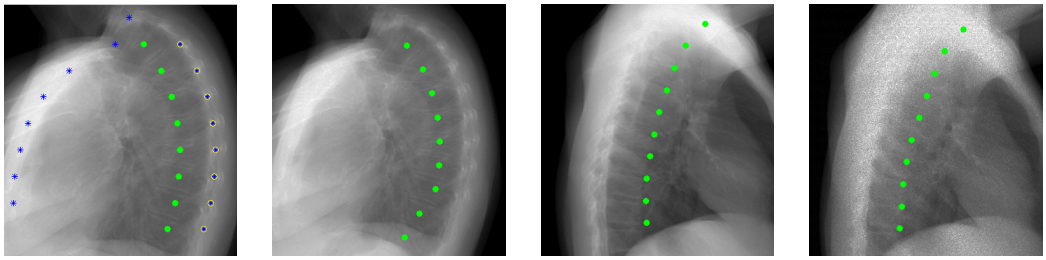
where a lateral view image is PA or AP oriented if the back of the patient is in the left- or right-hand side of the field of view, respectively. The cluster $\mathcal{L}_1$ is then composed by the points $t_j \in \mathcal{T}$ with x-coordinates $x_j \in [x_r - S, x_r + S]$, with x_r equal to $median\{x_j \mid t_j \in \mathcal{T}\}$. Note that in some cases the number of landmarks $|\mathcal{L}_1|$ could be very low because only a part of the head is acquired and is bent forward. We then consider this cluster only if $|\mathcal{L}_1| > 3$. Figure 5.13a shows an example of automatically detected landmarks $l_j \in \mathcal{L}_1$.

Pelvis. The landmarks $l_j \in \mathcal{L}_5$ simply correspond to the salient points $p_j \in \mathcal{P}_{(h,128)}$ that are sufficient to describe the pelvis from the lateral point of view. In some particular cases, e.g. in presence of gonad shields, $\mathcal{P}_{(h,128)}$ may present some outliers. However, these outliers can be simply removed by considering the other detected points. Formally, if, for a given y-coordinate y_f, $|\mathcal{P}_{y_f}| > 1$, the retained point $p_j \in \mathcal{P}_{y_f}$ is the nearest one to the points that lie on nearby lines y_i for which $|\mathcal{P}_{y_i}| = 1$. Figure 5.13b shows an example of automatically detected landmarks $l_j \in \mathcal{L}_5$.

Legs. The landmarks in the leg are selected among the salient points $p_j \in \mathcal{P}_{(h,128)}$ by selecting two points on each line $y = y_j$ according to Equation 5.10. If there is only one point p_j on the horizontal line $y = y_j$, then p_j is simply considered as a landmark l_j. Figures 5.13c, 5.13d and 5.13e show some examples of automatically detected landmarks $l_j \in \mathcal{L}_6$, $l_j \in \mathcal{L}_7$ and $l_j \in \mathcal{L}_8$, respectively. As discussed in the results (Section 5.7.2), the quality of the estimation of the EI values in these ROIs

highly depends on the posture of the patient and, in particular, whether the legs are superposed or not in the image.

5.6.2 Detection of the landmarks in the thoracic spine



(a) (b) (c) (d)

Figure 5.14: *Detection of the landmarks $l_j \in \mathcal{L}_2$ on lateral view images (a): a subset of points $t_j \in \mathcal{T}_0$ (yellow circle) is selected among the salient points $p_j \in \mathcal{P}_{(h,128)}$ to roughly capture the posterior side of the thoracic cage. Results: (b) adult morphotype; (c) young morphotype; (d) preview of the exam in (c).*

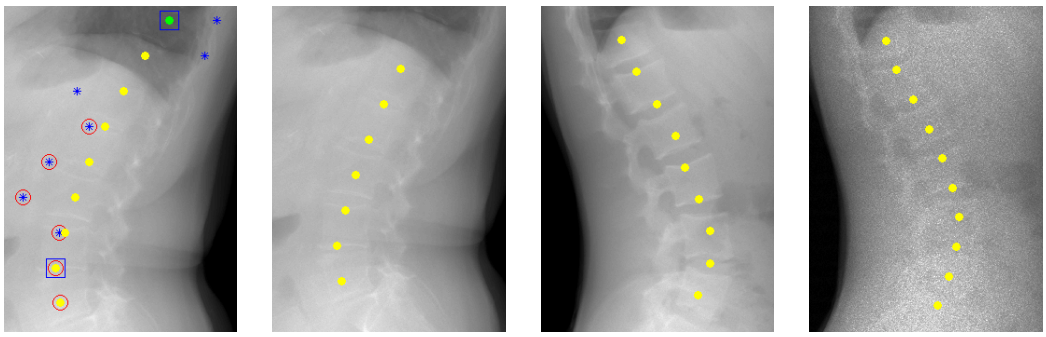
We consider a sub-image $\mathbf{k}$ that contains the chest according to the positions of the control points for the vertebrae T1 and T12. The vertebrae in the thoracic spine cannot be directly associated with salient points. However, the landmarks $l_j \in \mathcal{L}_2$ can be inferred from the thoracic cage that is well represented by the points $p_j \in \mathcal{P}_{(h,128)}$, i.e. peaks of absorption. According to our tests, while the arms of the patient may make the detection of the anterior part of the thoracic cage difficult, the posterior part is robustly identified independently from the posture of the patient. Figure 5.14a shows the set of points $p_j \in \mathcal{P}_{(h,128)}$ and a subset $\mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ (yellow circles). The points $t_j \in \mathcal{T}_0$ correspond to the posterior part of the thoracic cage and are formally defined by using Equation 5.19. Some outliers may be present among the points $t_j \in \mathcal{T}_0$. Nevertheless, this problem is rare and the correctness of a point $t_j \in \mathcal{T}_0$ can be easily assessed by measuring the spatial distance from nearby points. Formally, a point t_j is retained only if its x-coordinate respects the condition $|x_j - x_{j+1}| < S/2$, $S = 128$. Potential holes are then filled with linear interpolation.

The landmarks $l_j \in \mathcal{L}_2$ are defined by shifting the points $t_j \in \mathcal{T}_0$ towards the center of the body while preserving their y-coordinates, i.e. the movements are perpendicular to the vertical axis of the image. At each shift a new set $\mathcal{T}_1$ is obtained and the associated EI value is computed by using Equation 5.1. The cluster $\mathcal{L}_2$ is then defined as the set $\mathcal{T}_1$ whose EI estimate is the first local maximum because in the region in the proximity of the thoracic cage the X-ray absorption is higher than in the thoracic vertebrae. Therefore, as the set of points $\mathcal{T}_1$ depart from the thoracic cage, the associated EI value starts increasing. However, at a certain distance, the EI value starts decreasing again due to superposition of different tissues, e.g. the heart. Moreover, by selecting the first peak of EI value, we constraint the solution to belong to the neighborhood of $\mathcal{T}_0$, which enforces to respect the condition of the thoracic vertebrae to be posteriorly connected to the ribs. Figure 5.14a shows the points $t_j \in \mathcal{T}_1$ (green points) that result from the described process. Figure 5.14b

displays the final landmarks $l_j \in \mathcal{L}_2$ after the refinement of the position of the control point for T12, which is discussed in Section 5.6.4.

Figures 5.14b, 5.14c and 5.14d prove that by means of the proposed method the landmarks follow the arching of the spine regardless the posture of the patient. Besides, the detection is also robust to changes in the position of the arms, morphotype and radiation exposure. For the example in Figure 5.14c, the landmarks $l_j \in \mathcal{L}_2$ do not perfectly correspond to the centers of the vertebrae, and in the upper part of the spine the superposition with the shoulders generate some outliers. The latter aspect is the most important of the two and the contribution from these problematic points should be reduced by giving them lower weights $\omega(l_j)$. However, this option has not been implemented in the tested method as this issue has been observed only after the analysis of the results.

5.6.3 Detection of the landmarks in the lumbar spine



(a) (b) (c) (d)

Figure 5.15: *Detection of the landmarks $l_j \in \mathcal{L}_4$ on lateral view acquisitions (a): a subset $\mathcal{T}_0 \subset Pi_{(h,128)}$ (red circles) of the salient points (blue stars) is selected according to the geometrical model (Section 5.3.2). The most prominent point of the lordotic curve is then detected (blue square) and, with T12 and L5, used to estimate the arching of the spine. The results are then obtained by means of spline interpolation: (b) adult morphotype; (c) young morphotype; (d) preview of the exam in (c).*

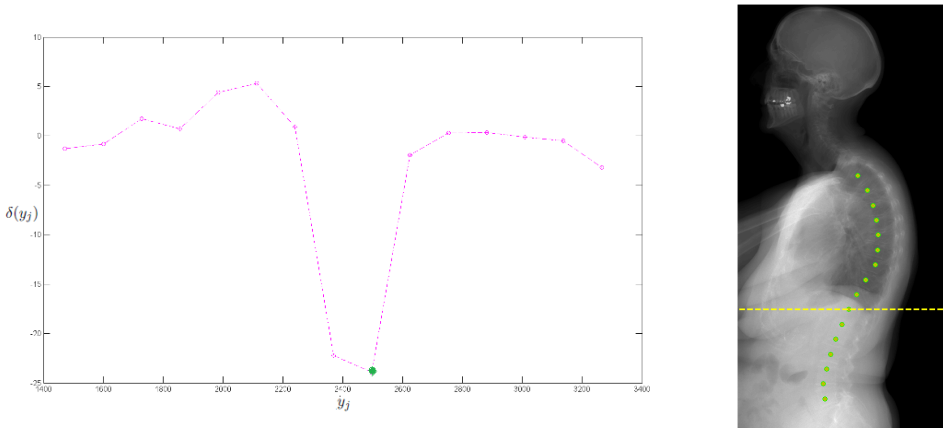
The definition of the landmarks $l_j \in \mathcal{L}_4$ is difficult because, as pointed out in the discussion of analysis of the pelvis on frontal view acquisitions, in regions where the absorption is very strong the signal is approximately homogeneous. Therefore, the salient points $p_j \in \mathcal{P}_{(h,s)}$ do not necessarily capture the lumbar spine. Figure 5.15a shows with an example that the superior part of the lumbar spine is particularly affected by this issue, whereas in the inferior part the points $p_j \in \mathcal{P}_{(h,128)}$ correspond quite well to the lumbar vertebrae.

The detection of the landmarks $l_j \in \mathcal{L}_4$ is based on the identification of three key landmarks that are: the vertebrae T12, the most prominent point of the lordotic curve and the vertebra L5. The vertebrae T12 is obtained from the cluster $\mathcal{L}_2$, the vertebra L5 is simply the nearest landmark to the bottom of the image and the most prominent point is defined by analyzing the positions of the points $p_j \in \mathcal{P}_{(h,128)}$. The most prominent point will correspond to the position of one among the local

minima or maxima of $\mathcal{P}_x$ depending on the orientation. The smallest peak is chosen because the transition is assumed to be smooth, i.e. the most prominent point and its neighbors are normally close, and the other peaks are considered as outliers. In the provided example (Figure 5.15a), there are two possible candidates for the most prominent point and the nearest one to the bottom of the image is chosen according to the aforementioned principle. Note that only a subset $\mathcal{T}_0 \subset \mathcal{P}_{(h,128)}$ is taken into account for this analysis (red circles in Figure 5.15a). The points $t_j \in \mathcal{T}_0$ are selected as a function of the prior geometrical position of the most prominent point with respect to T12 and L5 as discussed in Section 5.3.2. Finally, the output landmarks $l_j \in \mathcal{L}_4$ are obtained by spline interpolation of the three key points.

Figures 5.15b, 5.15c and 5.15d show that the detection of the landmarks $l_j \in \mathcal{L}_4$ is robust to changes in patient morphotype and signal strength.

5.6.4 Refinement of the control point for the vertebra T12



(a) (b)

Figure 5.16: *By analyzing how the local EI value decreases from a landmark l_j to the next one l_{j+1} (a), the control point for T12 is refined (b).*

The position of the control point for the vertebra T12 (c_2) can be refined by analyzing the landmarks $l_j \in (\mathcal{L}_2 \cup \mathcal{L}_4)$. The passage from $\mathcal{L}_2$ to $\mathcal{L}_4$ is marked by a strong reduction of the local EI estimate due to a higher density in the lumbar region than in the chest. Therefore, the local EI estimates $e(l_j)$ are computed at the landmarks $l_j \in (\mathcal{L}_2 \cup \mathcal{L}_4)$, and the following function is defined from these values:

$$\delta(y_j) = e(l_{j+1}) - e(l_j), \quad j = 1, \ldots, (|\mathcal{L}_2 \cup \mathcal{L}_4| - 1) \tag{5.20}$$

where $\delta(.)$ depends on y_j because there is one point for each line in the set $\mathcal{L}_2 \cup \mathcal{L}_4$. Figure 5.16a shows the function $\delta(y_j)$. Since each element of the function is associated with the drop of local EI value that occurs in the following one, the y-coordinate of the control point for T12 is estimated as follows:

$$y = \underset{y_j \in Y_m}{\arg\min}(\delta(y_j)) \tag{5.21}$$

where Y_m are the values that constrain the position of $y = y_2$ according to the model of global relations (see Table 5.2). Figure 5.16b displays how the landmarks

$l_j \in (\mathcal{L}_2 \cup \mathcal{L}_4)$ are then separated into two clusters, i.e. thoracic and lumbar spine. The labels of the landmarks in the spine on the corresponding frontal view acquisition are eventually updated after this analysis.

5.7 Evaluation of automatically computed exposure index values

We evaluate the method explained in the previous sections according to the DI values computed as in Equation 4.7, which assesses the accuracy of the automatically estimated EI values with respect to those computed from manually annotated landmarks. The precision in terms of positions of the landmarks is not discussed here because, as justified in Section 4.4.3, the proposed landmark-based approach allows for a certain degree of localization error. Please refer to Appendix B for a qualitative evaluation of the accuracy in terms of locations.

The database is composed by 82 full-body examinations and, in particular, the following subsets can be distinguished:

- $\mathcal{D}_a$: 24 diagnostic exams of adult patients aged between 54 and 82.
- $\mathcal{D}_b$: 29 diagnostic exams of young patients aged between 8 and 16.
- $\mathcal{D}_c$: 29 preview exams of the patients in the subset $\mathcal{D}_b$.

For each category of age the size of the patients varies from normal to obese, some acquisitions follow surgery, i.e. there are metallic objects in the field of view, the patients are not necessarily centered with respect to the image plane, and there are some pathological cases (e.g. scoliosis).

Sections 5.7.1 and 5.7.2 analyze the results corresponding to frontal and lateral view acquisitions, respectively. The descriptions are conducted by referring to the anatomical ROIs $\mathcal{A}_r$. The accuracy of the estimated EI values is evaluated according to the RMSE defined in Equation 4.8 and to the percentages ε_0, ε_1, ε_2, ε_3 and ε_4 that indicate the amount of samples in the database whose estimation errors are negligible, low, medium, high and severe (see Section 4.3.2), respectively, when compared with the ground truth values. Moreover, we will use $\varepsilon_0+\varepsilon_1$ as a measure of the amount of acceptable errors. Furthermore, specific observations are made on the three subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$ in order to understand if the quality of the estimation may change depending on the age of the patient or the amount of radiation exposure.

Sections 5.7.3 and 5.7.4 deal with cases that, according to the state of the art, make it difficult to obtain consistent EI measures, i.e. presence of metallic objects and particular conditions of the patient, respectively.

The proposed method is very efficient as the analysis is limited to some salient points that are sparsely distributed on the pixel space: by running our MATLAB code on a conventional machine (Intel Core 2.20 GHz, 4 GB RAM), it takes about 6.50 s and 3.50 s at most to detect and recognize the landmarks on both frontal and lateral view full-body acquisitions, respectively, i.e. a total of only about 10 s.

Our approach can be quantitatively compared to the results from the state of the art methods presented in Section 4.3.2 and, hence, the discussion conducted in this section completes the validation from the previous chapter. Moreover, we

consider another detection technique that consists in a landmark-based formulation of the method sROI (Section 4.3.1.2), where the landmarks are labeled depending on the control point positions, but they are detected according to the borders of the patient envelop. We denote these landmarks by $b_j \in \mathcal{B}_r$, $r = 1, \ldots, 8$ and they are obtained by regularly sampling the points that lie on the central axis of the envelop. All the landmarks are located according to this principle except those in the lungs ($b_j \in \mathcal{B}_3$) that are placed in the middle between the thoracic spine ($b_j \in \mathcal{B}_2$) and the borders of the patients (see Appendix A for details and examples). The EI values computed at the landmarks $l_j \in \mathcal{L}_r$ and $b_j \in \mathcal{B}_r$ are not compared in Section 5.7.3 because the problem caused by the presence of metallic objects does not specifically concern the position of the landmarks. For example, a landmark in the prosthetic femoral head is well placed, but the associated EI estimate does not correspond to significant X-ray absorption.

5.7.1 Quantitative evaluation on frontal view acquisitions

Table 5.3: *Evaluation on frontal view acquisitions in the whole database: comparison of the automatic EI estimates associated with the clusters $\mathcal{B}_r$ and $\mathcal{L}_r$, $r = 1, \ldots 8$, in terms of RMSE and percentages of negligible (ε_0), low (ε_1), medium (ε_2), high (ε_3) and severe (ε_4) estimation errors.*

Frontal		RMSE	ε_0	ε_1	ε_2	ε_3	ε_4
$\mathcal{A}_1$	$\mathcal{B}_1$	0.32	65.9%	25.6%	4.9%	1.2%	2.4%
	$\mathcal{L}_1$	**0.29**	67.1%	26.6%	4.9%	1.2%	1.2%
$\mathcal{A}_2$	$\mathcal{B}_2$	0.61	20.7%	31.7%	28.0%	13.4%	6.1%
	$\mathcal{L}_2$	**0.30**	61.0%	32.9%	4.9%	1.2%	0.0%
$\mathcal{A}_3$	$\mathcal{B}_3$	0.56	45.1%	30.5%	15.9%	3.7%	4.9%
	$\mathcal{L}_3$	**0.27**	85.4%	9.8%	2.4%	1.2%	1.2%
$\mathcal{A}_4$	$\mathcal{B}_4$	0.11	96.3%	3.7%	0.0%	0.0%	0.0%
	$\mathcal{L}_4$	**0.09**	97.6%	2.4%	0.0%	0.0%	0.0%
$\mathcal{A}_5$	$\mathcal{B}_5$	0.35	73.2%	18.3%	4.9%	1.2%	2.4%
	$\mathcal{L}_5$	**0.25**	86.6%	8.5%	2.4%	1.2%	1.2%
$\mathcal{A}_6$	$\mathcal{B}_6$	0.68	39.0%	36.6%	8.5%	3.7%	12.2%
	$\mathcal{L}_6$	**0.19**	86.6%	13.4%	0.0%	0.0%	0.0%
$\mathcal{A}_7$	$\mathcal{B}_7$	0.16	65.9%	25.6%	4.9%	1.2%	2.4%
	$\mathcal{L}_7$	**0.12**	85.4%	14.6%	0.0%	0.0%	0.0%
$\mathcal{A}_8$	$\mathcal{B}_8$	0.09	100.0%	0.0%	0.0%	0.0%	0.0%
	$\mathcal{L}_8$	**0.06**	100.0%	0.0%	0.0%	0.0%	0.0%

Tables 5.3 and 5.4 report the results corresponding to all the frontal view acquisitions in the database and to the three subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$, respectively.

5.7.1.1 Head and cervical spine

Since the landmarks $b_j \in \mathcal{B}_1$ and $l_j \in \mathcal{L}_1$ are defined in the same way except for a control on the distance from T1 (see Section 5.5.3), the EI values computed from the two sets are very similar. The results obtained with the proposed method are good

Table 5.4: *Evaluation over the frontal view acquisitions in the subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$: comparison of the automatic EI estimates associated with the clusters $\mathcal{B}_r$ and $\mathcal{L}_r$, $r = 1, \dots 8$ in terms of RMSE.*

Frontal		RMSE		
		$\mathcal{D}_a$	$\mathcal{D}_b$	$\mathcal{D}_c$
$\mathcal{A}_1$	$\mathcal{B}_1$	0.11	0.36	0.37
	$\mathcal{L}_1$	0.11	0.33	0.35
$\mathcal{A}_2$	$\mathcal{B}_2$	0.69	0.62	0.51
	$\mathcal{L}_2$	0.28	0.32	0.28
$\mathcal{A}_3$	$\mathcal{B}_3$	0.33	0.60	0.67
	$\mathcal{L}_3$	0.28	0.19	0.33
$\mathcal{A}_4$	$\mathcal{B}_4$	0.11	0.10	0.13
	$\mathcal{L}_4$	0.06	0.09	0.10
$\mathcal{A}_5$	$\mathcal{B}_5$	0.17	0.46	0.32
	$\mathcal{L}_5$	0.15	0.16	0.36
$\mathcal{A}_6$	$\mathcal{B}_6$	0.59	0.78	0.65
	$\mathcal{L}_6$	0.19	0.18	0.19
$\mathcal{A}_7$	$\mathcal{B}_7$	0.12	0.10	0.22
	$\mathcal{L}_7$	0.09	0.06	0.17
$\mathcal{A}_8$	$\mathcal{B}_8$	0.09	0.07	0.10
	$\mathcal{L}_8$	0.07	0.04	0.08

as shown by the RMSE equal to 0.29, and by 67.1% and 93.7% of samples whose estimation errors are negligible and acceptable, respectively. The accuracy is higher on $\mathcal{D}_a$ (RMSE equal to 0.11) than on $\mathcal{D}_b$ (RMSE equal to 0.33). This can be related to the position of the head for the patients in the databases $\mathcal{D}_b$ and $\mathcal{D}_c$, which is bent forward by hence leading to a partial superposition of the head with the neck. As it will become more evident from the analysis on the lateral view acquisitions (Section 5.7.2), the partial occlusion of an anatomical structure by another one can cause some problems for the estimation of the EI values. Nevertheless, a RMSE error equal to 0.33 is still a low error and, thus, this issue does not quantitatively concern this anatomical ROI. Finally, note that the strong reduction of radiation exposure does not affect the measure as the RMSE values associated with $\mathcal{D}_b$ and $\mathcal{D}_c$ are similar.

5.7.1.2 Thoracic spine

The accuracy of the automatic EI value in the anatomical ROI $\mathcal{A}_2$ significantly changes depending on which cluster of landmarks is used. The measure computed at the landmarks $b_j \in \mathcal{B}_2$ produces a RMSE equal to 0.61, and 20.7% and 52.4% of samples whose estimation errors are negligible and acceptable, respectively. These results are due to the misplacement of the thoracic spine with respect to the central axis that may be caused, for example, by a scoliosis (Section 5.7.4) or because the patient is side-bending. Moreover, this issue is not limited to a particular type of images as indicated by the RMSE values equal to 0.69, 0.62 and 0.51 for $\mathcal{D}_a$, $\mathcal{D}_b$

and $\mathcal{D}_c$, respectively. However, the performances are slightly better on the preview images because the signal levels in the thoracic spine and in the lungs are more similar than on diagnostic EOS acquisitions. Therefore, the contributions from the lung tissues that surround the thoracic spine bias slightly less the EI measure. On the other hand, the proposed method accounts for such cases that normally occur in clinical routine and allows lowering the RMSE to 0.30 and increasing the percentages of negligible and acceptable errors up to 61.0% and 93.9%, respectively. Moreover, the same quality of results is obtained on the subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$.

5.7.1.3 Lungs

The analysis of the results associated with the ROI $\mathcal{A}_3$ leads to observations similar to those corresponding to the ROI $\mathcal{A}_2$. Indeed, as the landmarks $b_j \in \mathcal{B}_3$ are inferred from the positions of $b_j \in \mathcal{B}_2$, the EI measures give a RMSE equal to 0.56, and 45.1% and 75.6% of samples whose estimation errors are negligible and acceptable, respectively. However, the results are slightly better because the lungs occupy a wider area than the thoracic spine, but this depends on the morphotype as justifiable by comparing adults ($\mathcal{D}_a$) to young patients ($\mathcal{D}_b$). The landmarks $l_j \in \mathcal{L}_3$ give a lower RMSE (0.27) and higher percentages of negligible (85.4%) and acceptable (95.2%) errors. By analyzing the performances in the subsets, it can be observed that the accuracy is higher on $\mathcal{D}_b$ (RMSE equal to 0.19) than on the corresponding preview acquisitions (RMSE equal to 0.33). This is due to a particular exam in which the arms of the patient are not well placed and partially cover the lungs by then generating DI values equal to -0.77 and -1.76 on the diagnostic and preview acquisitions, respectively. The performances are different depending on the radiation exposure because the lower the dose, the more problematic the superposition with thin structures is. This singular case should not be retained because the estimation of the EI value in the ROI $\mathcal{A}_3$ is not relevant if the region is covered by the arms. Nevertheless, this confirms that the automatic EI estimation is unstable if the anatomical object of interest is partially occluded by other structures, as highlighted in the former analysis on the ROI $\mathcal{A}_1$.

5.7.1.4 Lumbar spine

The ROI $\mathcal{A}_4$ is one of the easiest to describe in terms of EI and there is no significant difference depending on whether $b_j \in \mathcal{B}_4$ or $l_j \in \mathcal{L}_4$ are used. Quantitatively, the landmarks $b_j \in \mathcal{B}_4$ produce a RMSE equal to 0.11, and 96.3% and 100% of samples whose estimation errors are negligible and acceptable, respectively. The landmarks $l_j \in \mathcal{L}_4$ give practically the same results as the RMSE is equal to 0.09, and there are 97.6% and 100% of samples whose estimation errors are negligible and acceptable, respectively. The results computed over the subsets are coherent with those corresponding to the whole database.

5.7.1.5 Pelvis

In the pelvic region, even if some landmarks $b_j \in \mathcal{B}_5$ are located in the pelvic cavity, the results are quite good. The RMSE is equal to 0.35, and there are 73.2% and 91.5% of samples whose estimation errors are negligible and acceptable, respectively. Nevertheless, the landmarks $l_j \in \mathcal{L}_5$ allow for a slight improvement given

the RMSE equal to 0.25, and 86.6% and 95.1% of samples whose estimation errors are negligible and acceptable, respectively. For the exams in the subset $\mathcal{D}_a$, there is practically no difference between the two methods as on adult morphotypes the signal in the ROI $\mathcal{A}_5$ is quite homogeneous and, thus, the difference of intensity levels between bone and soft tissues is minimal. The same assessment is valid for $\mathcal{D}_c$, whereas for $\mathcal{D}_b$, as the amount of dose is higher with respect to $\mathcal{D}_c$ and smaller morphotype with respect to $\mathcal{D}_a$, the two intensity levels associated with soft and bone tissues in the pelvis can be distinguished and, hence, the proposed method performs better. Finally, note that the RMSE in $\mathcal{D}_c$ computed at the landmarks $l_j \in \mathcal{L}_5$ is equal to 0.36 due to one high error and one severe error. In these cases, the position of the control point L5 is not well estimated because the proximal femur is slightly shifted from its ground truth location and there is no local analysis to refine the position of the control point for L5 (see Section 5.5.1.3). The parameters of the ellipse are then wrongly estimated and many of the landmarks $l_j \in \mathcal{L}_5$ are not correctly placed. However, note that, according to our tests, this case is very rare and strictly limited to preview acquisitions.

5.7.1.6 Femurs

The proposed method significantly increases the accuracy of the EI estimation in the ROI $\mathcal{A}_6$. Indeed, by considering the landmarks $b_j \in \mathcal{B}_6$ the RMSE is equal to 0.68, and there are 39.0% and 75.6% of samples whose estimation errors are negligible and acceptable, respectively. On the other hand, the landmarks $l_j \in \mathcal{L}_6$ lower the RMSE value to 0.19 and increase up to 86.6% and 100% the amount of samples whose estimation errors are negligible and acceptable. Similar considerations can be made by analyzing the subsets. The bad performances obtained when the EI value is computed at the landmarks $b_j \in \mathcal{B}_6$, are due to a high body mass index as detailed in Section 5.7.4.

5.7.1.7 Knees

In the ROI $\mathcal{A}_7$ the difference between the two methods is very low and the performances are independent from the morphoype and the dose. Quantitatively, the landmarks $b_j \in \mathcal{B}_7$ give a RMSE equal to 0.16, and 65.9% and 91.4% of samples whose estimation errors are negligible and acceptable, respectively. The landmarks $l_j \in \mathcal{L}_7$ slightly improve the performances as the RMSE is equal to 0.12, and there are 85.4% and 100% of samples whose estimation errors are negligible and acceptable, respectively.

5.7.1.8 Tibiae

The estimation of the EI value associated with the ROI $\mathcal{A}_8$ is identical if the clusters $\mathcal{B}_8$ or $\mathcal{L}_8$ are considered because the tibia is at the center of the leg. Moreover, this region is the easiest one to process given the very low RMSE (0.09 and 0.06) and the 100% of samples whose estimation errors are negligible.

5.7.1.9 Global evaluation

The estimation of the sets of EI values on frontal view acquisitions with the proposed approach is very robust and overcomes some drawbacks that manifest when the landmarks are detected simply depending on the position with respect to the borders of the patient envelop. The improvement is particularly significant in the anatomical ROIs $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_6$, that are hence confirmed as the regions most affected by errors in the landmark positions as pointed out in Section 4.4.3. Moreover, the results indicate that the accuracy does not significantly change if the morphotype varies or if the EI values are computed over preview images.

5.7.2 Quantitative evaluation on lateral view acquisitions

Table 5.5: *Evaluation on lateral view acquisitions in the whole database, of the accuracy of the automatic EI estimates associated with the clusters $\mathcal{B}_r$ and $\mathcal{L}_r$, $r = 1, \dots 8$, in terms of RMSE and percentages of cases whose estimation errors are negligible (ε_0), low (ε_1), medium (ε_2), high (ε_3) and severe (ε_4).*

Lateral		RMSE	ε_0	ε_1	ε_2	ε_3	ε_4
$\mathcal{A}_1$	$\mathcal{B}_1$	0.68	62.7%	17.6%	5.9%	3.9%	9.8%
	$\mathcal{L}_1$	**0.57**	66.7%	17.6%	2.0%	5.9%	7.8%
$\mathcal{A}_2$	$\mathcal{B}_2$	0.89	23.2%	18.3%	19.5%	13.4%	25.6%
	$\mathcal{L}_2$	**0.64**	41.5%	17.1%	13.4%	14.6%	13.4%
$\mathcal{A}_4$	$\mathcal{B}_4$	0.16	90.2%	7.3%	2.4%	0.0%	0.0%
	$\mathcal{L}_4$	0.16	86.6%	13.4%	0.0%	0.0%	0.0%
$\mathcal{A}_5$	$\mathcal{B}_5$	0.28	72.0%	22.0%	2.4%	2.4%	1.2%
	$\mathcal{L}_5$	**0.15**	95.1%	3.7%	0.0%	1.2%	0.0%
$\mathcal{A}_6$	$\mathcal{B}_6$	0.44	50.0%	28.0%	14.6%	4.9%	2.4%
	$\mathcal{L}_6$	**0.42**	62.2%	22.0%	7.3%	4.9%	3.7%
$\mathcal{A}_7$	$\mathcal{B}_7$	0.75	25.6%	18.3%	17.1%	19.5%	19.5%
	$\mathcal{L}_7$	**0.58**	36.6%	22.0%	20.7%	11.0%	9.8%
$\mathcal{A}_8$	$\mathcal{B}_8$	0.47	56.1%	23.2%	13.4%	1.2%	6.1%
	$\mathcal{L}_8$	**0.45**	58.5%	20.7%	13.4%	1.2%	6.1%

Tables 5.5 and 5.6 report the results associated with all the lateral view acquisitions in the database and the three subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$, respectively.

5.7.2.1 Head and cervical spine

Both clusters $\mathcal{B}_1$ and $\mathcal{L}_1$ provide average estimations of the EI value associated with $\mathcal{A}_1$. The landmarks $b_j \in \mathcal{B}_1$ produce a RMSE equal to 0.68, and 62.7% and 80.3% of samples whose estimation errors are negligible and acceptable, respectively. The landmarks $l_j \in \mathcal{L}_1$ slightly improve the performances with an RMSE equal to 0.57, and 66.7% and 84.3% of samples whose estimation errors are negligible and acceptable, respectively. Nevertheless, it is possible to remark that the results are very good, on the subset $\mathcal{D}_a$ (RMSE equal to 0.12) and high errors are only found in the subsets $\mathcal{D}_b$ and $\mathcal{D}_c$. This is not due to the change of morphotype but rather to

Table 5.6: *Evaluation on lateral view acquisitions in the subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$, of the accuracy of the automatic EI estimates associated with the clusters $\mathcal{B}_r$ and $\mathcal{L}_r$, $r = 1, \dots 8$ in terms of RMSE.*

Lateral		RMSE		
		$\mathcal{D}_a$	$\mathcal{D}_b$	$\mathcal{D}_c$
$\mathcal{A}_1$	$\mathcal{B}_1$	0.11	1.00	0.84
	$\mathcal{L}_1$	0.12	0.83	0.70
$\mathcal{A}_2$	$\mathcal{B}_2$	1.16	0.72	0.79
	$\mathcal{L}_2$	0.22	0.70	0.80
$\mathcal{A}_4$	$\mathcal{B}_4$	0.17	0.17	0.15
	$\mathcal{L}_4$	0.11	0.19	0.16
$\mathcal{A}_5$	$\mathcal{B}_5$	0.21	0.34	0.27
	$\mathcal{L}_5$	0.14	0.14	0.17
$\mathcal{A}_6$	$\mathcal{B}_6$	0.42	0.42	0.47
	$\mathcal{L}_6$	0.35	0.40	0.48
$\mathcal{A}_7$	$\mathcal{B}_7$	0.50	0.91	0.75
	$\mathcal{L}_7$	0.47	0.68	0.55
$\mathcal{A}_8$	$\mathcal{B}_8$	0.58	0.53	0.26
	$\mathcal{L}_8$	0.59	0.46	0.25

the fact that, in these subsets, the scans start from half the head and, hence, the low amount of landmarks make the estimation unstable. Moreover, in some exams, the hands of the patients are placed on the neck and interfere then with the structures of interest. On the contrary, the images in the subset $\mathcal{D}_a$ are acquired from over the top of the head and there is no occlusion with the hands, which are the conditions that allow for a robust estimation of the EI values. Therefore, the ROI $\mathcal{A}_1$ should be taken into account only if it is important for the exam and, hence, acquired with the patient positioned in a suitable way.

5.7.2.2 Thoracic spine

The evaluation on the whole database indicates that the EI value in the ROI $\mathcal{A}_2$ is not easy to estimate automatically. Indeed, the landmarks $b_j \in \mathcal{B}_2$ produce a RMSE equal to 0.89, and 23.2% and 41.5% of samples whose estimation errors are negligible and acceptable, respectively. The landmarks $l_j \in \mathcal{L}_2$ improve the performances given the RMSE equal to 0.64, and 41.5% and 58.6% of samples whose estimation errors are negligible and acceptable, respectively, but is not satisfying yet. Nevertheless, the results substantially change according to the subset here too. In details, the landmarks $b_j \in \mathcal{B}_2$ give a low quality estimation regardless the morphotype and the amount of dose, whereas the RMSE associated with the landmarks $l_j \in \mathcal{L}_2$ for the exams in $\mathcal{D}_a$ is very low (0.22). Figures 5.14b and 5.14c show that the difference between the results associated with $\mathcal{D}_a$ and $\mathcal{D}_b$ ($\mathcal{D}_c$) is not due to the morphotype but it rather depends on the position of the arms: while for the patient in Figure 5.14b the arms cover the heart but do not occlude the vertebrae, in Figure 5.14c the superior vertebrae (T1-T4) are partially occluded by the shoulders. The thoracic spine is always a region of interest and, thus, the estimation of the associated EI

value should not depend on the position of the arms. A possible solution may consist in attributing higher importance to the local estimates $e(l_j)$ associated with landmarks l_j placed in the inferior part of the thoracic spine.

5.7.2.3 Lumbar spine

The EI value corresponding to the lumbar spine is the easiest one to estimate and does not depend on the morphotype or the amount of dose. Furthermore, the results indicate that there is practically not difference from computing the EI values at the landmarks $b_j \in \mathcal{B}_4$ and $l_j \in \mathcal{L}_4$. The landmarks $b_j \in \mathcal{B}_4$ do not capture the curvature of the lumbar spine, whereas the landmarks $l_j \in \mathcal{L}_4$ do, but this is not relevant because the signal in the abdomen is almost homogeneous.

5.7.2.4 Pelvis

The landmarks $b_j \in \mathcal{B}_5$ allow for good accuracy as the RMSE is equal to 0.28, and 72.0% and 94.0% of samples present negligible and acceptable errors, respectively. However, the landmarks $l_j \in \mathcal{L}_5$ improve the performances given the RMSE equal to 0.15, and 95.1% and 98.8% of samples whose estimation errors are negligible and acceptable, respectively. Similar observations can be made by analyzing the results obtained on the subsets $\mathcal{D}_a$, $\mathcal{D}_b$ and $\mathcal{D}_c$.

5.7.2.5 Femurs, knees and tibiae

The two methods give similar performances in the ROI $\mathcal{A}_6$. In details, the landmarks $b_j \in \mathcal{B}_6$ produce a RMSE equal to 0.44, and 50.0% and 78.0% of samples whose estimation errors are negligible and acceptable, respectively. The landmarks $l_j \in \mathcal{L}_6$ return almost the same RMSE (0.42), but increases to 62.2% and 84.2% the amount of samples whose estimation errors are negligible and acceptable, respectively. However, some high errors may also occur depending on the position of the legs. In particular, if one leg is superposed to the other one in the image, the estimation of a consistent EI value may be quite complicated. Similar considerations on the estimation accuracy are valid for the ROI $\mathcal{A}_8$.

In the knees, the measure computed at the landmarks $l_j \in \mathcal{L}_7$ is more precise than that computed at the landmarks $b_j \in \mathcal{B}_7$ because two peaks of absorption corresponding to the non superposed knees can be detected with the proposed approach. On the other hand, the landmarks $b_j \in \mathcal{B}_7$ necessarily fall at the center by potentially overestimating the EI value. This case mainly concerns the exams in the subsets $\mathcal{D}_b$ and $\mathcal{D}_c$ whose patients have a leg slightly in front of the other. However, even with the proposed method the performances are not very high as the RMSE is equal to 0.58 and, only 36.6% and 58.6% whose estimation errors are negligible and acceptable, respectively. These results confirm the aforementioned problem in estimating an EI value associated with an anatomical structure that is partially occluded by another one.

5.7.2.6 Global evaluation

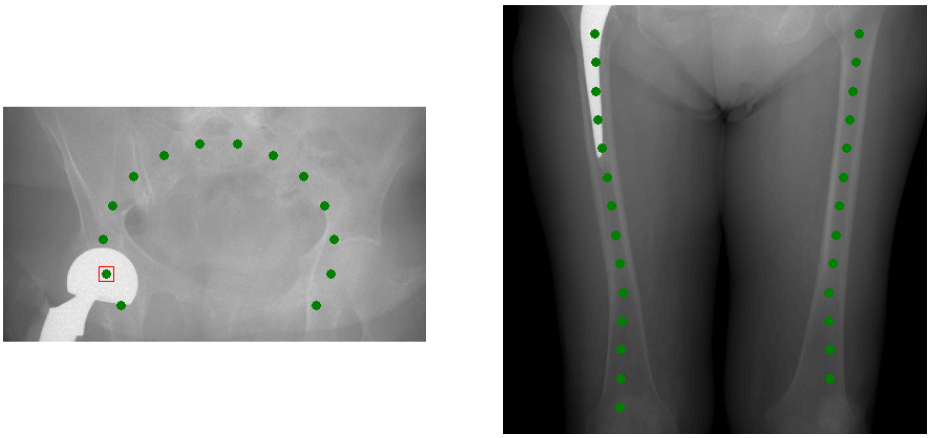
The results on lateral view acquisitions indicate that a level of accuracy similar to that registered on frontal view acquisitions can only be obtained for $\mathcal{A}_4$ and

$\mathcal{A}_5$. However, as for the anatomical structures in the leg, it is not sure that an EI value estimation in this region from lateral view is actually meaningful given the superposition of the legs in the image. As for $\mathcal{A}_1$ the results are mainly influenced by the fact that the full region is not necessarily contained in the images of $\mathcal{D}_b$ and $\mathcal{D}_c$, and the estimation is indeed good for the database $\mathcal{D}_a$. The most significant problem is then represented by the estimation of the EI value in the anatomical ROI $\mathcal{A}_2$ where the quality of the estimation depends on the position of the arms and, in particular, if they are superposed or not with part of the thoracic vertebrae in the image. This problem seems however easy to address by weighting the local estimates also according to prior assumptions on the positions of the landmarks in the cluster.

5.7.3 Metallic objects

The metallic objects should be neglected in the estimation of the EI values because they are not associated with anatomical information and significantly bias the measure given the strong X-ray absorption. Therefore, they should be removed from the whole pixel space Ω, just like the background where there is not attenuation. However, considering that the EI value $e(l_j)$ is computed at a landmark l_j, the non-anatomical objects do not necessarily need to be segmented, but the weight ω_j should be set to 0 if the local estimate $e(l_j)$ corresponds to a metallic object. A local estimate $e(l_j)$ is an outlier only if more than 50% of the circular patch P_j centered at l_j is covered by a metallic object. In other words, the presence of metallic objects in the field of view does not necessarily imply that some of the local estimates $e(l_j)$ need to be discarded and, even if it is the case, the number of outliers is most likely low. According to these observations, a simple threshold-based method is used to reject the contributions from potential metallic objects. Formally, the minima of the local estimates $\min(e(l_j))$, for the landmarks $l_j \in \mathcal{L}_r$, $r \in \{2, 4, 5, 6\}$, are computed. The median of these values is retained as the reference value that corresponds to anatomical exposure in strong absorption regions, which is why only the anatomical ROIs $\mathcal{A}_r$, $r \in \{2, 4, 5, 6\}$ are taken into account. The median value is computed because we assume that the outliers are at most present in two of the anatomical ROIs $\mathcal{A}_r$, $r \in \{2, 4, 5, 6\}$ as biased local estimates due to metallic objects are rare. The threshold τ_m is fixed at 80% of the reference value, i.e. a DI with respect to the reference superior to -1 is considered acceptable. The landmarks $l_j \in \mathcal{L}_r$, $r = 1, \ldots, 8$, such as $e(l_j) < \tau_m$ are considered as outliers and, hence, their weights $\omega(l_j)$ are fixed to zero.

Figure 5.17 shows two sub-images extracted from the same image with a femoral prosthesis. In details, Figure 5.17a displays the landmarks $l_j \in \mathcal{L}_5$ on the pelvic region and the one placed in the middle of the prosthetic femoral head (red square) is rejected. The DI value with respect to the ground truth is then equal to -0.07. Nevertheless, even if the outliers are taken into account, the measure is not significantly affected as the DI value becomes -0.12, which shows how the landmark-based method (see Section 4.4.2) is practically not biased by the presence of outliers as long as their number is limited. Figure 5.17b shows that the presence of metallic objects does not necessarily imply presence of outliers. Indeed, the EI values associated with the left and right legs are practically identical, i.e. 41.24 and 40.27, even if all the landmarks are taken into account. It is worth noting, however, that in



(a) (b)

Figure 5.17: *Robustness to the presence of the metallic objects. In Figure (a), among the landmarks $l_j \in \mathcal{L}_5$ (green points), the point that corresponds to the femoral prosthetic head is labeled as outlier (red square) and then rejected. In Figure (b), none of the landmarks $l_j \in \mathcal{L}_5$ is detected as outlier, but this does not affect the EI measure (see text for details).*

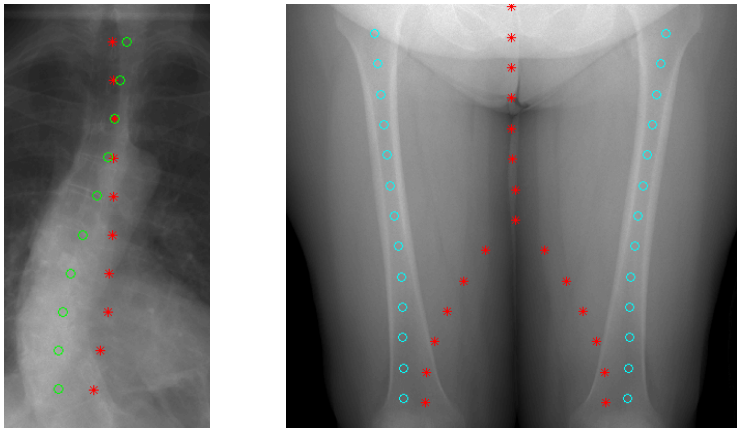
this example only 2 of the 5 landmarks placed in the prosthesis give wrong EI estimates because, as for the others, the metallic object is not represented by the central tendencies of the local gray level distributions. As a consequence, the number of outliers is too low to influence the EI estimated at the femoral region. On the other hand, the proposed technique to reject outliers could be useful when their number in a cluster $\mathcal{L}_r$ is relatively high with respect to the number of landmarks $|\mathcal{L}_r|$. This would be the case, for example, of the estimation of the EI value associated with the anatomical region $\mathcal{A}_7$ in presence of a knee prosthesis.

Finally, these results well summarize that our method is able to overcome the issues due to the presence of metallic objects, which is one of the advantages of the landmark-based formulation (Section 4.4.2).

5.7.4 Robustness to particular conditions of the patients

We address in this section the challenges brought by the presence of deformed anatomical structures (e.g. the scoliosis) and by changes in size of the patient. The measures computed from the landmarks $l_j \in \mathcal{L}_r$ and $b_j \in \mathcal{B}_r$ are compared.

Figure 5.18a displays the cluster $\mathcal{L}_2$ (green points) and $\mathcal{B}_2$ (red points) on the anatomical region $\mathcal{A}_2$ of a patient affected by scoliosis. The landmarks $b_j \in \mathcal{B}_2$ generate a DI index value equal to 1.36 because the central axis does not correspond to the thoracic spine but rather to the surrounding structures. On the contrary, the landmarks $l_j \in \mathcal{L}_2$ well respect the curvature of the spine and, hence, the EI value is robustly estimated as shown by the DI value that is equal to -0.16. In general, when a patient is affected by a scoliosis in the thoracic region, the measure computed at $b_j \in \mathcal{B}_2$ generates severe errors, whereas the proposed method overcomes this issue. This analysis justifies the results presented in Table 5.3 for the anatomical ROI



(a) (b)

Figure 5.18: *Particular cases due to patient conditions. In Figure (a) the landmarks $b_j \in \mathcal{B}_2$ do not follow the curvature of the spine due to scoliosis, whereas the landmarks $l_j \in \mathcal{L}_2$ allow overcoming this drawback. In Figure (b) as the legs are adjacent, the landmarks $b_j \in \mathcal{B}_2$ do not capture the femurs, whereas the landmarks $l_j \in \mathcal{L}_2$ well match the anatomical structures.*

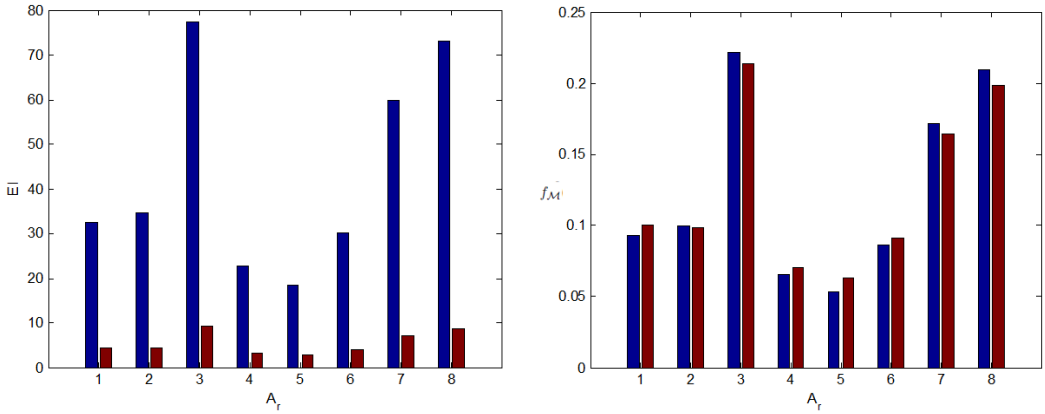
$\mathcal{A}_2$. On the other hand, if the scoliosis affects the lumbar spine the EI value is not considerably biased.

Figure 5.18b shows the femoral region of a patient with high body mass index. Since the legs are joined, the envelop of the patient does not well describe the underlying anatomy and the landmarks $b_j \in \mathcal{B}_6$ start separating into two branches only at half of the femur. As a consequence, the DI value rises up to 1.69. The proposed method does not depend on the morphology of the patient (see Section 5.5.2) and the EI value is perfectly estimated (DI equal to 0.01). These considerations are coherent with the results presented in Table 5.3 corresponding to the anatomical ROI $\mathcal{A}_6$.

5.7.5 Information extracted from preview images

The results presented in Sections 5.7.1 and 5.7.2 indicate that the EI, expressed as SNR, can be also correctly estimated on preview images in spite of the drastic reduction of radiation exposure. This is strictly related to the ability in detecting accurate landmark positions on preview images as shown by the examples in Sections 5.5 and 5.6 and in Appendix B.

EI values computed over preview acquisitions could be exploited to estimate the values of the acquisition parameters (e.g. kV and mA) to be used in diagnostic or follow-up examinations (see Section 4.2.2). In order to achieve this goal, the estimated EI values could be associated with the representative patient's thickness in an anatomical ROI according to a database of SNR measures computed over flat fields behind PMMA blocks of varying thickness. In this section, a theoretical study is provided in order to establish whether the information present in preview images is sufficient to achieve this goal or not. In other words, by assuming that the EI measures computed over images acquired at a normal amount of dose allow



(a) (b)

Figure 5.19: *EI values on a diagnostic image (blue bars) and on the corresponding preview acquisition (red bars): (a) measures; (b) normalized measures, i.e. the EI value distributions.*

estimating the patient's thickness, we evaluate how much the information extracted from a preview diverges from the one at higher radiation exposure.

Figure 5.19a compares the EI values computed at the landmarks manually annotated on a diagnostic image (blue bars) and on the corresponding preview image (red bars), respectively. It is not surprising that the two sets of EI values are very different as the X-ray doses at the detector are not comparable. Nevertheless, the two sets are very much similar if they are normalized, i.e. if the distributions of the EI values at the anatomical ROIs are observed. Figure 5.19b shows two EI distributions $f_{\mathcal{M}}(\mathcal{D}_b^i)$ and $f_{\mathcal{M}}(\mathcal{D}_c^i)$. The notation $f_{\mathcal{M}}(\mathcal{D}_b^i)$ indicates the distribution of EI values computed at the manually annotated landmarks ($\mathcal{M}$) on the frontal or lateral view image of a patient i that belongs to the database $\mathcal{D}_b$, i.e. the diagnostic examinations on young patients of our full dataset. By considering that the acquisition parameters do not change during the c-arm scan, if the EI values are distributed in the identical way on the preview and diagnostic images of the same patient, it then means that the identical information about the patient's thickness can be extracted from them. Figure 5.19b suggests that this is the case, but, in order to quantitatively validate this assessment, we compute the Kullback-Leibler divergence (Kullback and Leibler, 1951) of the two distributions. This measure is used in information theory to compare two probability distributions X and Y. The discrete case is considered here and the Kullback-Leibler divergence is then defined as follows:

$$\Delta(X \mid Y) = \sum_r X(r) \log \left(\frac{X(r)}{Y(r)} \right). \tag{5.22}$$

The Kullback-Leibler divergence is a non-symmetric measure, i.e. $\Delta(X \mid Y) \neq \Delta(Y \mid X)$, that allows quantifying the amount of missing information when Y is used instead of X, where Y is an approximation of the reference X. This measure is then exploited for our purpose by considering $f_{\mathcal{M}}(\mathcal{D}_b)$ as the reference and the following distributions as its approximations: $f_{\mathcal{M}}(\mathcal{D}_c)$, $f_{\mathcal{B}}(\mathcal{D}_c)$ and $f_{\mathcal{L}}(\mathcal{D}_c)$. The first tested EI value distribution allows establishing if preview acquisitions contain a sufficient

amount of information to estimate the patient's thickness, whereas the two others are evaluated to assess the ability in automatically extracting this information.

Table 5.7: *Average and maximum of the Kullback–Leibler divergence measures (Δ) computed between EI value distributions associated with diagnostic ($\mathcal{D}_b$) and corresponding preview ($\mathcal{D}_c$) frontal view acquisitions. The EI values are computed at landmarks manually annotated ($\mathcal{M}$) and automatically detected according to the patient's envelop ($\mathcal{B}$) or with the proposed approach ($\mathcal{L}$).*

	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{M}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{B}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{L}(\mathcal{D}_c)$
average$(\Delta \cdot 10^3)$	2.11	8.39	4.09
max$(\Delta \cdot 10^3)$	4.78	58.78	27.28

Table 5.7 shows the average and maximum Kullback–Leibler divergence values obtained by comparing the EI measure distributions for frontal view acquisitions of the patients in the subset $\mathcal{D}_b$ with the corresponding previews $\mathcal{D}_c$. The measures associated with manually annotated landmarks on previews well approximate the information on images acquired at normal dose as indicated by the average and maximum Δ values that are equal to $2.11 \cdot 10^{-3}$ and $4.78 \cdot 10^{-3}$, respectively. The landmarks $\mathcal{L}$ that are detected according to the automatic proposed approach give an average and a maximum Δ values that are equal to $4.09 \cdot 10^{-3}$ and $27.28 \cdot 10^{-3}$, respectively. Therefore, the divergence remains very low when the measures are automatically estimated. Acceptable results are also obtained when the landmarks $\mathcal{B}$ are used, but the Kullback–Leibler distance is twice higher than with the proposed technique that, thus, is preferred.

Table 5.8: *Average and maximum of the Kullback–Leibler divergence measures (Δ) computed between EI value distributions associated with diagnostic ($\mathcal{D}_b$) and corresponding preview ($\mathcal{D}_c$) lateral view acquisitions. The EI values are computed at landmarks manually annotated ($\mathcal{M}$) and automatically detected according to the patient's envelop ($\mathcal{B}$) or with the proposed approach ($\mathcal{L}$).*

	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{M}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{B}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{L}(\mathcal{D}_c)$
average$(\Delta \cdot 10^3)$	9.84	13.68	11.02
max$(\Delta \cdot 10^3)$	44.10	52.65	32.06

Table 5.9: *Average and maximum of the Kullback–Leibler divergence measures (Δ) computed between EI value distributions associated with diagnostic ($\mathcal{D}_b$) and corresponding preview ($\mathcal{D}_c$) lateral view acquisitions. The EI measures associated with anatomical structures in the legs are rejected. The EI values are computed at landmarks manually annotated ($\mathcal{M}$) and automatically detected according to the patient's envelop ($\mathcal{B}$) or with the proposed approach ($\mathcal{L}$).*

	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{M}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{B}(\mathcal{D}_c)$	$\mathcal{M}(\mathcal{D}_b) \mid \mathcal{L}(\mathcal{D}_c)$
average$(\Delta \cdot 10^3)$	2.95	8.26	7.47
max$(\Delta \cdot 10^3)$	7.83	39.69	43.60

Table 5.8 shows the average and maximum Kullback–Leibler divergence values obtained by comparing the EI measure distributions for lateral view acquisitions of

the patients in the subset $\mathcal{D}_b$ with the corresponding previews $\mathcal{D}_c$. We observe that the distances between the distributions $f_{\mathcal{M}}(\mathcal{D}_b)$ and $f_{\mathcal{M}}(\mathcal{D}_c)$ are higher than the ones obtained on frontal views. This is due to the issues caused by the superposition of the legs in the image (see Section 5.7.2), as shown in Table 5.9 that reports the same measures than the ones in Table 5.8, but excluding the EI values associated with the femur, knee and tibia from the distributions. The average and maximum Δ values are then equal to 2.95 and 7.83, i.e. comparable to the ones obtained on the frontal views, which suggests that it is preferable to neglect the anatomical structures in the legs for the analysis of lateral view acquisitions. As for the automatic estimation, the distributions of the EI values computed at the landmarks $\mathcal{B}$ and $\mathcal{L}$ diverge approximately to the same degree with respect to the ground truth on diagnostic images. The average distance between $f_{\mathcal{M}}(\mathcal{D}_b)$ and $f_{\mathcal{L}}(\mathcal{D}_c)$ is equal to 7.47, which is higher than the corresponding divergence measured on frontal view acquisitions (4.09). This is certainly because of the problems in estimating the EI values in the thoracic spine region from the lateral view (see Section 5.7.2).

These results prove the coherence between the EI values extracted from a diagnostic image and the corresponding preview acquisition. Furthermore, the measures can be computed in a completely automatic way while not significantly damaging the amount of information in the preview acquisition. Although, frontal view acquisitions are easier to process than lateral ones, which follows the logic of the results presented in Sections 5.7.1 and 5.7.2. Finally, we estimate that the type of information extracted after a preview acquisition is sufficient to robustly drive an automatic exposure control system. Nevertheless, this assessment remains to be empirically validated by linking the EI measure to the thickness of the patient.

5.8 Conclusion

In this chapter we have proposed an approach for the detection and recognition of anatomical landmarks that, combined with the technique introduced in Section 4.4.2, allows estimating the EI value associated with an anatomical ROI on frontal and lateral view clinical acquisitions. Our method is original, but it shares some principles with the approaches from the literature such as the combination of global and local analysis, and the sparse search for the solution. In particular, we exploit a global model that encodes the proportions between the structures and the relative positions of the clusters of landmarks to define. We then combine it with a local analysis that is driven by some salient points that are associated with peaks of absorption or signal at the detector.

The evaluation indicates a strong correlation between the estimated EI values and those computed at manually annotated landmarks. In details, by considering a heterogeneous set composed by 82 patients, the highest RMSE values on the frontal and lateral view acquisitions both correspond to the ROI $\mathcal{A}_2$ and are equal to 0.30 and 0.64, respectively. Therefore, on frontal view acquisitions the results are very satisfying, and on lateral view acquisitions the performances are good but the estimation needs to be improved in some anatomical ROIs. In particular, the EI estimate for $\mathcal{A}_2$ (thoracic spine) should not be affected by the position of the patient arms and the actual need for an estimation in the legs should be certified by accounting for the opinion of the users. We have then shown that the proposed approach

overcomes some drawbacks of the state of the art methods in certain challenging situations. Finally, the EI values computed over the preview images offer a sufficient amount of information for forecasting the optimal parameter setting for the diagnostic examination. Moreover, the information is practically not deteriorated if the automatically estimated EI values are used rather than the ground truth ones. The latter point further increases the interest of our approach in the optic of designing an automatic exposure control system coherent with ALARA guidelines.

The proposed detection approach can, however, be improved in certain aspects. For example, the order of the search and some parameters of the algorithm, e.g. the factors of scale, are fixed. It would be more efficient to dynamically adapt them according to features extracted from the image. Moreover, the detection of some landmarks progressively narrows the search space for the other ones. Nevertheless, the algorithm does not provide yet a backward check on the already located landmarks, which would help to avoid the propagation of errors due, for example, to the misplacement of control points. Please refer to Section 6.1.2 for a detailed discussion on the perspectives of improvement of our landmark detection method.

Chapter 6

Conclusion and perspectives

6.1 Contributions and open issues

In this thesis, we have addressed the reduction of radiation exposure for images acquired with the Stereo-Radiographic system EOS by means of two different approaches. The first one aims at optimizing the compromise between image quality and amount of dose by exploiting a joint denoising and contrast enhancement method. The goal of the second approach is to help users in daily clinical practice by providing them an immediate feedback on the image quality in order to be able to certify whether they are operating according to ALARA guidelines or not. Our contributions are detailed in the following discussion, which also presents some related open issues.

6.1.1 Optimization of the image quality

In the first part of the thesis we have studied how to optimize the quality of X-ray images by means of post-processing algorithms. The presence of noise makes it particularly challenging to optimize the balance between the increase of the visibility of anatomical features and the reduction of noise. We have then studied different solutions to understand which one better adapts for diagnosis. The proposed approaches and their comparison are, to the best of our knowledge, original contributions of our work.

The non-homoscedastic noise that affects EOS and, in general, X-ray images, is the first element to take into account for restoring anatomical information. Moreover, the noise level of the image to be processed is unknown and needs then to be estimated. The block-based formulation of the percentile method allows estimating the strength of signal-dependent noise. We have then extended the Non Local-means (NL-means) filter to denoise X-ray images by relying on the estimation provided by the percentile method. The proposed filter, called X-ray Non Local-means (XNL-means) filter, allows adapting to the noise level in a given region while avoiding manual tuning of parameters. The compromise between noise reduction and preservation of anatomical structures is then automatically optimized. This property is particularly relevant for processing full-body X-ray images given the high heterogeneity of tissue densities present in the human body. Since the degree of X-ray absorption changes depending on the density, the signal that reaches the detector, and hence the noise level, significantly varies. Furthermore, we have shown that for

a given amount of signal, the noise can be locally approximated by a Gaussian model. By making the assumption that the signal changes smoothly, the XNL-means filter can then be efficiently implemented by using differences of integral images.

We have then studied how to exploit the filtered image in combination with the contrast enhancement. The easiest option is to fully enhance the contrast of the filtered image. This solution assumes that the noise has been completely removed while preserving the anatomical structures. According to the evaluation, this approach entails a slight loss of resolution that does not make it suitable for diagnosis. We have hence proposed to use the filtered image to define non-parametric noise containment maps to limit the increase of noise in a multiscale framework. The general advantage of noise containment-based methods over denoising is that no information is lost, and, thus, they are more adapted for X-ray image processing. Generally, the main drawback of this kind of solutions lies in their dependency on user-defined parameters. Our method offers a solution to this issue by comparing at each scale the observed and filtered images to determine the probability for a coefficient to be associated with noise. This measure is computed according to the correlation between the two images and, thus, it depends on the amount of local redundant information that the XNL-means filter tends to preserve. We have also proposed a method for the evaluation on clinical images that consists in associating average local variance (ALV) measures with anatomical structures of interest. The measures so computed are coherent with visual comparison and readings from expert users (radiologists).

Despite the assessed good performances, some aspects could still be improved. First, the patch-based filters only remove the high frequency component of the noise, but the noise is uniformly distributed at all the frequencies. In other words, the noise is also scale-dependent and the XNL-means filter does not remove the noise at medium/low frequencies. In the majority of the cases, the contribution of the noise at these levels is negligible, but this is not true when the radiation exposure is very low. According to recent advances in the literature (Delbracio et al., 2014), patch-based filters that denoise at different scales may be interesting to address the mentioned issue. In our context, this would probably allow us to define more robust noise containment maps at medium/low levels for images acquired at very low dose. Secondly, other types of multiscale decomposition than the isotropic undecimated wavelets should be tested. An anisotropic decomposition would indeed give a better characterization of the noise given the oriented representation of image features. However, it has to be taken into account that the inverse transform after the enhancement of the band-pass coefficients may cause an excessive boosting (Dippel et al., 2002). Finally, the execution time is still too high to use the proposed method in clinical routine. Nevertheless, the code can be parallelized, leading to an efficient implementation.

6.1.2 Estimation of the image quality

In the second part of our work, we have studied how to automatically estimate the quality of non-processed EOS images to help users to control whether they are using the correct radiation exposure level or not. After having analyzed the exposure index (EI) algorithm, we have proposed an alternative approach that offers multiple advantages such as consistency between inter-examinations measures, possibility of

managing the presence of multiple anatomical structures in the field of view, and robustness to metallic objects and pathological cases. The problem formulation and the method introduced to solve it are original and represent, to the best of our knowledge, the first proposition that addresses the definition of relevant regions of interest for the EI application since the publication of the standard IEC 62494-1 (2008).

The method for the estimation of the noise level used in the first part of the thesis is also convenient to quantify the image quality in terms of signal to noise ratio (SNR). This measure is correlated to the EI as the latter is proportional to the squared SNR. Therefore, it can be used as alternative definition of EI as it also provides an image quality measure linked to the radiation exposure. The main issue is how to define the regions of the image that are meaningful to quantify the image quality depending on the clinical purpose of the undergoing examination. The structures of interest need to be detected on the field of view in order to provide robust measures. Since the EI has to represent the central tendency of a given region, the median value of the gray level distribution at the pixels belonging to the region is often retained as value of interest from which the EI is computed. We have shown that a more robust measure can, however, be obtained from multiple local estimates computed at some landmarks associated with the anatomical structures of interest. The impact of detection errors or the bias due to outliers, e.g. metallic objects, are significantly reduced with the landmark-based approach.

According to this formulation of the problem, the image quality estimation becomes a question of detection and recognition of anatomical landmarks. We have proposed an unsupervised method that addresses this task by combining the global information on the size and the positions of the anatomical structures to local analysis. The search of the position of the landmarks is sparsely conducted by visiting salient points that correspond to peaks of absorption or signal at the detector. The results indicate that the EI values computed at the automatically detected landmarks correspond very well to those associated with manual annotations on frontal view acquisitions. On lateral view acquisitions, the estimates in the thoracic spine may be biased by the superposition with the shoulder in the superior part of the region. Therefore, in a next version of the method, the contributions from the lower thoracic vertebrae should be considered as more important. An unstable estimation may also occur in the legs due to their superposition in a lateral image. We should then check with the users whether the estimation of the image quality in these regions is meaningful despite the superposition or not. Note that the validation has been conducted on a heterogeneous database composed by full-body images of patients of different morphotypes, acquired at different radiation exposure amounts, and including post-surgery and pathological cases. The proposed method has shown to be very robust to all these different cases, except for the aforementioned structures acquired from the lateral view for which the estimation needs to be improved. Even preview images have been included in the tests and we have proved that it is possible to extract from them consistent information that could be used to drive the parameters setting for the following diagnostic examinations.

The proposed detection method relies on the hypothesis that the anatomical structures present in the image are known. For example, all the tested images cover the full body. This allows initializing the method by detecting a set of control points from which the search for the solution can be sequentially narrowed and then locally

refined. Nevertheless, if this assumption is dropped, the proposed method would not work. A possible solution to overcome this drawback may consist in detecting the control points with a classification- or regression-based supervised approach. The control points could be then located despite the missing prior knowledge and the landmarks detected as proposed in our method. In our context, the main challenge is to define some descriptors to learn a robust model in spite of the changes in patient morphotype and radiation exposure. The spatial relations between the salient points may offer a solution given their consistency among heterogeneous cases. However, it remains to be established how to formally achieve this task.

The detection is sequentially conducted in a predefined order. However, as proposed by Fouquier et al. (2012), it may be preferable to conduct the search according to information extracted from the image to process. Moreover, while an already detected cluster of landmarks can constraint the sets corresponding to other anatomical structures, our method does not offer a backward check on the already detected landmarks. This option could however prevent eventual errors to propagate. The algorithm could also be improved by adaptive setting of the parameters, such as the scale at which the salient points are computed. Moreover, the statistical inferences that are used for the 3D skeleton reconstruction from the bi-planar EOS images may also be exploited in the detection method to remove some outliers from the sets of salient points.

A further perspective consists in adapting the shape of the patches to the associated anatomical regions. For example, we could register bone structure models to the automatically detected landmarks. This may allow estimating EI values that are even closer to those obtained from manually segmented ROIs. Furthermore, this would be interesting for other applications such as the initialization of automatic anatomical structure segmentation.

Finally, in the future we aim at working in collaboration with radiologists and physicists in order to extend the results and evaluate the application of the proposed method to clinical practice.

6.2 Perspectives

Future works will address the open issues that we have pointed out in Section 6.1. However, during this thesis we have also identified other research directions that are presented in this section.

6.2.1 More on image processing

6.2.1.1 Application to segmentation and registration

In Chapter 3 we have shown that by fully enhancing the filtered input image, the spatial resolution is slightly reduced. Therefore, this solution does not suit well the visualization of clinical images. On the other hand, this type of approach could be adapted to applications such as automatic segmentation and registration, for which it is not necessary to preserve fine bone texture. Besides, the denoised image enhanced in contrast is the best one in terms of contrast to noise ratio and the edges of the structures are preserved with the XNL-means filter. As a consequence, it would be interesting to quantify how much the precision of image segmentation and

registration algorithms improves when applied on this type of images rather than on typical display images that are richer in details but have a higher amount of noise as well.

The method proposed in Chapter 5 has been validated according to the EI values computed at the detected landmarks. However, the positions of the landmarks could also be used to reduce the amount of user interaction required by the software for 3D reconstruction of the patient's skeleton.

6.2.1.2 Halo artifacts

The halo artifacts are one of the problems in digital radiography image processing (see Chapter 2). This effect typically appears around metallic objects under the form of shadows that do not allow to properly evaluate whether the tissues that surround the prosthesis are in good health or some complications are manifesting after surgery. A similar effect may be observed at the borders of the lungs, i.e. not only in presence of metallic objects. This problem is due to the smoothing of strong edges in the linear multiscale decomposition. In our work, this problem has been limited by using proper activity maps. However, in order to completely avoid this issue, edge-preserving decomposition methods have been proposed in the literature. The use of these approaches for medical applications has not been studied yet and it would be interesting to evaluate them in this context. Nevertheless, for multiple reasons the extension of edge-preserving methods to the decomposition of X-ray images is not a trivial task. First, the degree of smoothing depends on the parameters that quantify the presence of discontinuities. Therefore, the effect of these parameters and the capacity in well representing the different levels of spatial frequency need to be evaluated. Secondly, the time required for the decomposition would certainly increase as well.

6.2.1.3 Anatomical average local variance to drive parameters setting

In the conclusion of Chapter 2 we have proposed an extension of the ALV measure and used it in Chapter 3 to compare different methods. This measure has been proved to be consistent, but it requires to manually segment anatomical structures of interest on the images. This procedure, beyond being time demanding, limits this tool to the evaluation of post-processed images. It would be then interesting to re-adapt the detection and recognition method proposed in Chapter 5 to automatically get ALV measures associated with anatomical structures. This would not only allow avoiding the manual annotations, but it could also be used to dynamically define the optimal gain parameters used for boosting the multiscale band-pass coefficients. In practice, a limited number of gain parameters settings could be tested to find which one maximizes the anatomical contrast measure by assuming, for example, that the main objective is to increase the visibility of anatomical features.

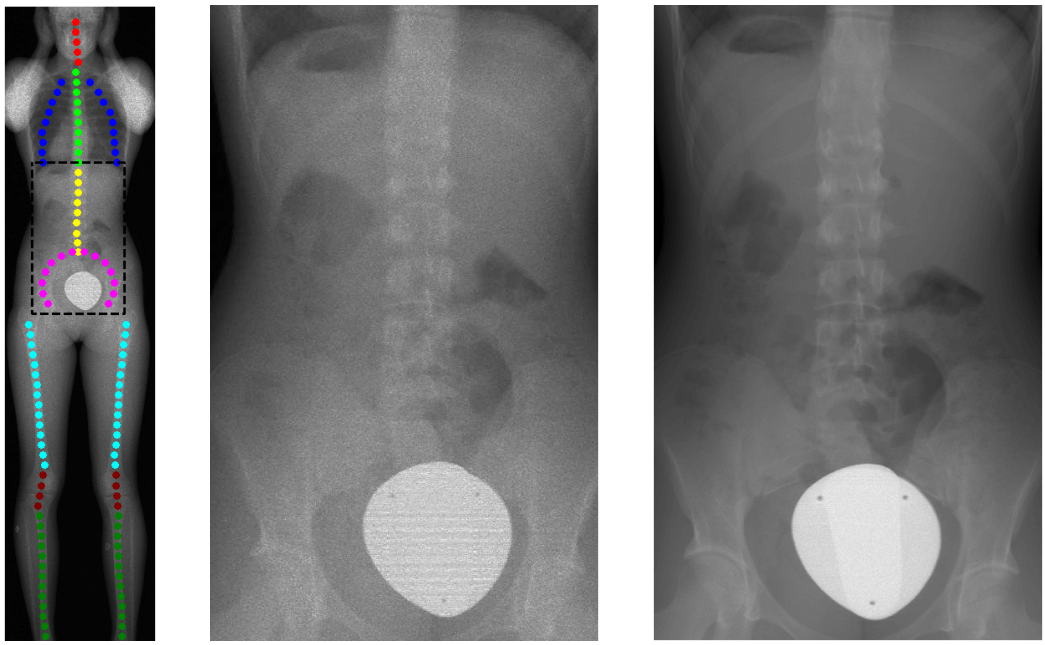
6.2.2 Applications with preview images

6.2.2.1 mA modulation

In Chapter 5 we have shown that a considerable amount of information can be extracted from preview images despite the very low radiation exposure. In particular,

we have suggested the possibility of using image quality measures computed from preview images to define the optimal parameters setting for the following diagnostic acquisition. A similar question that could be considered is whether the mA value needs to be fixed or rather dynamically adjusted during the c-arm scan. The second option seems to be more coherent with ALARA principles as, intuitively, different anatomical structures in the body do not need to be exposed to the same amount of radiations. For example, the mA should be higher in the pelvic region than in the knees where the X-ray absorption is low. The image quality measures computed over preview images allow capturing this information and can be hence used to drive the mA modulation for the corresponding diagnostic examination. In other words, just as the preview acquisitions could be used to balance the inter-patient variability (Chapter 5), they could also be exploited to harmonize the amount of absorbed doses in different anatomical regions, i.e. to account for the intra-patient variability. Nevertheless, the feasibility of such applications remains to be certified.

6.2.2.2 Definition of the exposed region



(a) (b) (c)

Figure 6.1: *Definition of the exposed region from preview images: (a) an example of preview image with the detected anatomical landmarks and the rectangle that defines the region of interest (e.g. the lumbar spine); (b) zoom in the rectangle on the preview image; (c) the diagnostic examination limited to the region of interest.*

According to ALARA guidelines, only the region that is relevant for the undergoing examination should be exposed to X-rays. The landmarks detected on preview images can be used to define the extension of the scanned region on the corresponding diagnostic acquisition. Figure 6.1 shows an example of this procedure. A full-body preview image is acquired and the anatomical landmarks are detected

by using the method presented in Chapter 5 (see Figure 6.1a). Let us consider for example an examination of the lumbar spine. In this case, the clusters corresponding to the lumbar spine and to the pelvis are used to define a rectangle that represents the region to be exposed to X-rays. The pelvis is also considered because the positions of the femoral heads are assumed to be relevant for this examination. Figure 6.1b shows the zoom in the selected region according to the positions of the anatomical landmarks in the clusters of interest. Figure 6.1c shows the cropped region from the diagnostic image that would correspond to the scanned region by using the inputs from the preview acquisition. All the structures of interest are well captured while limiting the exposed tissues. The procedure requires to acquire the full-body preview image, but the price to pay in terms of dose is widely justified by the gain in the same terms that is obtained by reducing the exposed region in the following diagnostic examination. The provided position and extension of the rectangle are although only a suggestion as it should not be excluded that the user is interested in having the visibility of a wider area even if the main objective of the exam is to study the lumbar spine. Therefore, the user should be left free to reject the proposed region if this does not completely satisfies the original clinical purpose. The management of the user interaction and the automatic suggestion of clinically relevant regions of interest remain to be properly established.

Finally it is worth noting that preview full-body images could offer a direct solution to one of the issues presented in Section 6.1.2. Indeed, by detecting the anatomical landmarks on the preview image we would be straight aware of which control points are present in the diagnostic acquisition and of their positions. The control points would just need to be refined to correct potential changes of locations due to slight movements of the patient between the two acquisitions.

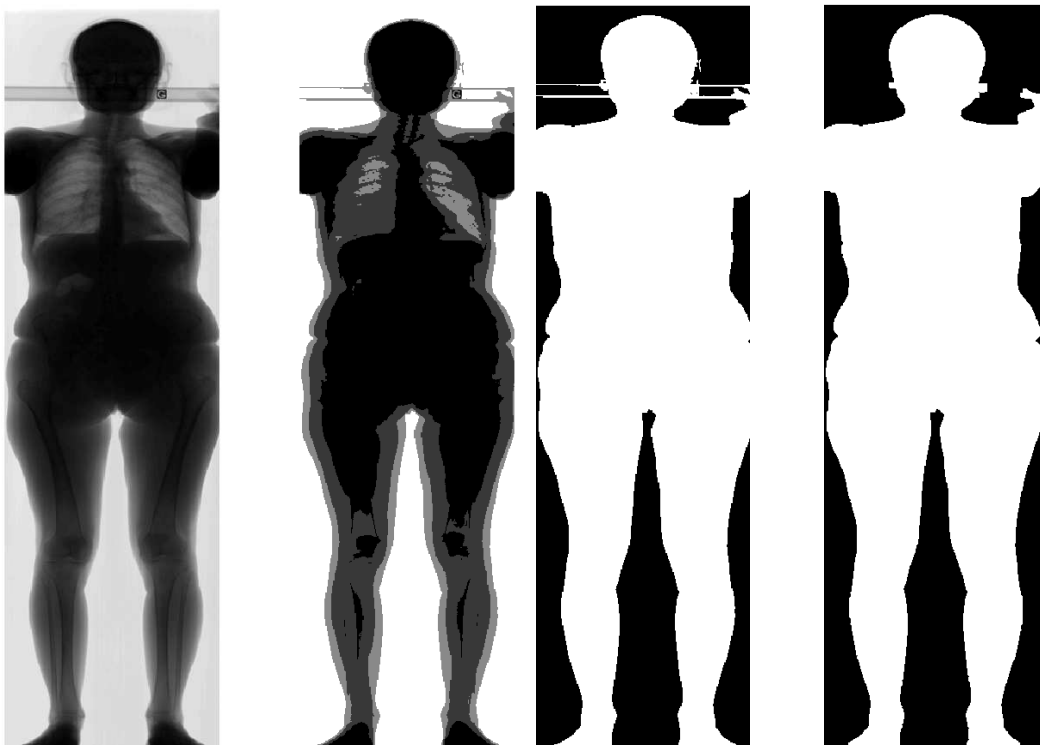
Appendix A

Computation of patient's envelop and landmarks

We present how to extract the envelop of the patient from a non-processed EOS image $\mathbf{u}$, i.e. how to separate the body of the patient from the background that is the set of pixels associated with X-rays that have not been attenuated. We then explain how to define the position of some landmarks depending on the envelop in order to use them as inputs to the algorithm, introduced in Chapter 4, that computes the exposure index from local estimates. These are the landmarks $b_j \in \mathcal{B}_r$, $r = 1, 2, \ldots, 8$ that have been compared in Chapter 5 to those obtained with our approach based on a global to local analysis.

A.1 Method for the extraction of the patient's envelop

Figure A.1 presents the sequence of steps that allows extracting the envelop of the patient. The image $\mathbf{u}$ (Figure A.1a) is considered because the X-ray attenuation follows an exponential law and, thus, there is a relevant difference between absorbed and non-absorbed X-rays in terms of gray level values. We then cluster the pixels according to their intensity levels. We use the k-means clustering method with 4 classes to label the pixels of the image in Figure A.1a. The corresponding result is reported in Figure A.1b where the values at the pixels correspond to the centroid of the cluster to which they are assigned. Since we are interested in separating the body from the background, the cluster with the highest centroid value is identified as the area where the X-rays are not absorbed. Although two classes have to be distinguished (patient's body and background), four clusters are considered in order to avoid including in the background very low density tissues, e.g. air in the lungs. Figure A.1c shows the mask obtained after the clustering. The latter is post-processed with the following morphological operations: an erosion with a circular structuring element of radius equal to 0.2 cm, the selection of the biggest connected component and, finally, a dilatation with the same structuring element as the one used for the erosion. The erosion allows eliminating some components external to the body such as the ones side to the head (Figure A.1c). The biggest connected component is retained as, intuitively, the structures of the human body are all connected together. The dilatation allows restoring the part of the mask connected to the body that has been eroded. Figure A.1d gives an example of the resulting envelop. According to



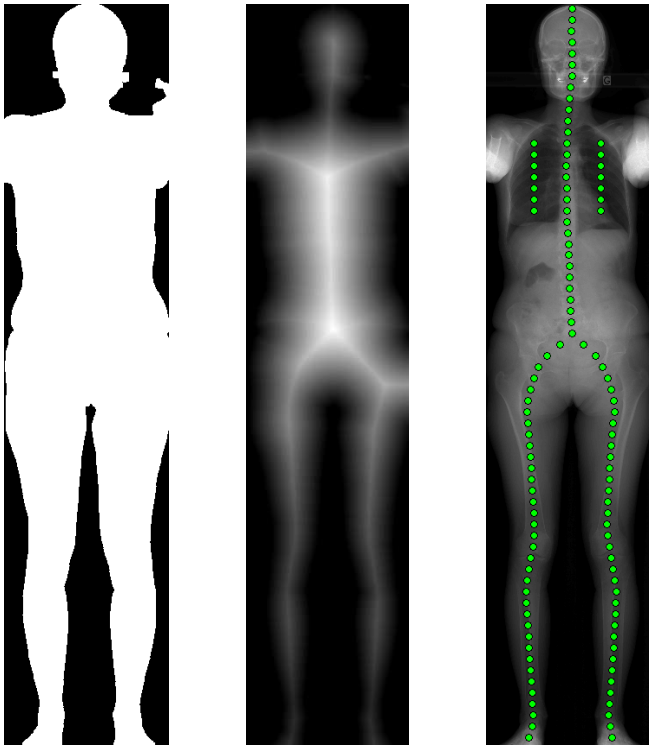
(a) (b) (c) (d)

Figure A.1: *Segmentation of the patient's envelop: (a) input image* $\mathbf{u}$*; (b) result of the clustering; (c) mask defined by excluding the cluster associated with the highest centroid; (d) mask after morphological operations.*

our tests on about 200 images, this procedure allows us to well extract the patient's envelop. Some marginal soft tissues may not be included in the mask, but this is not relevant for many application such as the one presented in the next section.

A.2 Landmarks inferred from the envelop

In the state of the art, the regions of interest for the exposure index are often placed at the center of the image because it is where significant information is assumed to be. A transposition of this idea into the landmark-based formulation leads to place the landmarks $b_j \in \mathcal{B}_r$ at the center of the mask. Formally, given the patient's envelop (Figure A.2a), the Euclidean distance transform (Figure A.2b) is computed. The values of this map increase with the distance from the borders of the mask and are equal to zero outside of it. On each line, the position of the maximum of the Euclidean distance transform profile on a line y_j corresponds to the x-coordinate of a landmark b_j. Two points per line, i.e. two local maxima, may also be considered if symmetrical to the central vertical axis of the mask (e.g. the legs on frontal view acquisitions). Figure A.2c shows an example of the resulting landmarks. Note that the landmarks in the lungs ($b_j \in \mathcal{B}_3$) are located as a function of the positions of those in the thoracic spine ($b_j \in \mathcal{B}_2$) and of the borders of the patient. Formally, if p_f is the nearest point on the border at the left (right) of the



(a) (b) (c)

Figure A.2: *Landmarks defined from the patient's envelop: (a) patient's envelop; (b) Euclidean distance map; (c) landmarks* $b_j \in \mathcal{B}_r$.

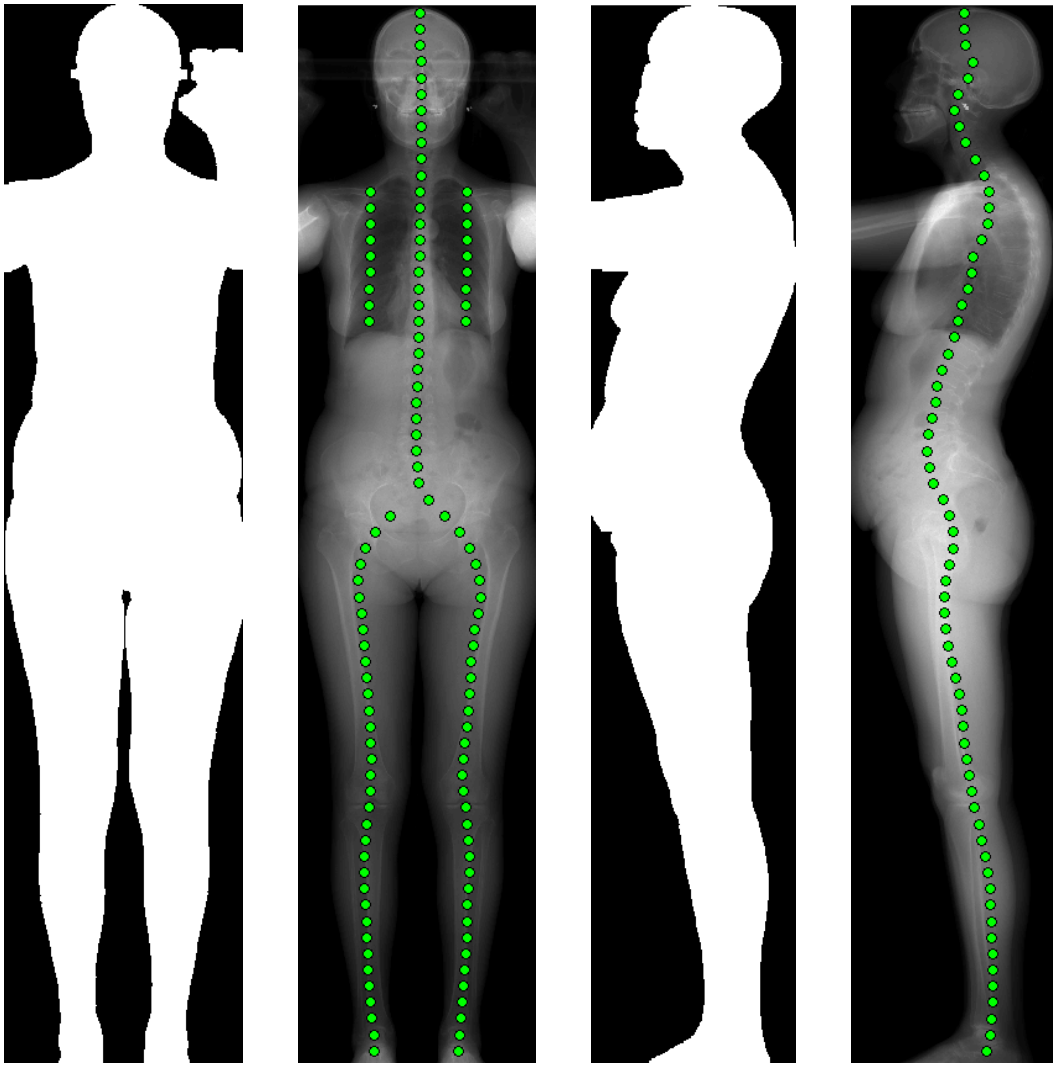
landmarks $b_f \in \mathcal{B}_2$ on a line $y = y_f$, the landmark $b_j \in \mathcal{B}_3$ in the left (right) lung will be placed at the coordinates $0.5(p_f + b_f)$. Finally, the landmarks are assigned to the specific anatomical structures depending on the positions of the control points (see Chapter 5).

A.3 Examples

We present here a sequence of figures that summarize the results obtained for the patient's envelop and the corresponding landmarks $b_j \in \mathcal{B}_r$. The following main conclusions are drawn from the analysis of these images:

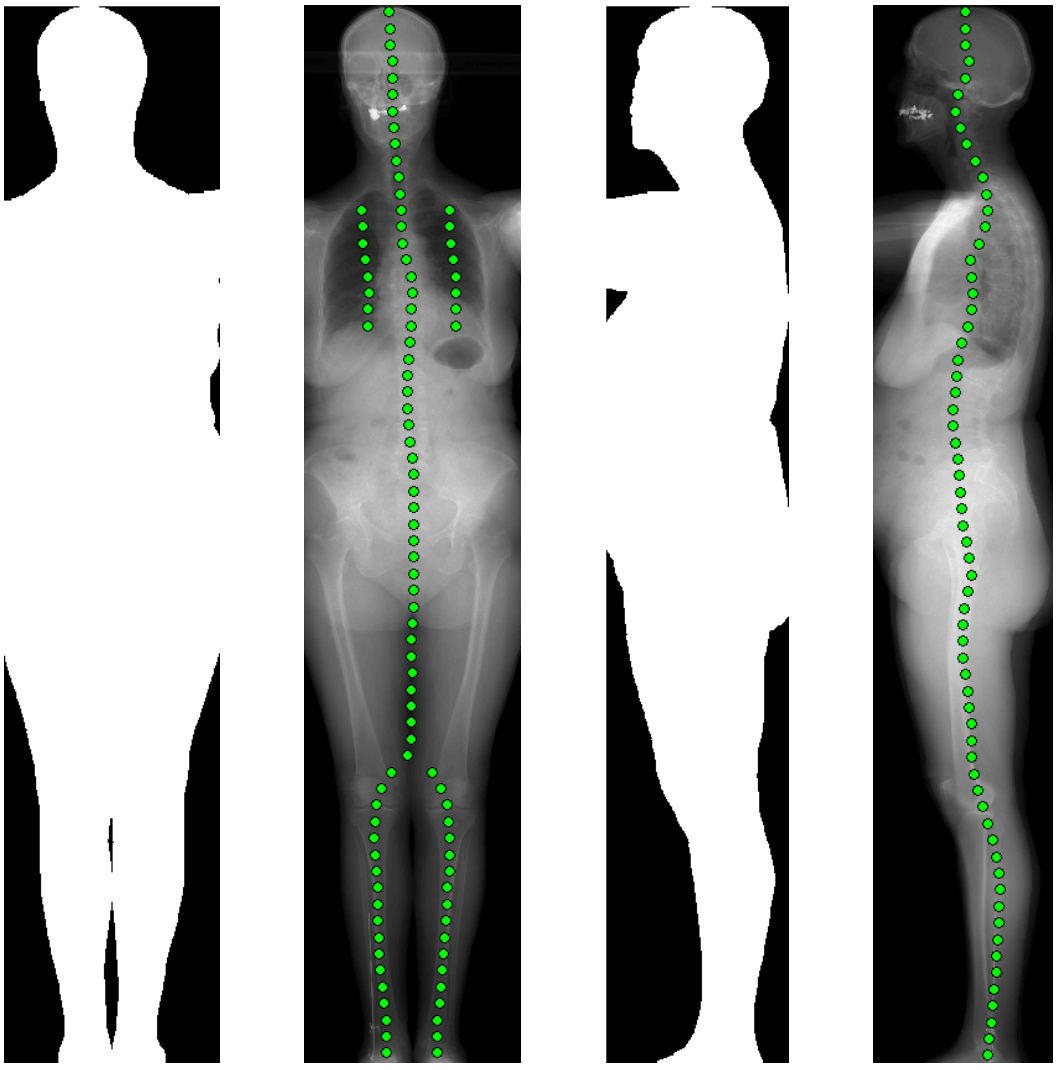
- the envelop of the patient significantly varies depending on the size of the patient, his/her posture and position in the device;
- the segmentation is not affected by changes in terms of radiation exposure, e.g. on preview images the method works as well;
- the landmarks $b_j \in \mathcal{B}_r$ do not consider deformations or misplacement of the center of gravity of the patient with respect to the center of the image;
- on frontal view acquisitions the size of the patient significantly influences the positions of the landmarks $b_j \in \mathcal{B}_r$, $r = 5, 6$ (pelvis and femur);

- on lateral view acquisitions the posture of the patient can significantly influence the positions of the landmarks $b_j \in \mathcal{B}_2$ (thoracic spine).



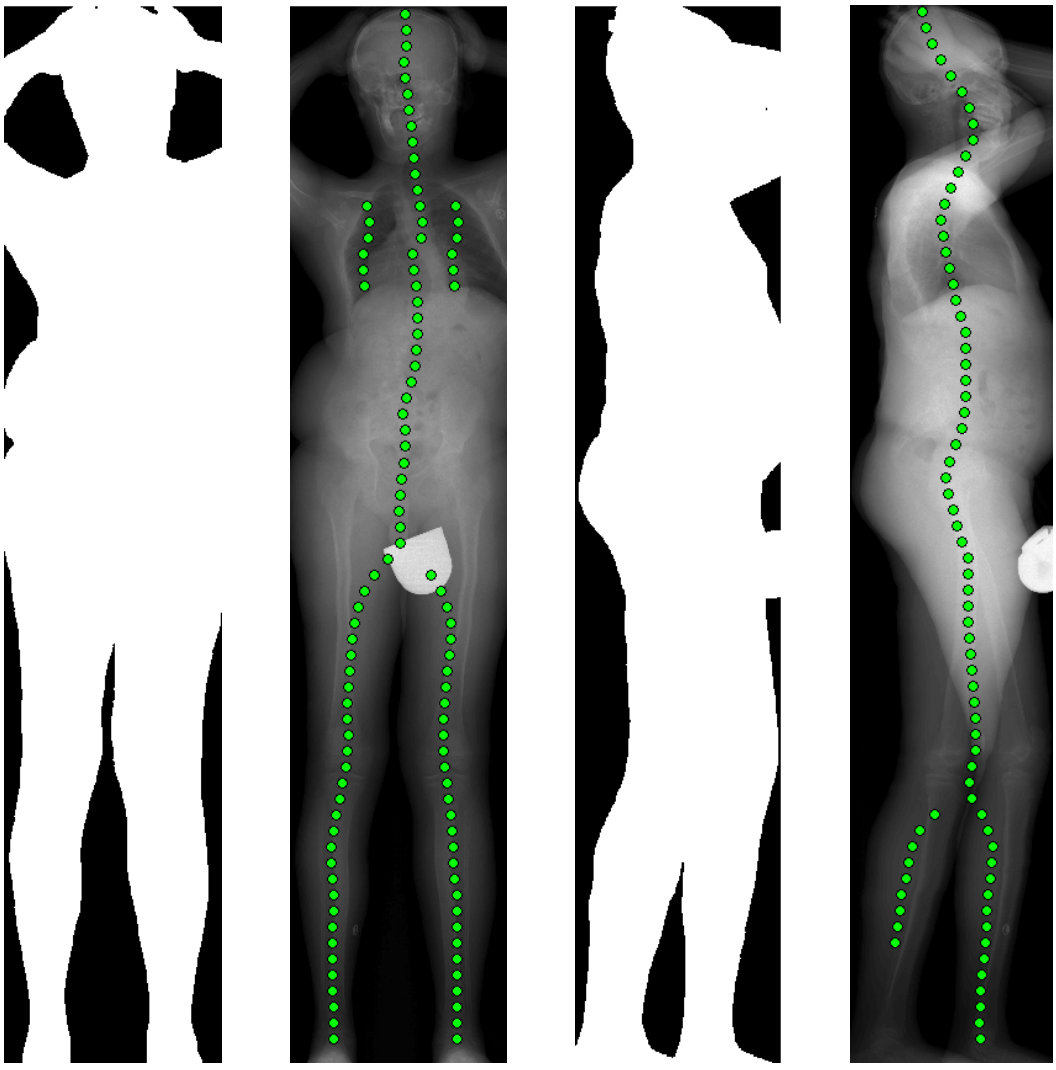
(a) (b) (c) (d)

Figure A.3: *Patient 64: (a) mask of the envelop on the frontal view; (b) landmarks computed from (a); (c) mask of the envelop on the lateral view; (d) landmarks computed from (c).*



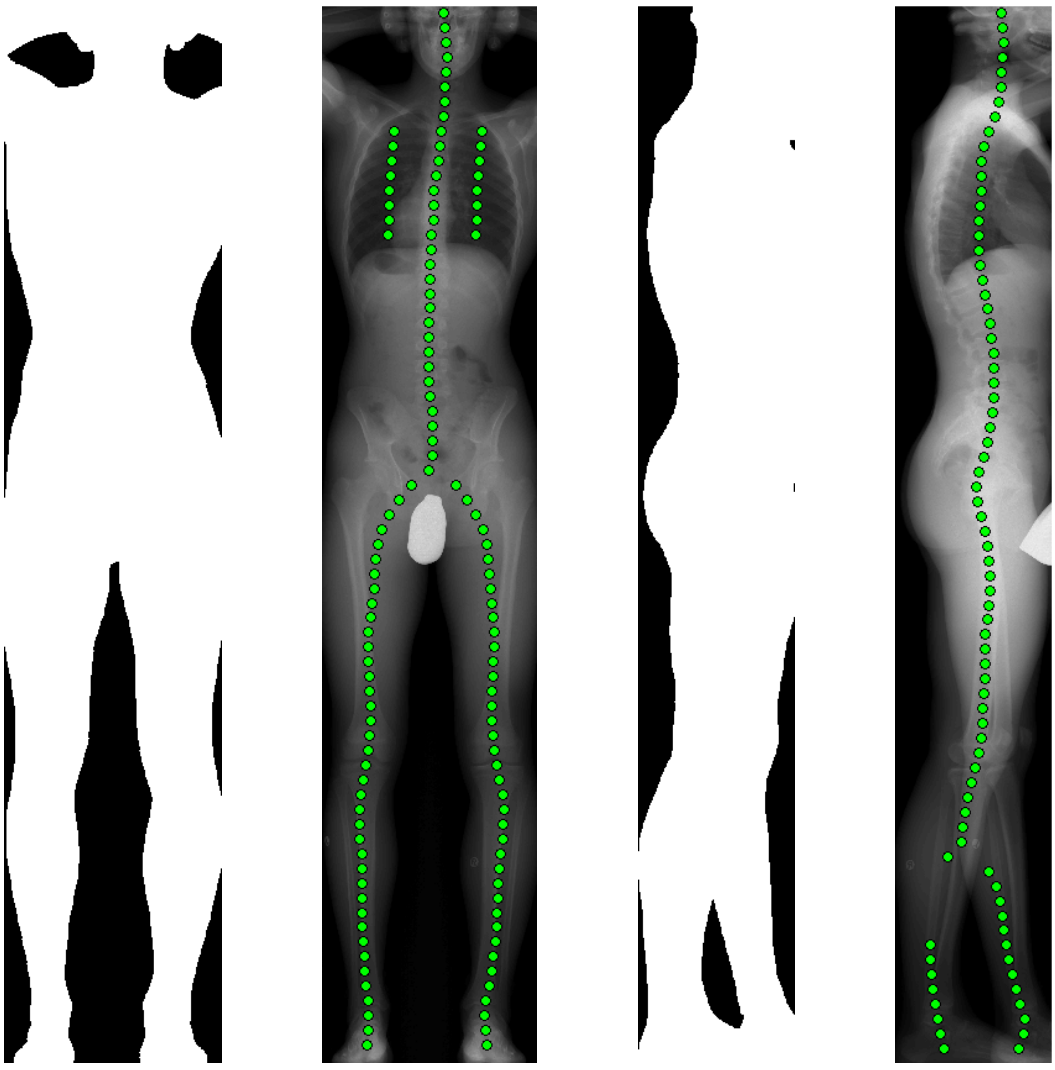
(a) (b) (c) (d)

Figure A.4: *Patient 80: (a) mask of the envelop on the frontal view; (b) landmarks computed from (a); (c) mask of the envelop on the lateral view; (d) landmarks computed from (c).*



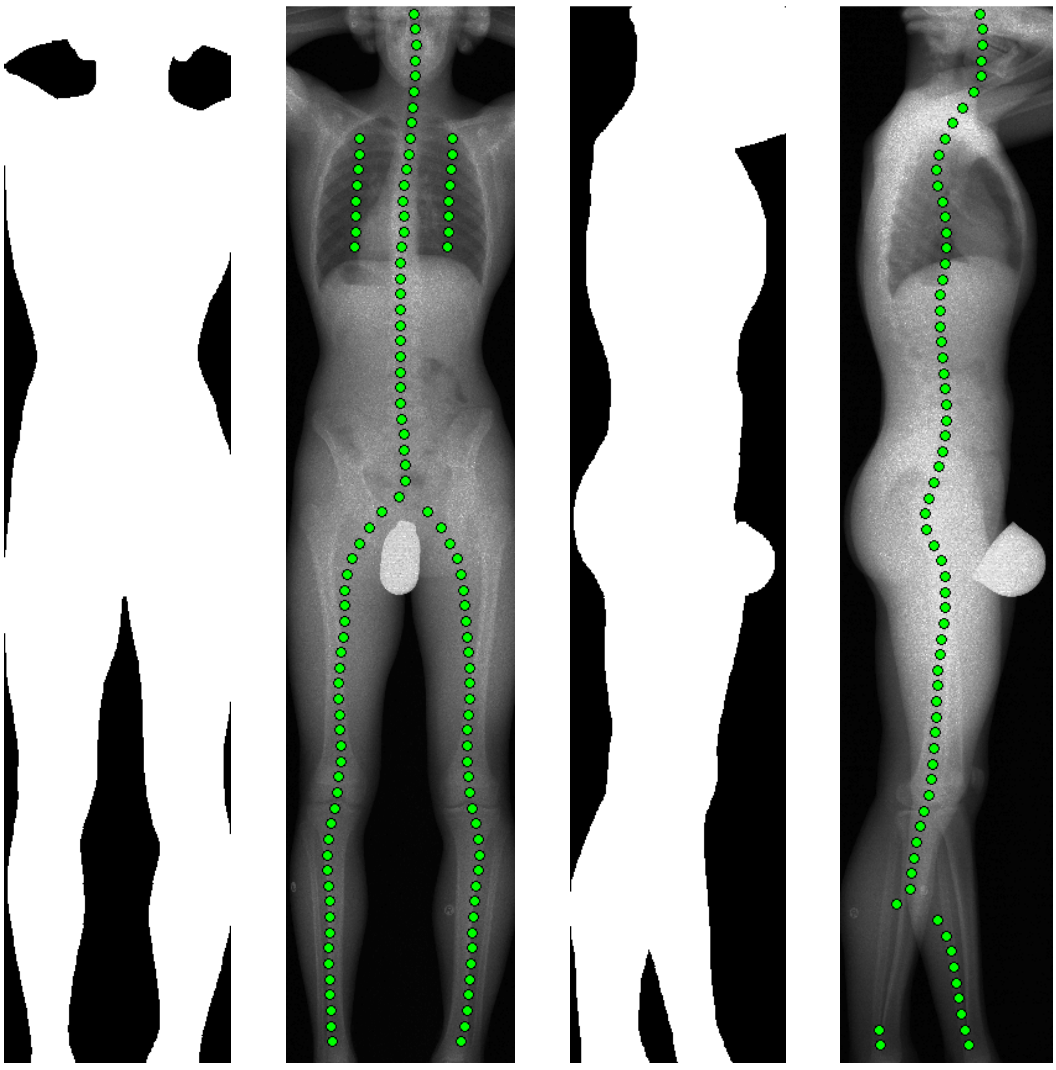
(a) (b) (c) (d)

Figure A.5: *Patient 27: (a) mask of the envelop on the frontal view; (b) landmarks computed from (a); (c) mask of the envelop on the lateral view; (d) landmarks computed from (c).*



(a) (b) (c) (d)

Figure A.6: *Patient 23: (a) mask of the envelop on the frontal view; (b) landmarks computed from (a); (c) mask of the envelop on the lateral view; (d) landmarks computed from (c).*



(a) (b) (c) (d)

Figure A.7: *Preview of the patient 23: (a) mask of the envelop on the frontal view; (b) landmarks computed from (a); (c) mask of the envelop on the lateral view; (d) landmarks computed from (c).*

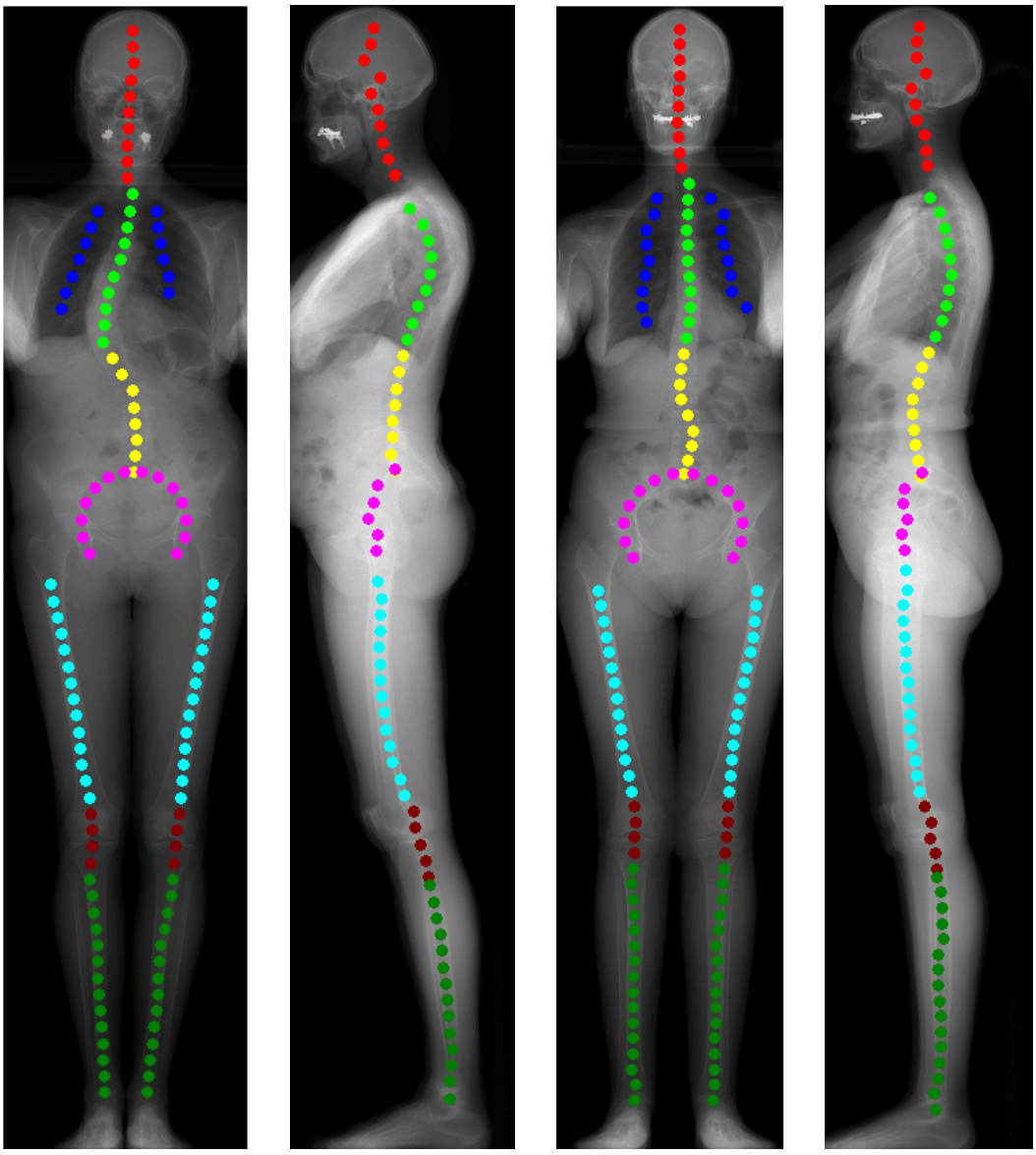
Appendix B

Qualitative evaluation of landmarks detection and recognition

We present here a sequence of examples that display on EOS images the positions of the landmarks automatically detected with the method presented in Chapter 5. Sections B.1 and B.2 show these results on a random selection of 6 patients from the adult subset ($\mathcal{D}_a$), and of 6 patients from the young subset ($\mathcal{D}_b$) with the 6 corresponding preview acquisitions ($\mathcal{D}_c$), respectively. The qualitative analysis of the results bring to the following observations:

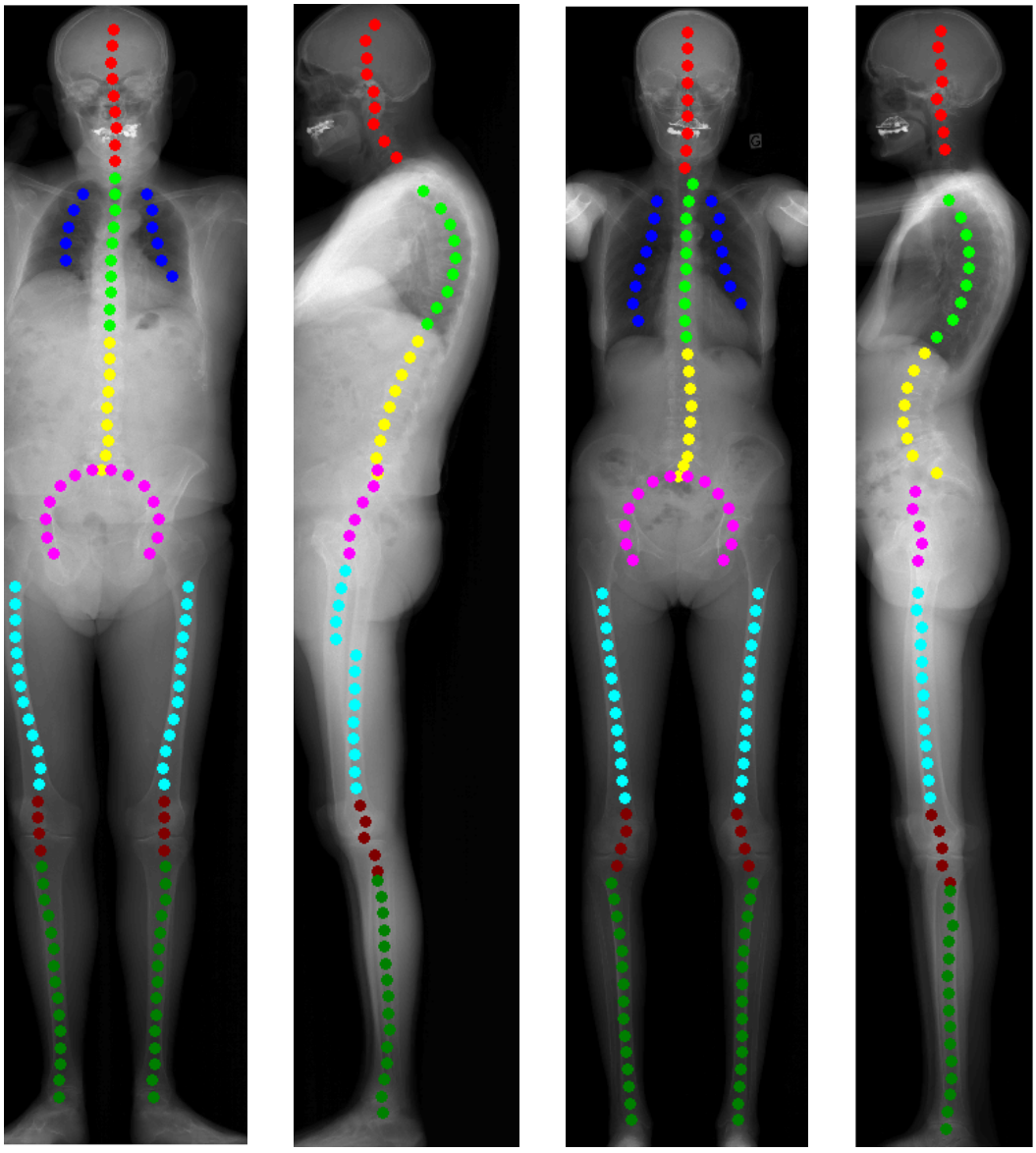
- Adult patients:
  - the landmarks on both frontal and lateral view acquisitions are well located despite the heterogeneity of the data;
  - occasionally, some landmarks could be not well placed but not so far from the corresponding structures of interest to cause significant EI estimation errors.
- Young patients:
  - the landmarks on the frontal view acquisitions are well placed in all regions;
  - in spite of a possible bias on the measure associated with the thoracic spine seen from the lateral view (see Section 5.7.2), the landmarks in this region are generally well placed. Potential estimation errors are not due to the detection, but rather to the outliers caused by the superposition of some thoracic vertebrae with the shoulders;
  - the detection is however quite unstable when the legs are partially superposed in the image. A better solution should be then conceived if the users consider important to have an EI estimation in this particular case.

B.1 Detection on adult patients



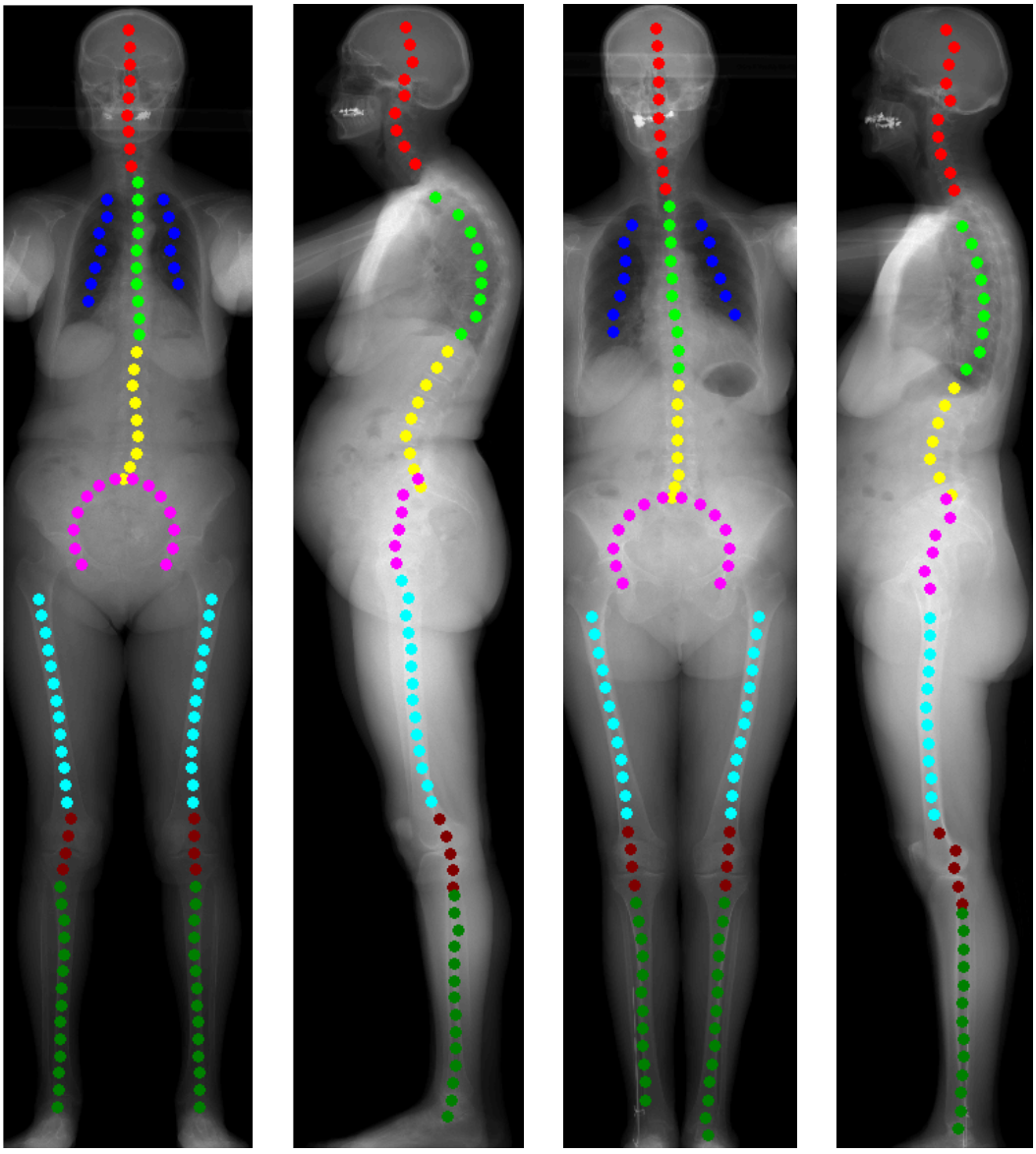
(a) (b) (c) (d)

Figure B.1: *Database $\mathcal{D}_a$: (a,b) patient number 8; (c,d) patient number 62.*



(a) (b) (c) (d)

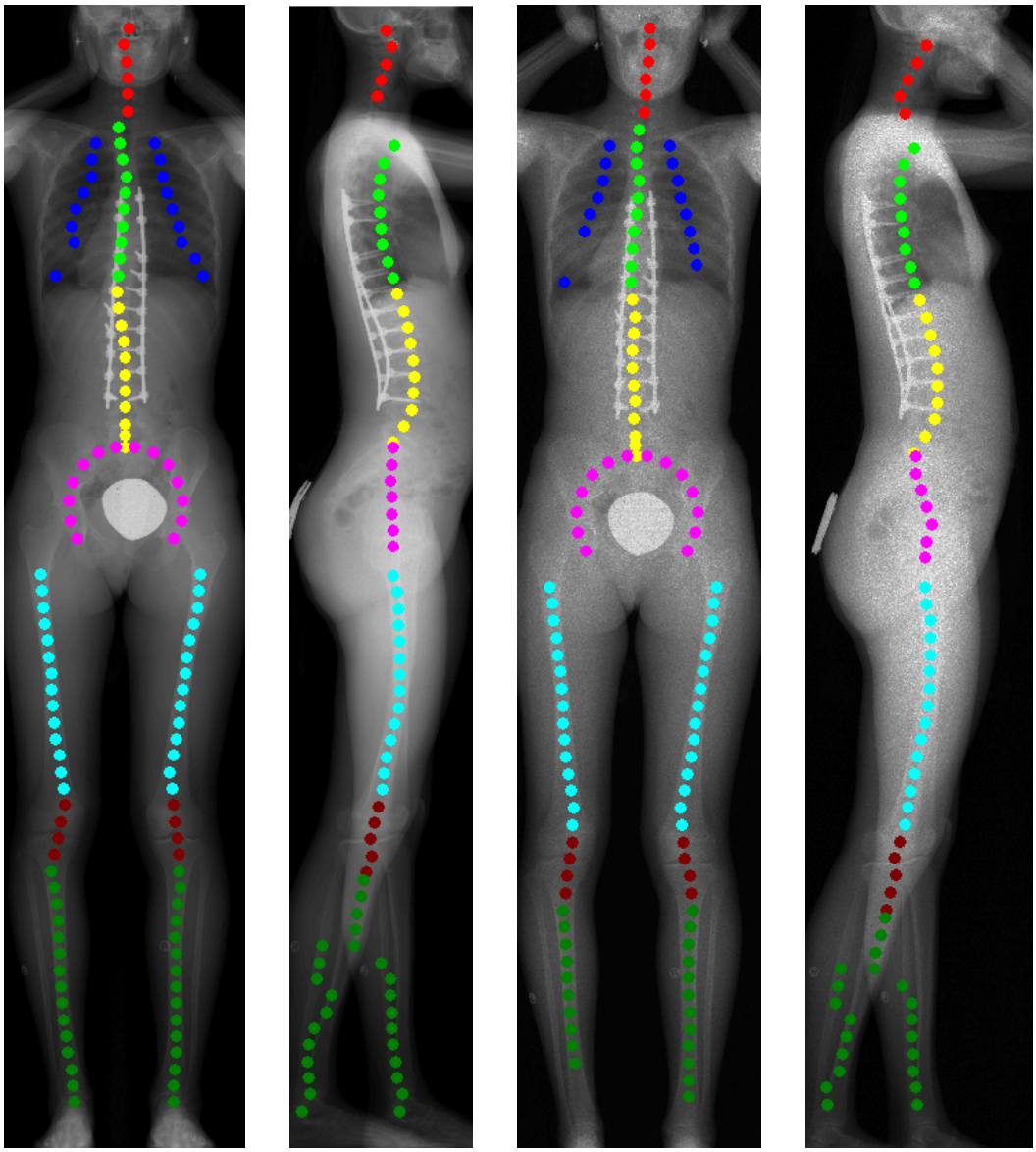
Figure B.2: *Database $\mathcal{D}_a$: (a,b) patient number 77; (c,d) patient number 78.*



(a) (b) (c) (d)

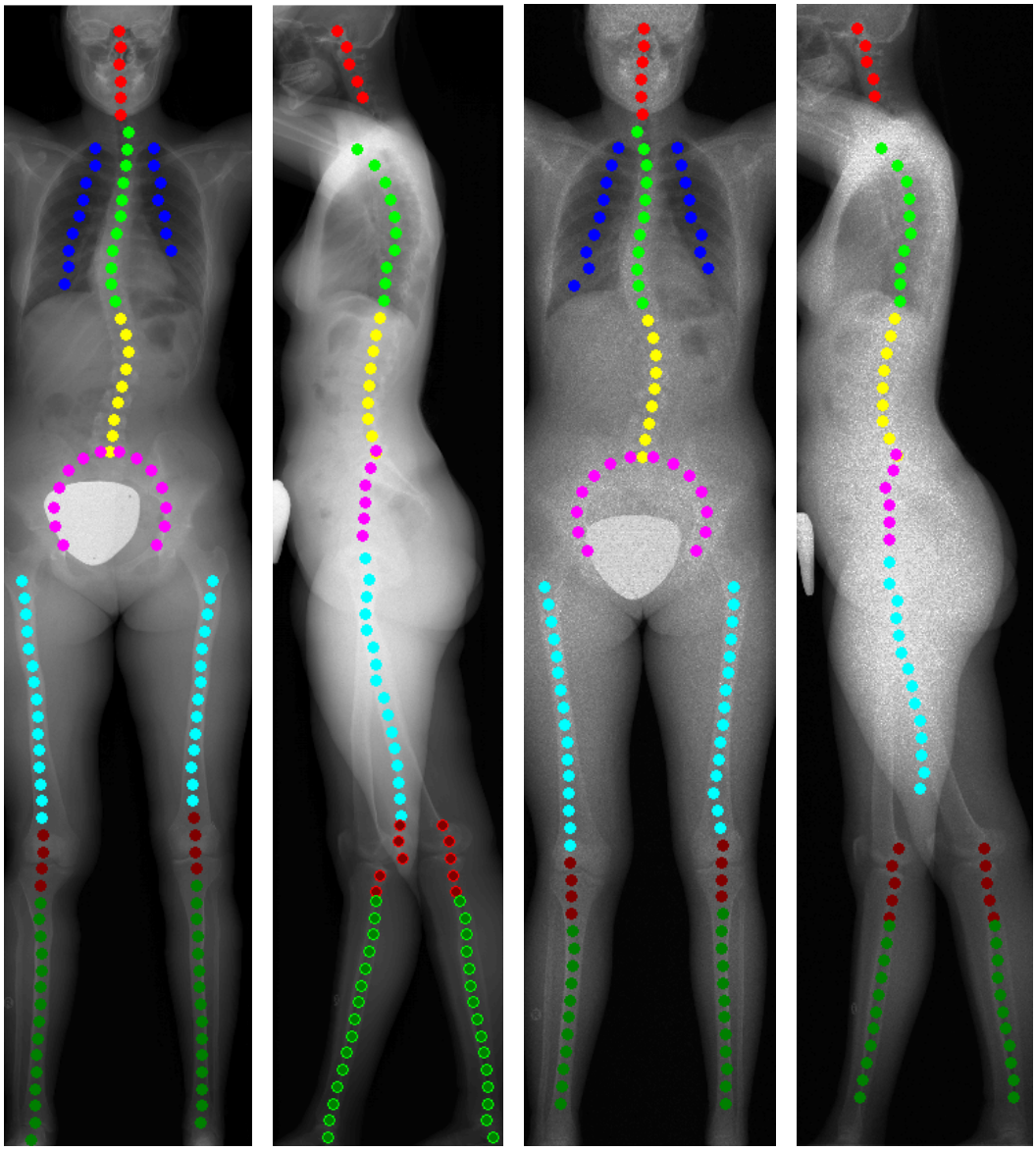
Figure B.3: *Database $\mathcal{D}_a$: (a,b) patient number 10; (c,d) patient number 80.*

B.2 Detection on young patients and corresponding preview acquisitions



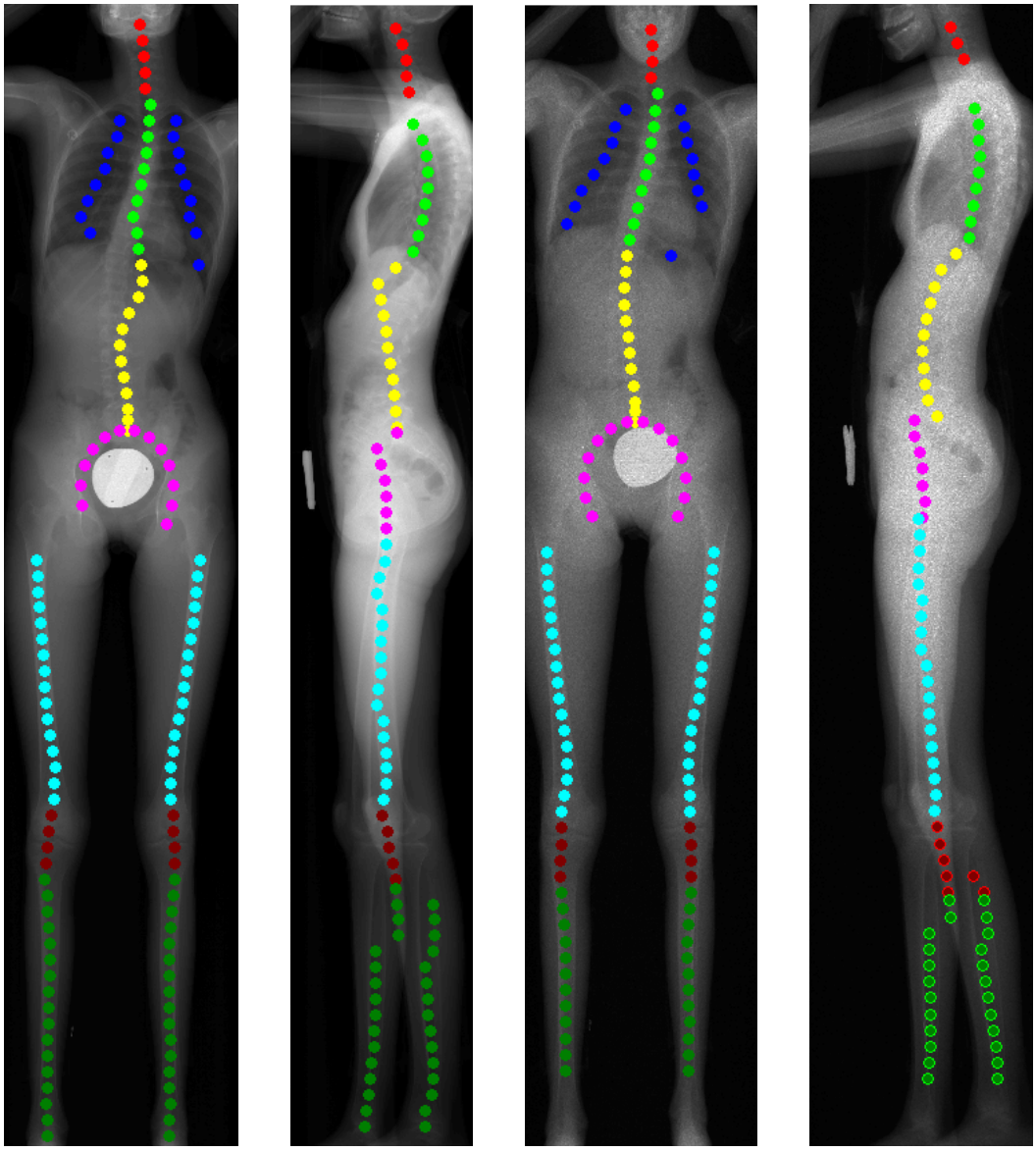
(a) (b) (c) (d)

Figure B.4: *Patient number 3: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*



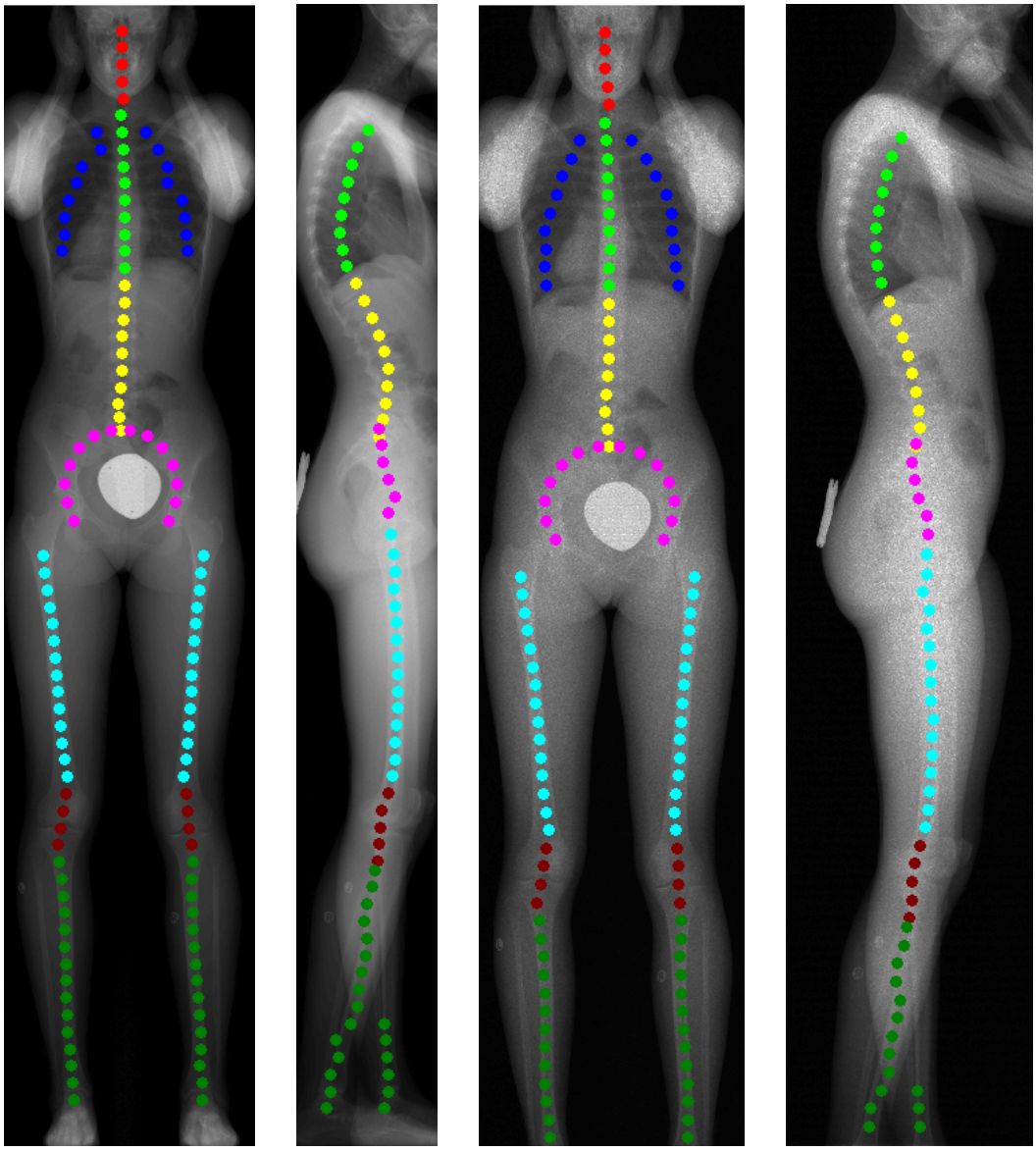
(a) (b) (c) (d)

Figure B.5: *Patient number 9: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*



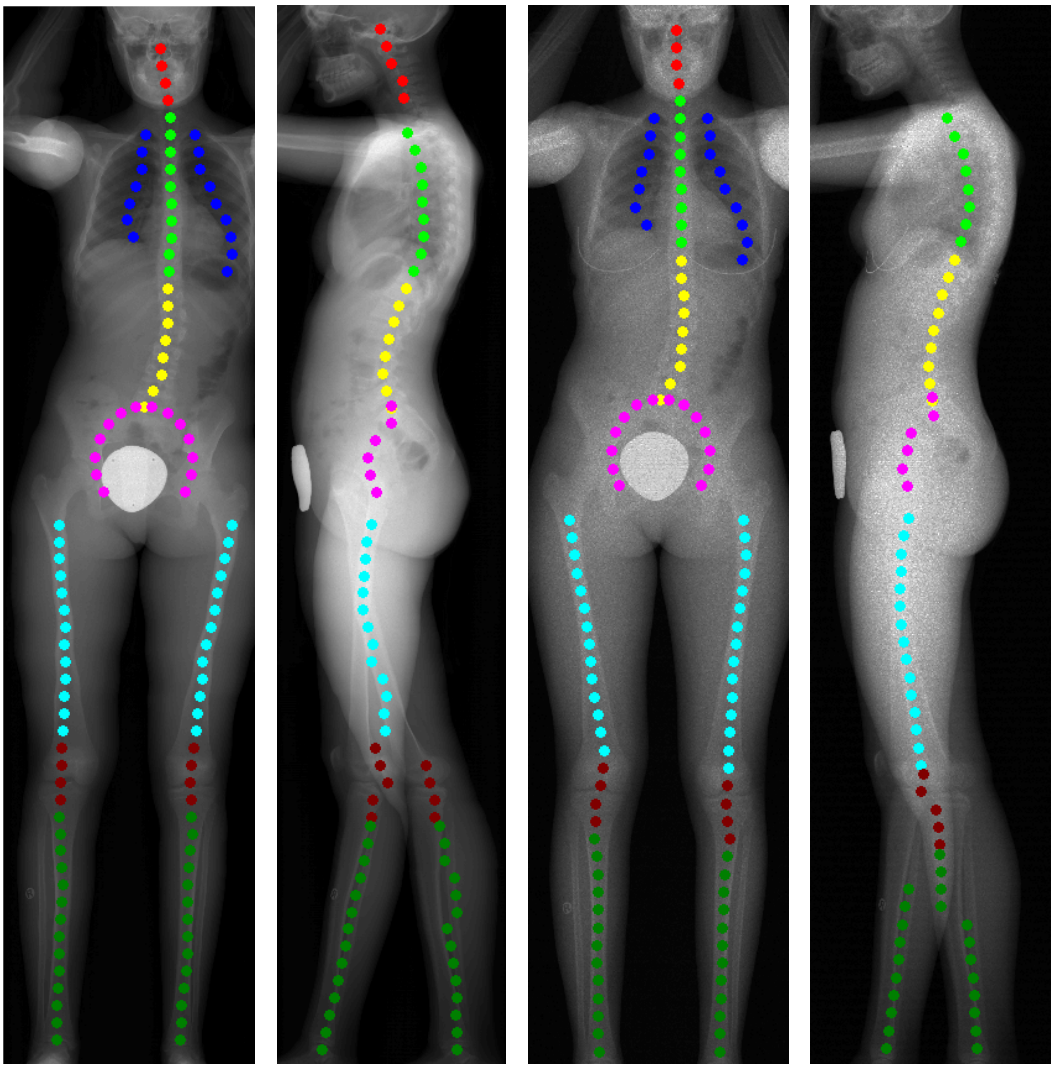
(a) (b) (c) (d)

Figure B.6: *Patient number 33: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*



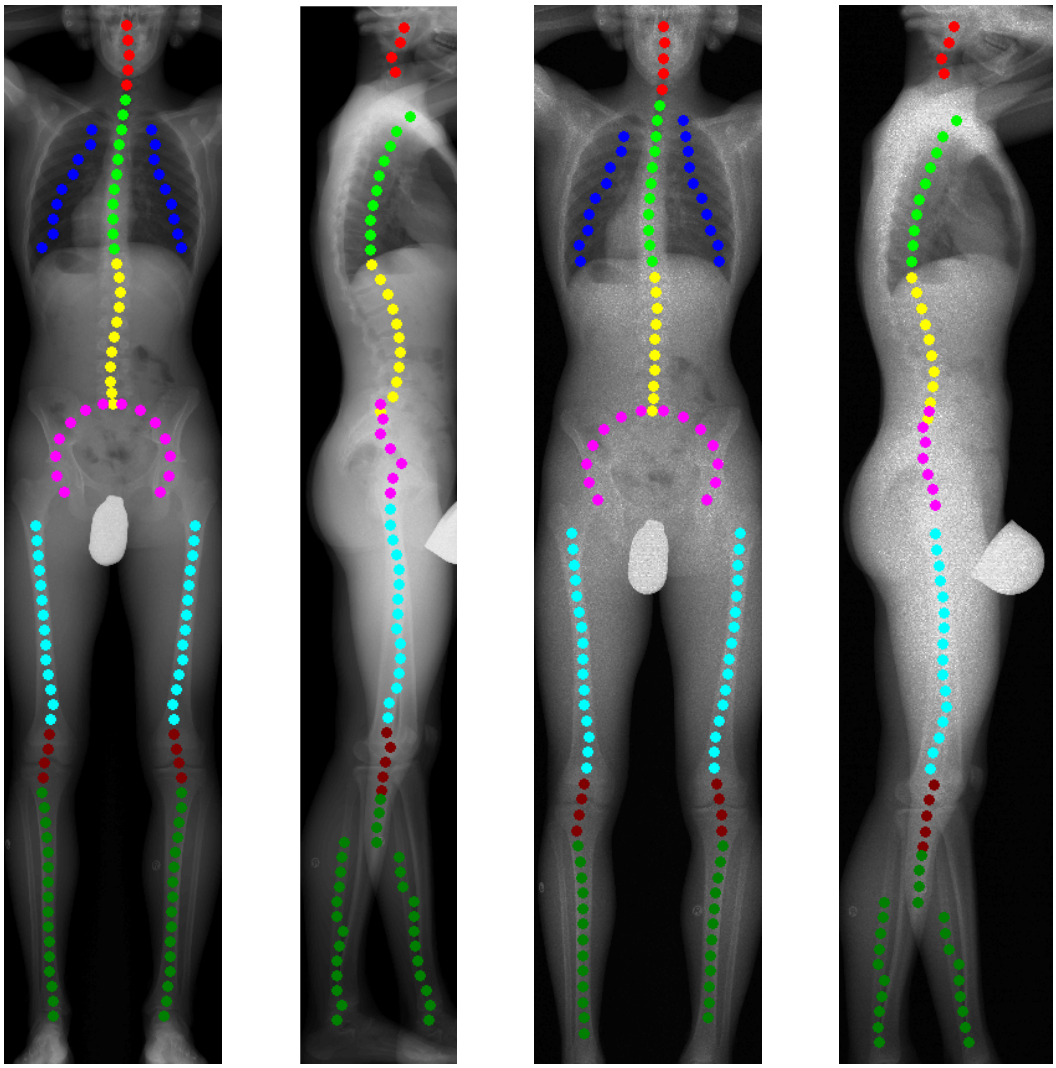
(a) (b) (c) (d)

Figure B.7: *Patient number 47: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*



(a) (b) (c) (d)

Figure B.8: *Patient number 49: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*



(a) (b) (c) (d)

Figure B.9: *Patient number 23: (a,b) diagnostic ($\mathcal{D}_b$); (c,d) preview ($\mathcal{D}_c$).*

Appendix C

List of the main acronyms and notations

- Chapter 1
  - kV: peak kilo Voltage applied to the X-ray tube
  - mA: X-ray tube current
  - C: scanning speed of the c-arm
  - PACS: Pictural Archival and Communication System
  - DR: Digital Radiography
  - CR: Computed Radiography
  - CT: Computed Tomography
  - ALARA: As Low As Reasonably Achievables
- Chapter 2
  - AWGN: Additive White Gaussian Noise
  - MSE: Mean Square Error
  - MAP: Maximum A Posteriori
  - NL-means: Non Local-means filter
  - MS: Multiple Scale levels
  - LP: Laplacian Pyramid
  - FWT: Fast Wavelet Transform
  - IUWT: Isotropic Undecimated Wavelet Transform
  - PMMA: Polymethyl Methacrylate
  - ROI: Region Of Interest
  - SNR: Signal to Noise Ratio
  - CNR: Contrast to Noise Ratio
  - DYN: Dynamic of the gray level distribution
  - ALV: Average Local Variance

  - CII: Contrast Improvement Index
  - AC: Contrast in Anatomical regions of interest
  - SCII: Signal Contrast Improvement Index
  - NCII: Noise Contrast Improvement Index
  - UCII: Unbiased Contrast Improvement Index

- Chapter 3
  - RMSE: Root Mean Square Error
  - XNL-means: X-ray Non Local-means filter
  - LNC-maps: Local Noise Containment-maps
  - NE: contrast enhancement algorithm that takes as input the noisy image
  - DE: contrast enhancement algorithm that takes as input the image filtered with the XNL-means filter
  - LNCE: contrast enhancement algorithm where the noise is contained by using the LNC-maps
  - EOSE: contrast enhancement algorithm used in clinical routine with the EOS system
  - BMI: Body Mass Index

- Chapter 4
  - EI: Exposure Index
  - DI: Deviation Index
  - AEC: Automatic Exposure Control
  - hROI: Definition of a region of interest based on measures from the histogram
  - sROI: Definition of a regions of interest at the center of the image
  - bbDet: Detection of bounding boxes
  - lDet: Detection of landmarks

- Chapter 5
  - $\mathcal{P}_{h,S}$: set of salient points at scale S that correspond to peaks of absorption
  - $\mathcal{P}_{l,S}$: set of salient points at scale S that correspond to peaks of signal at the detector
  - $\mathcal{A}_r$: anatomical region of interest r
  - $\mathcal{M}_r$: set of manually annotated landmarks associated with the region of interest r
  - $\mathcal{L}_r$: set of automatically detected landmarks associated with the region of interest r

Bibliography

M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, October 2006.

S. Alliney. Digital filters as absolute norm regularizers. *IEEE Transactions on Signal Processing*, 40(6):1548–1562, Jun 1992.

F. J. Anscombe. The transformation of Poisson binomial and negative-binomial data. *Biometrika*, 35(3):246–254, December 1948.

J. Atif, C. Hudelot, G. Fouquier, I. Bloch, and E. Angelini. From generic knowledge to specific reasoning for medical image interpretation using graph based representations. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 224–229, 2007.

I. Bloch. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89 – 110, 2005.

I. Bloch, T. Géraud, and H. Maître. Representation and fusion of heterogeneous fuzzy information in the 3d space for model-based structural recognition—application to 3d brain imaging. *Artificial Intelligence*, 148(1–2):141 – 175, 2003.

A. Buades, B. Coll, and J. M. Morel. A Non-Local Algorithm for Image Denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005a.

A. Buades, B. Coll, and J-M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling and Simulation*, 4:490–530, 2005b.

F. M. Buck, R. Guggenberger, P. P. Koch, and C. Pfirrmann. Femoral and tibial torsion measurements with 3D models based on low-dose biplanar radiographs in comparison with standard CT measurements. *American Journal of Roentgenology*, 199(5):W607–W612, 2012.

P. J. Burt and E. H. Adeldon. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540, 1983.

E. Candes. *Ridgelets: theory and applications*. PhD thesis, Standford University, 1998.

E. J. Candes and D. L. Donoho. Curvelets and curvilinear integrals. *Journal of Approximation Theory*, 113(1):59 – 90, 2001.

E. J. Candes and M. B. Wakin. An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008. ISSN 1053-5888.

T. Cerciello, P. Bifulco, M. Cesarelli, and A.Fratini. A comparison of denoising methods for x-ray fluoroscopic images. *Biomedical Signal Processing and Control*, 7(6):550 – 559, 2012.

D. C. Chang and W. R. Wu. Image Contrast Enhancement Based on a Histogram Transformation of Local Standard Deviation. *IEEE Transactions on Medical Imaging*, 17(4):518 –531, August 1998.

G. Charpak, R. Bouclier, T. Bressani, J. Favier, and Č. Zupančič. The use of multiwire proportional counters to select and localize charged particles. *Nuclear Instruments and Methods*, 62(3):262 – 268, 1968.

P. Chatterjee and P. Milanfar. Patch-based near-optimal image denoising. *IEEE Transactions on Image Processing*, 21(4):1635–1649, April 2012.

O. Colliot, O. Camara, and I. Bloch. Integration of fuzzy spatial relations in deformable models—application to brain {MRI} segmentation. *Pattern Recognition*, 39(8):1401 – 1414, 2006.

M. Colom and A. Buades. Analysis and Extension of the Percentile Method, Estimating a Noise Curve from a Single Image. *Image Processing On Line*, 3:332–359, 2013.

S. Cotter, R. Adler, R. Rao, and K. Kreutz-Delgrado. Forward Sequential Algorithms for the Basis Selection. *IEE Proceedings Vision, Image and Signal Processing*, 146(5):235–244, 1999.

P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, 2008.

A. Criminisi, J. Shotton, and S. Bucciarelli. Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In *MICCAI workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA)*, 2009.

A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis*, 17(8):1293 – 1303, 2013.

M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, Apr 1998.

K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian. Image Denoising by Sparse 3-D Transform- Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian. BM3D Image Denoising with Shape-Adaptive Principal Component Analysis. In *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations*, pages 1–6, 2009.

J. Damet, P. Fournier, P. Monnin, M. Sans-Merce, D. Ceroni, T. Zand, F. R. Verdun, and S. Baechler. Occupational and patient exposure as well as image quality for full spine examinations with the eos imaging system. *Medical Physics*, 41(6):1–12, 2014.

J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G. J. Jensen. Fast nonlocal filtering applied to electron cryomicroscopy. In *International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1331–1334. IEEE, 2008.

I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

M. Delbracio, P. Musé, A. Buades, J. Chauvier, N. Phelps, and J-M. Morel. Boosting Monte Carlo Rendering by Ray Histogram Fusion. *ACM Transaction on Graphic*, 33(1):8:1–8:15, February 2014.

C. Deledalle. *Image denoising beyond additive Gaussian noise*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, November 2011.

C. Deledalle, F. Tupin, and L. Denis. Poisson NL Means: Unsupervised Non Local Means for Poisson Noise. In *IEEE International Conference on Image Processing (ICIP)*, pages 801–804, Hong Kong, China, September 2010.

C. Deledalle, V. Duval, and J. Salmon. Anisotropic non-local means with spatially adaptive patch shapes. In *Scale Space and Variational Methods in Computer Vision*, pages 231–242, 2011.

C. Deledalle, L. Denis, and F. Tupin. How to compare noisy patches? Patch similarity beyond Gaussian noise. *International Journal of Computer Vision*, 99(1): 86–102, 2012. ISSN 0920-5691.

S. Deschênes, G. Charron, and G. Beaudoin. Diagnostic imaging of spinal deformities: reducing patients radiation dose with a new slot-scanning X-ray imager. *The Spine Journal*, 35(9):989–994, 2010.

P. Despres, G. Beaudoin, P. Gravel, and J. A. de Guise. Physical characteristics of a low-dose gas microstrip detector for orthopedic x-ray imaging. *Medical physics*, 32(4):1193–1204, 2005.

S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert. Multiscale Contrast Enhancement for Radiographies: Laplacian Pyramid Versus Fast Wavelet Transform. *IEEE Transactions on Medical Imaging*, 21(4):343 –353, April 2002.

D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

J. Dubousset, G. Charpak, I. Dorion, W. Skalli, F. Lavaste, J. Deguise, G. Kalifa, and S. Ferey. A new 2D and 3D imaging approach to musculoskeletal physiology and pathology with low-dose radiation and the standing position: the EOS system. *Bulletin de l'Académie nationale de médecine*, 189(2):287–97, 2005.

V. Duval, J. Aujol, and Y. Gousseau. A Bias-Variance Approach for the Nonlocal Means. *SIAM Journal on Imaging Sciences*, 4(2):760–788, 2011.

A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, volume 2, pages 1033–1038 vol.2, 1999.

B. Fan and H. Han. Medical image enhancement based on modified lut-mapping derivative and multi-scale layer contrast modification. In *4th International Congress on Image and Signal Processing (CISP)*, pages 696 – 703, October 2011.

Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transaction on Graphics*, 27(3):67:1 – 67:10, August 2008.

R. Fattal. Edge-avoiding wavelets and their applications. *ACM Transaction on Graphics*, 28(3):1–10, August 2009.

R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale Shape and Detail Enhancement from Multi-light Image Collections. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3):1–9, August 2007.

M. Fenchel, S. Thesen, and A. Schilling. Automatic labeling of anatomical structures in mr fastview images using a statistical atlas. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, volume 5241. Springer Berlin Heidelberg, 2008.

A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, Oct 2008.

D. H. Foos, X. Wang, and W. J. Sehnert. Low-dose automatic exposure control system for digital portable x-ray imaging, 04 2012.

G. Fouquier, J. Atif, and I. Bloch. Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. *Computer Vision and Image Understanding*, 116(1):146 – 165, 2012.

M. Gao, Y. Zhan, G. H. Valadez, Y. Shinagawa, D. N. Metaxas, and X. S. Zhou. Saliency-based rotation invariant descriptor for wrist detection in whole body ct images. In *ISBI*, pages 117–120, 2014.

R. Gauriau, R. Cuingnet, R. Prevost, B. Mory, R. Ardon, D. Lesage, and I. Bloch. A Generic, Robust and Fully-Automatic Workflow for 3D CT Liver Segmentation. In *Abdominal Imaging. Computation and Clinical Applications*, volume 8198, pages 241–250. Springer Berlin Heidelberg, 2013.

R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch. Multi-organ Localization Combining Global-to-Local Regression and Confidence Maps. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, volume 8675, pages 337–344. Springer Berlin Heidelberg, 2014.

G. Gilboa, N. Sochen, and Y. Y. Zeevi. Variational denoising of partly textured images by spatially varying constraints. *IEEE Transactions on Image Processing*, 15(8):2281–2289, Aug 2006.

B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu. Vertebrae Localization in Pathological Spine CT via Dense Classification from Sparse Annotations. In *MICCAI (3)*, volume 7512, pages 590–598. Springer, 2012.

J. W. Gofman. *Radiation and human health*. Sierra Club Books,San Francisco, CA, 1981.

M. J. Goske, K. E. Applegate, J. Boylan, P. F. Butler, M. J. Callahan, B. D. Coley, S. Farley, D. P. Frush, M. Hernanz-Schulman, and D. Jaramillo. The 'image gently'campaign: increasing ct radiation dose awareness through a national education and awareness program. *Pediatric radiology*, 38(3):265–269, 2008.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion . *NeuroImage*, 33(1):115 – 126, 2006.

M. Hensel, B. Lundt, T. Pralow, and R-R. Grigat. Robust and Fast Estimation of Signal-Dependent Noise in Medical X-Ray Image Sequences. *Bildverarbeitung für die Medizin 2006*, pages 46–50, 2006.

IEC 62494-1. Medical electrical equipement - Exposure index of digital X-ray imaging systems - Part 1: Definitions and requirements for general radiography, 2008.

B. Ilharreborde, C. Vidal, W. Skalli, and K. Mazda. Sagittal alignment of the cervical spine in adolescent idiopathic scoliosis treated by posteromedial translation. *European Spine Journal*, 22(2):330–337, 2013.

T. Illés, M. Tunyogi-Csapó, and S. Somoskeöy. Breakthrough in three-dimensional scoliosis diagnosis: significance of horizontal plane view and vertebra vectors. *European Spine Journal*, 20(1):135–143, 2011.

L. Itti and C. Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

Y. Jin, L. M. Fayad, and A. F. Laine. Contrast enhancement by multiscale adaptive histogram equalization. In *Proc. SPIE*, pages 206–213, 2001.

K. R. Kase, B. E. Bjärngard, and F. H. Attix. *The dosimetry of ionizing radiation*, volume 3. Elsevier, 2012.

V. Katkovnik, K. Egiazarian, and J. Astola. Adaptive Window Size Image De-noising Based on Intersection of Confidence Intervals (ICI) Rule. *Journal of Mathematical Imaging and Vision*, 16(3):223–235, 2002.

C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *International Journal of Computer Vision*, 79(1):45–69, 2008. ISSN 0920-5691.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 03 1951.

A. Laine, J. Fan, and W. Yang. Wavelets for Contrast Enhancement of Digital Mammography. *IEEE Engineering in Medicine and Biology Magazine*, 14:536–550, September/October 1995.

N. Lay, N. Birkbeck, J. Zhang, and S. K. Zhou. Rapid Multi-organ Segmentation Using Context Integration and Discriminative Models. In *Information Processing in Medical Imaging*, volume 7917 of *Lecture Notes in Computer Science*, pages 450–462. Springer Berlin Heidelberg, 2013.

J-Y. Lazennec, A. Brusson, and M-A. Rousseau. Hip–spine relations and sagittal balance clinical consequences. *European spine journal*, 20(5):686–698, 2011.

J-C. Le Huec, P. Leijssen, M. Duarte, and S. Aunoble. Thoracolumbar imbalance analysis for osteotomy planification using a new method: Fbi technique. *European Spine Journal*, 20(5):669–680, 2011.

M. Lebrun. An analysis and implementation of the BM3D image denoising method. *Image Processing On Line*, 2:175–213, 2012.

M. Lebrun, M. Colom, A. Buades, and J-M. Morel. Secrets of image denoising cuisine. *Acta Numerica*, 21:475–576, 5 2012. ISSN 1474-0508.

M. Lebrun, A. Buades, and J-M. Morel. A Nonlocal Bayesian Image Denoising Algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013. doi: 10.1137/120874989.

Y. Li, L. Sharan, and E. H. Adelson. Compressing and Companding High Dynamic Range Images with Subband Architecture. *ACM Transactions on Graphics*, 24 (3):836–844, 2005.

C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, and T. F. Cootes. Accurate Fully Automatic Femur Segmentation in Pelvic Radiographs Using Regression Voting. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, volume 7512, pages 353–360. Springer Berlin Heidelberg, 2012.

C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, and T. F. Cootes. Accurate Bone Segmentation in 2D Radiographs Using Fully Automatic Shape Model Matching Based On Regression-Voting. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, volume 8150, pages 181–189. Springer Berlin Heidelberg, 2013.

D. Liu, K.S. Zhou, D. Bernhardt, and D. Comaniciu. Search strategies for multiple landmark detection by submodular maximization. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2831–2838, June 2010.

D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal on Computer Vision*, 60(2):91–110, November 2004.

A. Loza, D. Bull, N. Canagarajah, and A. Achim. Non-Gaussian Model-based Fusion of Noisy Images in the Wavelet Domain. *Computer Vision and Image Understanding*, 114(1):54–65, 2010.

A. Loza, P. Hill, P. Verkade, D. Bull, Al-Mualla M., and A. Achim. Joint denoising and contrast enhancement for light microscopy image sequences. In *International Symposium on Biomedical Imaging*, pages 1083–1086, April 2014.

F. Luisier, C. Vonesch, T. Blu, and M. Unser. Fast Interscale Wavelet Denoising of Poisson-Corrupted Images. *Signal Processing*, 90(2):415–427, February 2010.

F. Luisier, T. Blu, and M. Unser. Image Denoising in Mixed Poisson-Gaussian Noise. *IEEE Transactions on Image Processing*, 20(3):696–708, March 2011.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, 1967.

J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-Local Sparse Models for Image Restoration. In *IEEE 12th International Conference on Computer Vision*, pages 2272–2279, 2009.

M. Makitalo and A. Foi. Optimal Inversion of the Anscombe Transformation in Low-Count Poisson Image Denoising. *IEEE Transactions on Image Processing*, 20(1):99–109, January 2011.

S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008.

S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transitions of Pattern Analysis and Machine Intelligence*, 11(7): 674–693, 1989.

J.F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In *In Proceedings 1st MICCAI*, pages 1230–1241, 1998.

R. Mantiuk, K. Myszkowski, and H-P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception*, 3(3):286–308, 2006.

J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm). *NeuroImage*, 2(2, Part A):89 – 101, 1995.

C. H. McCollough. Automatic Exposure Control in CT: Are We Done Yet? *Radiology*, 237(3):755–756, 2005.

D. Menotti, L. Najman, J. Facon, and A. A. De Araujo. Multi-Histogram Equalization Methods for Contrast Enhancement and Brightness Preserving. *IEEE Transactions on Consumer Electronics*, 53(3):1186–1194, 2007.

U. Mothiram, P. C. Brennan, J. Robinson, S. Lewis, and B. Moran. Retrospective evaluation of exposure index (ei) values from plain radiographs reveals important considerations for quality improvement. *Journal of Medical Radiation Sciences*, 60(4):115–122, 2013.

F. Murtagh, J-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics Supplement Series*, 112:179–189, 1995.

M. Nikolova. Minimizers of Cost-Functions Involving Nonsmooth Data-Fidelity Terms. Application to the Processing of Outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2002.

T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.

A. Papadakis, K. Perisinakis, and J. Damilakis. Automatic exposure control in ct: the effect of patient size, anatomical region and prescribed modulation strength on tube current and image quality. *European Radiology*, 24(10):2520–2531, 2014.

S. Paris, S.W. Hasinoff, and J. Kautz. Local Laplacian Filters: Edge-aware Image Processing with a Laplacian Pyramid. *ACM Transaction on Graphics*, 30(4):68:1 – 68:12, July 2011.

O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. Möller, S. Nekolla, and N. Navab. Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, volume 6893, pages 239–247. Springer Berlin Heidelberg, 2011.

E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7(10):2032–2040, October 1990.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3): 355 – 368, 1987.

G. G. Polkowski, R. M. Nunley, E. L. Ruh, B. M. Williams, and R. L. Barrack. Does standing affect acetabular component inclination and version after THA? *Clinical Orthopaedics and Related Research®*, 470(11):2988–2994, 2012.

N. N. Ponomarenko, V. V. Lukin, M. S. Zriakhov, A. Kaarna, and J. Astola. An automatic approach to lossy compression of aviris images. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS.*, pages 472–475, 2007.

V. Portilla, J.and Strela, M. J. Wainwright, and E. P. Simoncelli. Image Denosing Using Scale Mixture of Gaussians in the Wavelet Domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.

G. Ramponi, N. Strobel, S. K. Mitra, and T-H. Yu. Nonlinear unsharp masking methods for image contrast enhancement. *Journal of Electronic Imaging*, 5(3): 353–366, 1996.

C. M. Ronckers, M. M. Doody, J. E. Lonstein, M. Stovall, and Land C. E. Multiple Diagnostic X-rays for Spine Deformities and Risk of Breast Cancer. *Cancer Epidemioligy, Biomarkers and Prevention*, 17(3):605–613, March 2008.

L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992. ISSN 0167-2789.

M. Sakata and K. Ogawa. Noise Reduction and Contrast Enhancement for Small-Dose X-ray Images in Wavelet Domain. In *IEEE Nuclear Science Symposium Conference Record*, pages 2924 – 2929, October 2009.

J. Salmon, C. Deledalle, R. Willett, and Z. T. Harmany. Poisson Noise Reduction with Non-Local PCA. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP12)*, pages 1109 – 1112, March 2012.

A. Seibert and M. Richard. The standardized exposure index for digital radiography: an opportunity for optimization of radiation dose to the pediatric population. *Pediatric Radiology*, 41(5):573–581, 2011.

A. Seibert, D. K. Shelton, and E. H. Moore. Computed radiography x-ray exposure trends. *Academic Radiology*, 3(4):313–318, 1996.

I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

S. J. Shepard, J. Wang, M. Flynn, E. Gingold, L. Goldman, K. Krugh, D. L. Leong, E. Mah, K. Ogden, D. Peck, E. Samei, J. Wang, and C. E. Willis. An exposure indicator for digital radiography: Aapm task group 116 (executive summary). *Medical Physics*, 36(7):2898–2914, 2009.

I. Shuryak, R. K. Sachs, and D. J. Brenner. Cancer Risks After Radiation Exposure in Middle Age. *Journal of the National Cancer Institute*, 102:1628–1636, 2010.

R. Smith-Blindman, D. L. Miglioretti, and E. Johnson. Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 1996-2010. *The Journal of American Medical Association*, 307(22):2400–2409, June 2012.

M. Söderberg and M. Gunnarsson. Automatic exposure control in computed tomography - an evaluation of sytems from different manufactures. *Acta Radiologica*, 51 (6):625–634, 2010.

S. Somoskeöy, M. Tunyogi-Csapó, C. Bogyó, and T. Illés. Accuracy and reliability of coronal and sagittal spinal curvature data based on patient-specific three-dimensional models created by the eos 2d/3d imaging system. *The Spine Journal*, 12(11):1052–1059, 2012.

A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging,*, 32(7):1153–1190, 2013.

M. Stahl, T. Aach, S. Dippel, T. Buzug, R. Wiemker, and U. Neitzel. Noise-Resistant Weak-Structure Enhancement for Digital Radiography. In K. M. Hanson, editor, *SPIE Vol. 3661: Medical Imaging 99: Image Processing*, pages 1406–1417, San Diego, USA, February 1999.

J-L. Starck, J. Fadili, and F. Murtagh. The Undecimated Wavelet Decomposition and its Reconstruction. *IEEE Transactions on Image Processing*, 16(2):297–309, Feb 2007.

C. M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

H. Takeda, S. Farsiu, and P. Milanfar. Kernel Regression for Image Processing and Reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366, 2007.

P. Than, K. Szuper, S. Somoskeöy, V. Warta, and T. Illés. Geometrical values of the normal and arthritic hip and knee detected with the eos imaging system. *International orthopaedics*, 36(6):1291–1297, 2012.

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 839–847. IEEE Computer Society, 1998.

A. Treisman. Preattentive processing in vision. *Computer Vision Graphics and Image Processing*, 31(2):156–177, 1985.

P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, volume 1, pages I–511–I–518 vol.1, 2001.

T. Vrtovec, F. Pernuš, and B. Likar. A review of methods for quantitative evaluation of spinal curvature. *European Spine Journal*, 18(5):593–607, 2009.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert. Automated Abdominal Multi-Organ Segmentation With Subject-Specific Atlas Generation. *IEEE Transactions on Medical Imaging*, 32(9):1723–1730, 2013.

M. Wybier and P. Bossard. Musculoskeletal imaging in progress: the eos imaging system. *Joint Bone Spine*, 80(3):238–243, 2013.

Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-Dose X-ray CT Reconstruction via Dictionary Learning. *IEEE Transactions on Medical Imaging*, 31(9):1682–1697, 2012.

L. P. Yaroslavsky, K. O. Egiazarian, and J. T. Astola. Transform domain image restoration methods: review, comparison, and interpretation. In *Proceedings of SPIE*, volume 4304, pages 155–169, 2001.

P. Yin, A. Criminisi, J. Winn, and I. Essa. Tree-based Classifiers for Bilayer Video Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1–8, June 2007.

Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan. Active Scheduling of Organ Detection and Segmentation in Whole-Body Medical Images. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) - Part I*, pages 313–321, 2008.

Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. *IEEE Transactions on Medical Imaging*, 27(11):1668–1681, November 2008.

Y. Zheng, B. Georgescu, and D. Comaniciu. Marginal space learning for efficient detection of 2d/3d anatomical structures in medical images. In *Information Processing in Medical Imaging*, volume 5636, pages 411–422. Springer Berlin Heidelberg, 2009.

S. K. Zhou. Shape regression machine and efficient segmentation of left ventricle endocardium from 2d b-mode echocardiogram. *Medical Image Analysis*, 14(4):563 – 581, 2010.

S. K. Zhou. Discriminative anatomy detection: Classification vs regression. *Pattern Recognition Letters*, 43(0):25 – 38, 2014.

S. K. Zhou, J. Zhou, and D. Comaniciu. A boosting regression approach to medical anatomy detection. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1–8, June 2007.