Algorithme de correspondance de superpatchs multi-échelles basé sur des descripteurs duals de superpixels

Rémi Giraud, Merlin Boyer, Michaël Clément

Algorithme de correspondance de superpatchs multi-échelles basé sur des descripteurs duals de superpixels

Rémi Giraud[1] Merlin Boyer[1] Michaël Clément[2]

[1] Bordeaux INP, Univ. Bordeaux, CNRS, IMS, UMR 5218, F-33400 Talence, France
[2] Bordeaux INP, Univ. Bordeaux, CNRS, LaBRI, UMR 5800, F-33400 Talence, France

remi.giraud@ims-bordeaux.fr

Résumé

Plusieurs travaux ont proposé d'utiliser des sur-segmentations irrégulières en superpixels pour un traitement rapide et dense des images. Néanmoins, ces méthodes ont du mal à fournir des descripteurs précis, car elles ne calculent les descripteurs qu'à l'intérieur de chaque superpixel, ignorant les informations structurelles à leurs frontières. Dans ce travail, nous introduisons le superpatch dual, un descripteur contenant à la fois l'information dans les superpixels et à leurs interfaces, afin de capturer explicitement la structure du voisinage. Un algorithme rapide de correspondance non local multi-échelles est également introduit pour la recherche de descripteurs similaires à plusieurs résolutions dans une base d'images. Enfin, nous démontrons les performances de cette nouvelle approche duale sur les applications de mise en correspondance et d'étiquetage supervisé.

Mots Clef

Superpixels ; Descripteur dual de superpatch ; Correspondance non-locale multi-échelles.

Abstract

Several works have proposed to use irregular over-segmentation into superpixels for fast dense image processing. Nevertheless, they struggle to provide accurate descriptors, since they only compute features within each superpixel, ignoring structural information at their borders. In this work, we address these limitations by introducing the dual superpatch, a novel superpixel-based descriptor containing features from superpixel regions, and their interfaces to explicitly capture contour structure information. A fast multi-scale non-local matching framework is also introduced for the search of similar descriptors at several resolutions in an image dataset. Finally, we demonstrate the performances of this new dual strategy on matching and supervised labeling applications.

Keywords

Superpixels ; Superpatch dual descriptor ; Multi-scale non-local matching.

1 Introduction

Pour de nombreuses applications de vision par ordinateur, il existe un besoin important de résultats rapides et automatiques. Une des stratégies employées consiste à s'inspirer d'autres données, de façon supervisée lorsque l'on dispose d'annotations issues d'une vérité terrain. Dans ce contexte, les méthodes non locales ont permis d'obtenir des résultats précis pour de nombreuses applications. Dans ces méthodes, les régions de l'image sont considérées de manière indépendante, avec généralement un patch carré défini pour chaque pixel, capturant un motif local [4]. Pour une segmentation ou classification basée exemples, des algorithmes de correspondance peuvent alors être utilisés pour trouver des motifs similaires dans les données, et ensuite transférer l'information associée, à l'échelle du pixel, ou de l'image après un processus de décision global.
Cette recherche de motifs similaires est généralement effectuée pour chaque patch de l'image, par des algorithmes de correspondance rapide, par ex., PatchMatch [3], TreeCANN [24] ou FLANN [22], permettant d'exploiter efficacement un grand nombre d'images en des temps de calcul réduits. Pour trouver des contenus similaires, des descripteurs de patchs doivent être extraits de l'image, qui sont généralement conçus pour être robustes à diverses transformations telles que les changements d'échelle ou d'illumination. On retrouve notamment, par ex., des descripteurs comme SIFT [21], HoG [7, 8], ou BRIEF [5].
Plus récemment, les réseaux de neurones convolutifs permettent aussi d'extraire des caractéristiques pertinentes, et ont donné des résultats prometteurs dans de nombreuses applications liées au traitement d'images [19]. Cependant, bien que certaines architectures récentes peuvent apprendre à partir de jeux de données relativement réduits [26], ces méthodes s'appuient sur des stratégies d'apprentissage supervisé coûteuses, et nécessitent souvent de grands ensembles de données annotés. Dans de nombreux domaines, comme l'imagerie médicale, ces inconvénients peuvent réduire les performances de ces méthodes, en plus de limiter l'interprétabilité des résultats. Il existe donc un besoin de méthodes rapides, sans apprentissage, efficaces avec peu de données d'entraînement ou de puissance de calcul.

Dans ce contexte de correspondance rapide entre images, de de nombreux travaux ont considéré des approches de sur-segmentation en grilles régulières, par ex. [18]. D'autres méthodes ont proposé de regrouper les pixels en composantes connexes de couleurs homogènes, appelées superpixels, réduisant considérablement le nombre d'éléments à traiter tout en respectant les contours [28]. Un traitement appliqué à cette échelle est alors proche du résultat attendu à l'échelle pixellique. Plusieurs travaux ont utilisé des superpixels dans des modèles non locaux, par ex., [13, 29], ou bien de manière non supervisée en utilisant des forêts aléatoires [6, 17]. Néanmoins, l'irrégularité de forme de ces représentations [10] peut devenir un problème pour structurer l'information de voisinage, nécessaire pour calculer des correspondances pertinentes.

D'autres approches ont tenté d'utiliser le voisinage des superpixels, par ex., [25, 27]. Parmi elles, la méthode SuperPatchMatch (SPM) [9] résout partiellement ce problème avec une structure de voisinage de superpixel appelée superpatch, ainsi qu'une métrique pour comparer deux structures ayant une géométrie et un nombre d'éléments différents. Cependant, SPM reste sous-optimal en termes de complexité et de précision de correspondance. La méthode permet uniquement la recherche de superpatch à la même échelle et ne calcule des descripteurs qu'à l'intérieur de chaque région, ignorant ainsi l'information de contour. Plusieurs approches ont en effet mis en évidence le besoin de descripteurs pertinents à l'échelle des superpixels [23, 31, 30], tandis que la plupart de ceux de la littérature sont calculés localement sur un voisinage carré.

Contributions

Dans ce travail [12], nous abordons certaines des limites des méthodes non locales récentes, qui se focalisent uniquement sur les informations intra-région au sein des superpixels ainsi que leur voisinage [9]. Nous introduisons un nouveau descripteur de voisinage de superpixels appelé superpatch dual (DSP pour *dual superpatch*), contenant deux ensembles de descripteurs indépendants (voir Sec. 3).

Premièrement, les caractéristiques intra-superpixels capturent des informations de couleur ou de texture au sein de superpixels rognés, ceci afin d'éviter l'influence des pixels aux contours ou des frontières de superpixels imprécises (voir Sec. 3.1). Ensuite, pour capturer la structure du voisinage considéré, par exemple en termes d'orientations de contours, nous extrayons une grille relativement régulière de descripteurs spécifiques aux interfaces entre les superpixels (voir Sec. 3.2). Afin de comparer efficacement de tels descripteurs duals irréguliers, ayant notamment une géométrie et un nombre d'éléments différents, nous proposons également de nouvelles distances et optimisations, réduisant considérablement la complexité de calcul.

L'algorithme SuperPatchMatch (SPM) [9] est alors appliqué avec ces nouveaux descripteurs (DSPM), et permet d'obtenir des correspondances entre superpixels plus pertinentes. Par ailleurs, nous étendons également DSPM à la recherche de correspondances à plusieurs échelles (voir Sec.4), et nous proposons un *framework* permettant d'effectuer un étiquetage automatique de superpixels à partir d'une base d'images avec vérité terrain. Dans ce *framework*, la comparaison de DSP à différentes échelles peut être facilement réalisée car nous considérons des informations spatiales réduites *i.e.*, des ensembles de barycentres. Ainsi, nous pouvons rechercher des objets similaires de différentes tailles dans des jeux de données hétérogènes.
Enfin, pour montrer la robustesse de notre approche, en particulier par rapport à [9], nous présentons plusieurs expériences sur une base d'images de visages standard [15], où nous effectuons un étiquetage automatique des régions à partir d'exemples (voir Sec. 5).

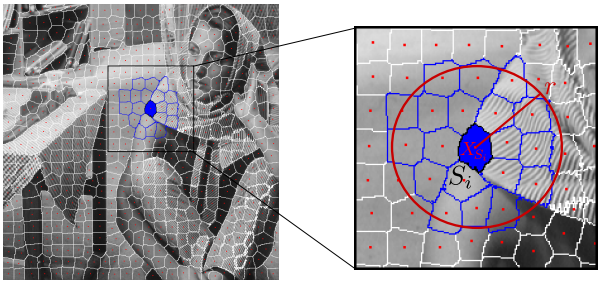


FIGURE 1 – Définition du superpatch. Pour un superpixel S_i (bleu plein), les superpixels voisins (contours bleus) avec leur barycentre (point rouge) dans un rayon r, centré sur X_{S_i} le barycentre de S_i, font partie du superpatch $\mathbf{S_i}$.

2 SuperPatchMatch

Dans cette section, nous présentons la méthode SuperPatchMatch (SPM) [9], constituant la base notre approche.

2.1 La structure de SuperPatch

Pour généraliser les méthodes standards basées patchs aux décompositions d'images irrégulières, [9] a proposé la structure de *superpatch*. Comme pour les patchs carrés définis autour de chaque pixel, un superpatch $\mathbf{S_i}$, associé à un superpixel S_i, contient les voisins d'un superpixel S_i par rapport à un rayon fixe r. La proximité est simplement calculée selon les barycentres spatiaux des superpixels :

$$\mathbf{S_i} = \{S_{i'}, \text{ tels que } ||X_{S_i} - X_{S_{i'}}||_2 \leq r\}, \qquad (1)$$

où $X_{S_i} = [x_{S_i}, y_{S_i}]$ et $X_{S_{i'}} = [x_{S_{i'}}, y_{S_{i'}}]$ représentent respectivement les barycentres spatiaux des superpixels S_i et $S_{i'}$. De cette façon, la structure de superpatch ne comprend que les superpixels voisins les plus proches, en utilisant une information spatiale réduite. En Fig. 1, nous montrons un exemple de superpatch, défini pour un superpixel S_i, contenant ses superpixels adjacents pour fournir un descripteur de superpixel intégrant le voisinage.

2.2 Distance entre SuperPatchs

Le principal problème pour concevoir une telle distance est que les deux structures sont très susceptibles d'avoir

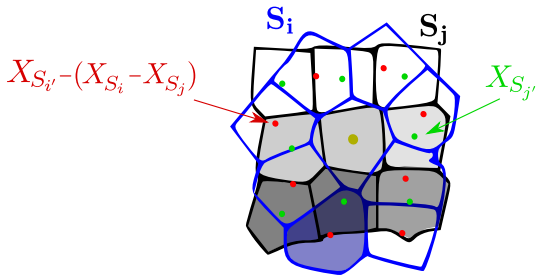


FIGURE 2 – Comparaison entre 2 superpatchs $\mathbf{S_i}$ et $\mathbf{S_j}$ (Eq. (3)). $\mathbf{S_i}$ et $\mathbf{S_j}$ sont recalés selon les barycentres X_{S_i} et X_{S_j} de leur superpixel central. Les poids w dans l'Eq. (3) favorisent la comparaison aux plus proches superpixels et la valeur de ceux correspondant au superpixel inférieur dans $\mathbf{S_i}$ (bleu plein) sont représentés dans chaque superpixel de $\mathbf{S_j}$ (une couleur sombre signifiant un poids plus fort).

une géométrie et un nombre d'éléments différents. Par conséquent, il n'y a pas de correspondance exacte entre les superpixels de deux superpatchs, contrairement aux pixels dans les patchs réguliers. Pour comparer de manière pertinente de telles structures, la spatialité doit être prise en compte, et [9] a proposé de considérer simplement la proximité des barycentres superpixels après recalage sur les superpixels centraux.

Dans ce qui suit, nous considérons deux superpatchs $\mathbf{S_i}$ et $\mathbf{S_j}$, issus par exemple de deux images A et B. Un poids $w(X_{S_{i'}}, X_{S_{j'}}) = \exp(-\|X_{S_{j'}} - (X_{S_{i'}} - (X_{S_i} - X_{S_j}))\|_2^2/\sigma_1^2)$, mesure le déplacement relatif entre les barycentres recalés $X_{S_{i'}}$ et $X_{S_{j'}}$ des superpixels $S_{i'} \in \mathbf{S_i}$ et $S_{j'} \in \mathbf{S_j}$, par rapport aux superpixels centraux S_i et S_j, et σ_1 est un paramètre d'échelle valant $1/2\sqrt{|I|/K}$, avec $|I|$ et K respectivement le nombre de pixels et de superpixels dans l'image. Ainsi un superpixel $S_{i'}$ ne se compare qu'aux plus proches dans $\mathbf{S_j}$. La distance D entre deux superpatchs $\mathbf{S_i}$ et $\mathbf{S_j}$ est finalement définie comme :

$$W(S_{i'}, S_{j'}) = w(X_{S_{i'}}, X_{S_{j'}}) w_s(X_{S_{i'}}) w_s(X_{S_{j'}}), \quad (2)$$

$$D(\mathbf{S_i}, \mathbf{S_j}) = \frac{\sum_{S_{i'} \in \mathbf{S_i}} \sum_{S_{j'} \in \mathbf{S_j}} W(S_{i'}, S_{j'}) d(F_{S_{i'}}, F_{S_{j'}})}{\sum_{S_{i'} \in \mathbf{S_i}} \sum_{S_{j'} \in \mathbf{S_j}} W(S_{i'}, S_{j'})}, \quad (3)$$

où $w_s(X_{S_{i'}})$ pondère également l'influence de $S_{i'}$ par rapport à sa distance spatiale à S_i de sorte que $w_s(X_{S_{i'}}) = \exp(-\|X_{S_{i'}} - X_{S_i}\|_2^2/(2r^2))$, et d est une distance quelconque entre les descripteurs des superpixels $F_{S_{i'}}$ et $F_{S_{j'}}$. La comparaison entre deux superpatchs ayant un nombre d'éléments et une géométrie différents est illustré en Fig. 2. Les poids w dans Eq. (3) pondèrent la distance d entre les descripteurs F entre un superpixel dans $\mathbf{S_i}$ et un superpixel dans $\mathbf{S_j}$ après recalage sur leurs barycentres centraux. Ceux correspondant au superpixel inférieur dans $\mathbf{S_i}$ sont représentés dans chaque superpixel de $\mathbf{S_j}$.

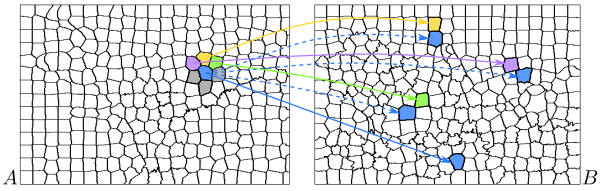


FIGURE 3 – Principe de l'algorithme SPM. Les lignes pleines correspondent aux meilleures correspondances. Les voisins adjacents du superpixel bleu sont considérés pour proposer de nouveaux candidats (lignes pointillées). Les orientations entre superpixels dans A tentent d'être respectées dans B, par ex., le superpixel jaune reste au dessus du bleu dans A et B. Les voisins restants (gris) qui n'ont pas encore été traités à cette itération seront considérés à la suivante, dont le sens de parcours sera inversé.

2.3 Algorithme de correspondance SuperPatchMatch

Les méthodes non locales ont rapidement mis en évidence le besoin d'algorithmes de recherche rapides basés patchs pour calculer des correspondances dans de grandes zones, par ex., dans des bibliothèques d'images d'exemples. Une avancée significative a été obtenue avec PatchMatch (PM) [3], un algorithme de correspondance rapide, en partie aléatoire, fournissant pour chaque patch d'une image A, une correspondance dans une image B. PM a des propriétés très intéressantes : il ne nécessite aucune étape d'apprentissage et sa complexité ne dépend que de la taille de l'image à traiter A, permettant de rechercher des correspondances dans un grand nombre d'images.

L'algorithme PM est initialisé à partir de correspondances aléatoires et les raffine itérativement avec un traitement des patchs séquentiel. Ce processus est principalement basé sur la propagation rapide de bonnes correspondances à partir des voisins adjacents. De larges régions sont en effet très susceptibles de se correspondre entre les images. Selon le sens de parcours, qui est inversé à chaque itération, les correspondances décalées de deux patchs spatialement adjacents sont considérées comme de nouveaux candidats. De plus, des patchs aléatoires sont testés près de la meilleure correspondance actuelle en B. Notons que PM étant partiellement aléatoire, des exécutions parallèles peuvent fournir facilement différentes correspondances.

La méthode SuperPatchMatch (SPM) généralise PM aux superpatchs [9], pour fournir un algorithme de correspondance rapide de superpixels. La méthode nécessite principalement d'adapter la propagation des correspondances depuis les voisins car il n'y a plus de géométrie régulière entre les éléments adjacents, contrairement au cas standard de la grille de pixels. L'étape de propagation de SPM est illustrée dans la Fig. 3. Les voisins adjacents sont considérés comme conduisant à de nouvelles correspondances tout en respectant l'orientation relative entre les superpixels en A et B, pour favoriser la correspondance de régions plus grandes.

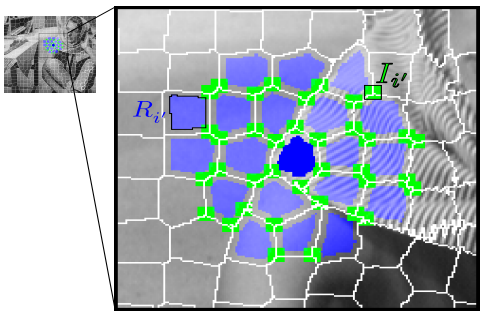


FIGURE 4 – Descripteur dual de superpatch (DSP). L'information intra-region ($R_{i'}$) dans chaque superpixel $S_{i'}$ avec un décalage de $\beta = 3$ pixels depuis la frontière (régions bleues) est considérée avec l'information des interfaces entre superpixels $I_{i'}$ (carrés verts) dans le même rayon r.

Limitations

La distance SPM par défaut de l'Eq. (3) a une complexité quadratique : chaque superpixel d'un superpatch est comparé à tous ceux de l'autre superpatch, ce qui peut entraîner un temps de calcul important. En outre, SPM ne prend en compte que des descripteurs intra-superpixels. Le superpatch ne capture donc pas précisément l'information de contours ou de gradient entre les régions, qui se trouve généralement aux frontières des superpixels, et qui peut être partagée entre deux régions. Enfin, SPM ne prend pas en compte d'aspect multi-échelles qui permettrait de capturer des objets similaires mais de tailles différentes.
Dans les sections suivantes, nous répondons à tous ces problèmes avec la méthode proposée de correspondance de superpatchs multi-échelles, qui utilise de nouveaux descripteurs duals basés superpixels.

3 Descripteurs duals de superpatchs

Dans cette section, nous proposons une approche pour extraire de manière pertinente les descripteurs dans un voisinage de superpixels. Nous introduisons un descripteur dual qui permet de capturer efficacement, autour d'un superpixel, à la fois le contenu de la région et l'information de structure des contours, généralement située à leurs frontières. Des descripteurs sont donc calculés à l'intérieur de chaque région et aux interfaces entre les superpixels adjacents. Ce descripteur dual est appelé *Dual SuperPatch* (DSP), est noté $\bar{\mathbf{S}}_\mathbf{i}$ pour un superpixel S_i, et est représenté dans la Fig. 4 sur le même exemple de décomposition utilisé dans la Fig. 1. Dans ce qui suit, nous présentons l'approche d'extraction des descripteurs de région (R), *i.e.*, intra-superpixel et d'interfaces (I), et nous proposons une méthode générale pour comparer différents DSP.

3.1 Descripteurs intra-superpixel

La formulation de superpatch de [9] considère des descripteurs calculés sur l'entièreté de chaque région de superpixel dans le voisinage. Cependant, les superpixels ont tendance à capturer des régions homogènes. Des pixels situés sur des contours fins ou flous peuvent donc être arbitrairement associés aux superpixels, ce qui peut perturber le calcul des descripteurs. Définir un bloc régulier ou une pondération spatiale depuis le barycentre du superpixel ne permettraient pas d'extraire de manière pertinente l'information lorsque les superpixels ont des formes très irrégulières [23].
Dans ce travail, nous proposons de considérer l'information intra-superpixel avec un décalage de β pixels à ses bordures. De cette façon, nous prenons en compte la quasi-totalité de la région, tout en étant robuste aux frontières imprécises ou aux contours fins, qui seront pris en compte dans un autre descripteur dédié aux interfaces au sein de notre DSP (voir Sec. 3.2). Dans la Fig. 4, les régions considérées $R_{i'}$ pour les superpixels $S_{i'}$ sont représentées en bleu. Pour chaque région $R_{i'}$, les informations $F_{R_{i'}}$ et spatiales $X_{R_{i'}}$ sont prises en compte, donc le superpatch dual contient un ensemble de couples $\mathbf{R_i} = \{F_{R_{i'}}, X_{R_{i'}}\}$. Pour démontrer le problème de considérer toute la région du superpixel pour extraire des descripteurs, nous considérons deux décompositions d'images contenant des régions avec 16 textures orientées différentes (voir la Fig. 5). Le rayon de superpatch est défini à $r = 0$ pour ne considérer que les informations d'un seul superpixel, où des descripteurs HoG [7] sont calculés. Pour chaque superpixel de l'image de gauche, nous calculons de manière exhaustive sa plus proche correspondance dans l'image de droite. Nous rapportons dans la Tab. 1 la précision de correspondance moyenne sur tous les superpixels de l'image de gauche, selon différentes valeurs β, et pour plusieurs niveaux de bruit Gaussien appliqués aux deux images après décomposition. Cette évaluation est faite sur des décompositions calculées avec [11], et sur les décompositions vérité terrain, capturant parfaitement les changements de texture. Cette expérience met en évidence la nécessité de restreindre la zone pour extraire des informations de superpixels car les décompositions de superpixels peuvent ne pas être parfaitement précises. De plus, même sur des décompositions parfaites, les informations de gradient inexactes situées sur les bordures de superpixels sont prises en compte et peuvent dégrader les résultats, ce qui peut être évité si $\beta > 0$.

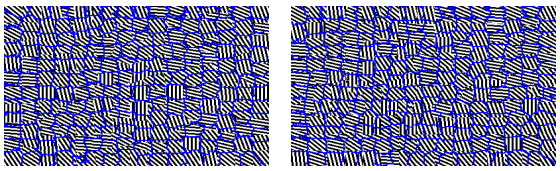
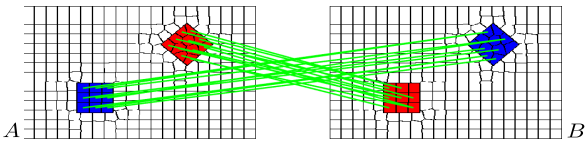
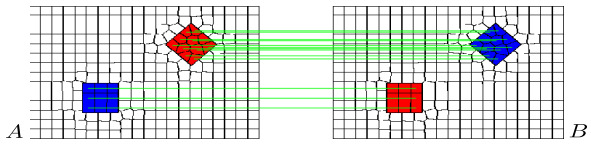


FIGURE 5 – Deux images synthétiques contenant 16 textures orientées et décomposées en superpixels avec [11].

Distance de comparaison rapide. La comparaison entre deux ensembles de descripteurs de région peut être effectuée d'une manière plus efficace en termes de complexité qu'avec l'Eq. (3). Nous proposons de ne sélectionner qu'un seul superpixel $S_{j'} \in \mathbf{S_j}$ à comparer pour chaque superpixel $S_{i'} \in \mathbf{S_i}$. Pour ce faire, le bary-



(a) Correspondance avec l'information couleur intra-superpixels ($\alpha = 1$)



(b) Correspondance avec l'information contour aux interfaces des superpixels ($\alpha = 0$)

FIGURE 6 – Impact du paramètre α sur la distance DSP (6). (a) Seuls les descripteurs de région $R_{i'}$ sont utilisés ($\alpha = 1$), avec des informations de couleur moyennes. (b) Seuls les descripteurs d'interface HoG [7] $I_{i'}$ sont utilisés ($\alpha = 0$) et permettent de capturer les informations de structure. Le rayon r est défini de sorte à capturer le premier anneau d'adjacence.

TABLE 1 – Influence du décalage β par rapport aux frontières des superpixels pour le descripteur intra-region. Les résultats de correspondance de texture entre les images de la Fig. 5 sont donnés pour différents β et variances de bruit Gaussien ajouté aux images, pour des décompositions calculées par [11] et vérité terrain. Les meilleurs et seconds résultats sont respectivement en gras et soulignés.

	Décompositions en superpixels				Décompositions vérité terrain			
Variance	0	50	100	125	0	50	100	125
$\beta = 0$	0.526	0.634	0.516	0.366	**1.000**	**1.000**	0.980	0.907
$\beta = 1$	0.616	0.654	**0.558**	0.419	**1.000**	**1.000**	**0.993**	**0.913**
$\beta = 2$	0.679	0.665	0.558	**0.482**	**1.000**	**1.000**	0.987	0.913
$\beta = 3$	**0.711**	**0.675**	0.521	0.482	**1.000**	**1.000**	0.953	0.900

centre $X_{S_{i'}}$ est d'abord recalé par le déplacement entre les superpixels centraux S_i et S_j, et nous notons cette nouvelle position $X_{S^j(i')}$, calculée telle que $X_{S^j(i')} = X_{S_{i'}} - (X_{S_i} - X_{S_j})$. Dans la Fig. 2, ceux-ci correspondent aux barycentres de superpixels rouges. Ensuite, nous projetons les barycentres recalés sur la décomposition de l'image d'où $\mathbf{S_j}$ est extrait. Ainsi, dans la Fig. 2, le superpixel noir contenant un point rouge $X_{S^j(i')}$ serait sélectionné pour être comparé au superpixel $S_{i'}$ de région $R_{i'}$. Ce superpixel correspondant contenant $X_{S^j(i')}$ dans l'image comparée, est noté $S^j(i')$, et son intra-région associée est notée $R^j(i')$. De cette façon, nous réduisons considérablement la complexité de la distance, tout en augmentant potentiellement la précision de la comparaison (voir Sec. 5). La comparaison entre deux descripteurs de région $\mathbf{R_i}$ et $\mathbf{R_j}$ est définie par des projections barycentriques telle que :

$$d_p(\mathbf{R_i}, \mathbf{R_j}) = \frac{\sum_{R_{i'} \in \mathbf{R_i}} w_s\left(X_{R_{i'}}\right) d\left(F_{R_{i'}}, F_{R^j(i')}\right)}{\sum_{R_{i'} \in \mathbf{R_i}} w_s\left(X_{R_{i'}}\right)}. \quad (4)$$

Notons que les barycentres recalés en dehors des limites de l'image sont projetés pour sélectionner le superpixel le plus proche à la frontière de l'image.

Une distance projetée similaire a été suggérée dans [9], mais avec une formulation non symétrique. Dans notre modèle de comparaison de superpatch dual, nous considérons une distance projetée symétrique D_p sur les descripteurs intra-région définie comme :

$$D_p(\mathbf{R_i}, \mathbf{R_j}) = \frac{1}{2}\left(d_p\left(\mathbf{R_i}, \mathbf{R_j}\right) + d_p\left(\mathbf{R_j}, \mathbf{R_i}\right)\right). \quad (5)$$

3.2 Descripteurs inter-superpixels

Pour capturer efficacement l'information de structure et de contours dans l'image, nous proposons de considérer également des descripteurs spécifiques aux interfaces entre superpixels. De telles interfaces peuvent être facilement extraites en considérant les points où au moins trois superpixels sont présents dans un voisinage de 3×3 pixels. De cette façon, nous obtenons directement une grille relativement régulière de points d'intérêt potentiels en termes de contours, sans introduire d'autres paramètres. Dans la Fig. 4, ces régions d'interface notées $I_{i'}$ sont représentées par des carrés verts. Sur ces régions, des descripteurs de contour spécifiques peuvent être calculés, par ex., HoG [7].

Accélération de la distance quadratique. Étant donné que les interfaces ne fournissent pas une décomposition dense du domaine de l'image, la distance Eq. (4) à l'aide de projections ne peut pas être utilisée pour comparer rapidement deux ensembles de descripteurs d'interface. Une distance quadratique un-à-plusieurs telle que Eq. (3) pourrait être utilisée, mais serait très coûteuse. Pour résoudre ce problème, nous proposons une association un-à-un pour chaque descripteur d'interface $I_{i'}$. Chaque $I_{i'}$ n'est comparé qu'à celui $I_{j'}$ spatialement le plus proche dans l'autre superpatch double. Ainsi la méthode ne calcule exhaustivement que les distances spatiales entre les points d'interface. La distance est calculée comme pour l'Eq. (4), où $I^j(i')$, le descripteur d'interface sélectionné dans $\mathbf{I_j}$ pour $I_{i'}$ est défini comme $I^j(i') = \underset{I_{j'}}{\operatorname{argmin}} \|X_{I_{i'}} - X_{I_{j'}}\|_2$. Enfin, comme pour l'Eq. (5), la distance est également calculée de $\mathbf{I_j}$ à $\mathbf{I_i}$ pour obtenir une distance symétrique.

3.3 Modèle général de comparaison de superpatchs duals

Notre superpatch dual (DSP) $\bar{\mathbf{S}}_\mathbf{i}$, pour un superpixel S_i, est donc décrit par un ensemble de régions intra-superpixels $\mathbf{R_i} = \{F_{R_{i'}}, X_{R_{i'}}\}$ et d'interfaces entre superpixels $\mathbf{I_i} = \{F_{I_{i'}}, X_{I_{i'}}\}$ tel que $\bar{\mathbf{S}}_\mathbf{i} = [\mathbf{R_i}, \mathbf{I_i}]$. Pour mesurer de manière pertinente la similitude de deux DSP $\bar{\mathbf{S}}_\mathbf{i}$ and $\bar{\mathbf{S}}_\mathbf{j}$ ayant une géométrie et un nombre d'éléments différents, nous proposons la distance de comparaison DSP générale suivante :

$$D(\bar{\mathbf{S}}_\mathbf{i}, \bar{\mathbf{S}}_\mathbf{j}) = \alpha\, D_p(\mathbf{R_i}, \mathbf{R_j}) + (1-\alpha)\, D_p(\mathbf{I_i}, \mathbf{I_j}), \quad (6)$$

avec D_p la distance rapide sur les descripteurs, en utilisant des projections de barycentre de l'Eq. (5) pour l'intra-région R, et la sélection du descripteur le plus proche pour

les interfaces I, et $\alpha \in [0,1]$ un paramètre de compromis, qui peut être réglé de manière intuitive en utilisant des descripteurs identiques ou normalisés pour R et I.
Dans la Fig. 6, nous montrons les résultats obtenus avec notre modèle utilisant la couleur moyenne comme descripteur intra-région $F_{R_{i'}}$ (Fig. 6 (a), $\alpha = 1$) et des descripteurs HoG pour les interfaces $F_{I_{i'}}$ (Fig. 6 (b), $\alpha = 0$). Nous mettons donc en évidence l'aspect général de notre approche qui permet de se concentrer soit sur l'information intra-régions (Fig. 6 (a)), soit sur l'information inter-régions (Fig. 6 (b)). Dans la Sec. 5, nous détaillons les performances obtenues à l'aide de ces descripteurs complémentaires.

4 DSPM multi-échelles

Dans cette section, nous étendons la méthode SPM avec notre descripteur dual (DSPM), pour effectuer la recherche de DSP à plusieurs échelles. Nous montrons d'abord comment comparer deux DSP de tailles différentes, puis nous proposons une stratégie de fusion multi-échelles.

4.1 Mise à l'échelle de DSP

Dans la Sec. 3, nous avons montré comment comparer deux DSP extraits avec la même taille de rayon r. Néanmoins, la distance proposée Eq. (6) peut facilement s'adapter à des DSP de différentes tailles, car l'information spatiale est concentrée dans les barycentres X. Nous considérons deux DSP $\bar{\mathbf{S}}_\mathbf{i}$ et $\bar{\mathbf{S}}_\mathbf{j}$, avec différents rayons d'extraction de voisinage r^i et r^j dans l'Eq. (1). Pour les comparer, toutes les informations spatiales contenues dans $\bar{\mathbf{S}}_\mathbf{j}$ peuvent être ajustées en fonction du rapport entre les rayons, tel que :

$$\bar{\mathbf{S}}_\mathbf{j} = \left[\left\{ \left(F_{R_{j'}}, X_{R_{j'}} \frac{r^i}{r^j} \right) \right\}, \left\{ \left(F_{I_{j'}}, X_{I_{j'}} \frac{r^i}{r^j} \right) \right\} \right]. \quad (7)$$

Ainsi, un DSP similaire peut être trouvé à différentes échelles, *par ex.*, dans une large base d'images d'exemples. Notons que les descripteurs $F_{R_{j'}}$ et $F_{I_{j'}}$ restent inchangés par cette transformation de mise à l'échelle.

4.2 Méthode multi-échelles basée exemples

La plupart des méthodes non locales effectuent une recherche de contenu similaire dans un ensemble de données hétérogènes, sans information préalable sur la taille de l'objet ciblé. Dans [9], aucune stratégie multi-échelles n'est proposée car l'expérience d'étiquetage basée exemples est effectuée sur des images déjà recalées linéairement [15].
Ici, nous introduisons une approche multi-échelles générale, basée exemples, qui permet de rechercher des DSP similaires de différents rayons, afin de capturer des objets de différentes tailles. La structure DSP proposée permet en effet d'effectuer une simple mise à l'échelle automatique, présentée dans la Sec. 4.1. Par conséquent, plusieurs tailles de DSP peuvent être considérées dans un ensemble d'images d'exemples $\mathbf{B}$. Un ensemble $\mathbf{r^B} = \{r^B\}$ est considéré pour définir le rayon de DSP dans $\mathbf{B}$.

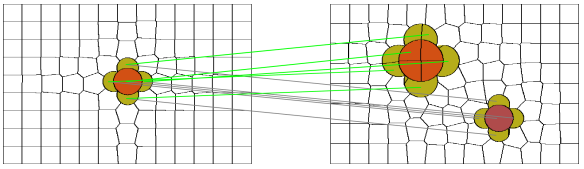


FIGURE 7 – Exemple de résultat de correspondance sans (lignes grises) et avec (lignes vertes) une recherche de correspondance multi-échelles. Plus de détails dans le texte.

Dans [9], une approche d'étiquetage supervisé basé sur l'algorithme des moyennes non locales [4] a été introduit pour fusionner les informations issues de plusieurs correspondances de superpatch calculées dans une bibliothèque d'images d'exemples pour une image A à traiter. Une carte d'étiquettes $L_m^{r^B}(S_i)$ est calculée pour un superpixel S_i, pour toutes les M étiquettes différentes m, telle que :

$$L_m^{r^B}(S_i) = \frac{1}{W} \sum_{S_j \in \mathbf{B}_i^{m,r^B}} \omega(S_i, S_j), \quad (8)$$

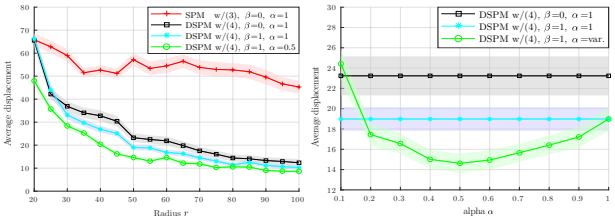
où $\mathbf{B}_i^{m,r^B}$ est l'ensemble des correspondances de S_i calculées à l'échelle r^B et ayant une étiquette de vérité terrain m, et W est le facteur de normalisation $W = \sum_{m=1}^{M} \sum_{S_j \in \mathbf{B}_i^{m,r^B}} \omega(S_i, S_j)$, avec ω un poids dépendant de la similarité des DSP [9]. L'étiquette finale d'un superpixel S_i est calculée par $\mathcal{L}(S_i) = \underset{m \in \{1,\dots,M\}}{\operatorname{argmax}} \left(\underset{r^B \in \mathbf{r^B}}{\operatorname{argmax}} \left(L_m^{r^B}(S_i) \right) \right)$.
Dans la Fig. 7, nous représentons les résultats de correspondance obtenus sans (lignes grises) et avec (lignes vertes) notre stratégie multi-échelles. Nous pouvons voir que la meilleure correspondance entre les recherches aux échelles $\mathbf{r^B} = [0.5, 1, 2, 4] \times r^A$ permet de capturer la plus grande fleur avec des couleurs similaires, au lieu de celle à la même échelle.

5 Validations expérimentales

Dans cette section, nous présentons plusieurs expériences quantitatives afin de démontrer l'intérêt de notre *framework* DPSM. Nous validons d'abord le comportement de notre modèle sur des images standards par rapport aux paramètres de la méthode. Ensuite, nous proposons des expériences de segmentation et d'étiquetage à plus grande échelle sur une base d'images de visages standard.

5.1 Réglage des paramètres

La méthode proposée a été implémentée en MATLAB et en utilisant du code C-MEX sur un ordinateur Linux standard avec un processeur 4 coeurs à $1,90$ GHz et 16 Go de RAM. Le nombre d'itérations de DSPM est fixé à 5, et nous utilisons la norme ℓ_2 pour calculer la distance d entre les caractéristiques F comme dans [9]. Pour éviter des sur-détections, les interfaces entre régions sont détectées au plus uniquement tous les 4 pixels.



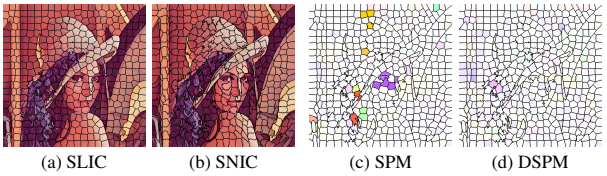
(a) (b)

FIGURE 8 – Influence du rayon r (a) et du coefficient de compromis α (b) pour la méthode DSPM proposée, par rapport à SPM. Nous rapportons la distance moyenne entre chaque barycentre de superpixel et celui de sa correspondance la plus proche dans une autre décomposition de la même image. Voir le texte pour plus de détails.

Par ailleurs, par défaut nous avons fixé le rognage des régions à $\beta = 1$; le paramètre de compromis entre les distances intra-région et d'interface dans Eq. (6) a été fixé à $\alpha = 0,5$; et le rayon d'un superpatch à $r = 50$. Pour R et I, différents types de descripteurs sont utilisés, en fonction de l'application considérée.

5.2 Influence des paramètres

Pour démontrer l'intérêt de chaque contribution, nous considérons une première expérience de mise en correspondance sur des images standard *Baboon, Barbara, House, Lena* et *Peppers*. Chaque image est décomposé avec deux méthodes de segmentation en superpixels : SLIC [1] et SNIC [2]. Pour chaque superpixel dans une décomposition donnée, nous calculons la correspondance DSP la plus proche dans l'autre. Un descripteur robuste devrait en effet être robuste aux variations entre différentes segmentations pour une même image. Pour les caractéristiques intra-régions $F_{R_{I'}}$, nous calculons l'histogramme RVB cumulé normalisé avec 9 classes par canal. Pour les interface entre les régions $F_{I_{I'}}$, nous calculons des descripteurs HoG [7] sur une fenêtre locale de 9×9 pixels. Nous évaluons la précision de l'appariement en fonction de la distance moyenne entre les barycentres des superpixels et ceux de leur correspondance dans l'autre décomposition. Dans les Fig. 8 (a) et Fig. 8 (b), nous rapportons respectivement la distance moyenne par rapport au paramètre de rayon r et par rapport au paramètre α pour $r = 50$. D'une part, la Fig. 8 (a) montre que la précision augmente logiquement avec le rayon de superpatch, et avec chacune de nos contributions : en utilisant la distance projetée symétrique (Eq. (5)), en appliquant le rognage des frontières ($\beta > 0$), et en utilisant les descripteurs d'interface. D'autre part, la Fig. 8 (b) illustre l'intérêt d'utiliser la combinaison des descripteurs d'interface (Eq. (6)) avec les régions recadrées lorsque $\alpha > 0,2$. En particulier, un compromis équilibré avec $\alpha \approx 0,5$ offre la meilleure précision de correspondance pour ces images. Dans la Fig. 9, nous montrons également un exemple de résultat de correspondance pour DSPM avec des paramètres par défaut par rapport à SPM.



(a) SLIC (b) SNIC (c) SPM (d) DSPM

FIGURE 9 – Correspondance de DSPM par rapport à SPM. (a) et (b) : décompositions avec SLIC et SNIC. (c) et (d) : résultats de correspondance pour $r = 50$. Les distances entre les barycentres des superpixels sont représentées à l'aide du code couleur de flot optique standard (des couleurs fortes représentes des déplacements plus grands).

TABLE 2 – Précision de l'étiquetage pour l'expérience multi-échelle. Les images d'entraînement ont été soit sous-échantillonnées, soit sur-échantillonnées par un facteur de 1.5 ou 2 et r^A est réglé sur 50. Les colonnes argmax correspondent à la stratégie de fusion proposée dans Sec. 4.2 pour différentes échelles combinées.

Rayon r^B	50^*	25^*	33^*	75^{*+}	100^{*+}	argmax*	argmax$^+$
sans (7)	94.08%	94.05%	93.93%	94.11%	94.07%	94.25%	94.16%
avec (7)	94.08%	94.22%	94.13%	94.87%	94.80%	94.78%	**94.96%**

5.3 Résultats de segmentation et étiquetage

Validation. Nous évaluons l'approche DSPM proposée sur la même expérience d'étiquetage de visages présentée dans [9]. Le jeu de données LFW (*Labeled Faces in the Wild*) [15] considéré contient 1500 images d'entraînement et 927 images de test de 250×250 pixels, recalées linéairement [14], et déjà segmentées en environ 250 superpixels. LFW étant fourni avec une vérité terrain des différentes régions et des étiquettes associées, les comparaisons entre différentes méthodes ne dépendent donc pas de l'algorithme de segmentation initial. Par ailleurs, afin d'assurer des comparaisons équitables avec [9], nous utilisons la même implémentation HoG sur une grille régulière [8], et nous calculons 50 correspondances de DSP par 50 processus DSPM indépendants pour chaque superpixel, qui sont ensuite fusionnés à partir de l'Eq. (8).
Validation multi-échelles. Dans cette expérience, l'objectif est de valider l'intérêt de notre stratégie d'appariement multi-échelles présentée dans la Sec. 4. Ainsi, nous avons appliqué manuellement des changement d'échelle aléatoires aux images d'entraînement initialement recalées. Chaque image et sa décomposition ont été soit sous-échantillonnées, soit sur-échantillonnées aléatoirement par un facteur de $1,5$ ou 2 (sans interpolation). Les visages représentés sur les images peuvent donc apparaître jusqu'à deux fois plus grands ou plus petits par rapport à leur échelle initiale. Suite à ces transformations, la base d'exemples contient des motifs de visages à différentes échelles, qui ne seraient pas capturés en utilisant le même rayon de DSP pour A et B. Nous appliquons alors DSPM avec $r^A = 50$ et $r^B \in \{25, 33, 50, 75, 100\}$.

TABLE 3 – Précision d'étiquetage sur la base LFW.

Méthode	Précision (superpixels)	Précision (pixels)
Spatial CRF [16]	93.95%	*non rapporté*
CRBM [16]	94.10%	*non rapporté*
GLOC [16]	94.95%	*non rapporté*
DCNN [20]	*non rapporté*	95.24%
SPM [9] avec (3)	91.88%	92.21%
SPM [9] avec (4)	95.08%	95.43%
DSPM	**95.24%**	**95.59%**

La Tab. 2, rapporte la précision d'étiquetage pour chaque rayon et pour la fusion d'étiquettes multi-échelles proposée dans la Sec. 4.2. D'après ces résultats, lorsque $r^A \neq r^B$, nous pouvons observer que les performances pour des rayons plus petits ou plus grands sont toujours meilleures après avoir appliqué la stratégie de mise à l'échelle, *i.e.*, avec l'Eq. (7). La fusion multi-échelles faisant la moyenne du résultat sur les différentes tailles de rayon fonctionne mieux que la valeur par défaut $r^B = 50$ et que la plupart des échelles simples. De plus, en considérant uniquement des échelles plus grandes, *i.e.*, $r^B = 75, 100$, donc des comparaisons de DSP plus précises, nous obtenons la meilleure précision, démontrant la capacité de DSPM à fusionner des informations issues de plusieurs échelles.

Comparaison aux méthodes de l'état de l'art. Dans la Tab. 3, nous comparons également les performances de la méthode DSPM proposée avec les résultats des méthodes de l'état de l'art, principalement basées sur des approches d'apprentissage profond supervisé. Dans [16], plusieurs approches sont utilisées pour étiqueter l'ensemble de données LFW, basées sur les champs aléatoires conditionnels (CRF) et les machines de Boltzmann conditionnelles restreintes (CRBM). La méthode GLOC (GLObal et LOCal) [16] est également proposée pour utiliser conjointement les approches CRF et CRBM pour introduire des *a priori* de forme globale dans le processus d'entraînement. Enfin, dans [20], un réseau de neurones convolutif profond (DCNN) est proposé et dédié à cette application d'étiquetage de visages. Les résultats du tableau 3 sont ceux rapportés par les auteurs.

Pour SPM, les résultats correspondent à l'approche initiale utilisant des comparaisons quadratiques coûteuses avec l'Eq. (3), et les résultats rapportés par les auteurs utilisant des distances projetées non symétriques avec l'Eq. (4). DSPM donne la meilleure précision d'étiquetage à l'échelle superpixellique et pixellique (où l'on tient compte de la taille de chaque superpixel).

Des exemples d'étiquetage sont également représentés dans la Fig. 10. L'approche DSP proposée permet de capturer de manière pertinente le contexte d'un superpixel en termes de texture et de structure. De plus, sans optimisations particulières du code, et avec un environnement non entièrement multithreadé, DSPM s'applique en moins de 3s par sujet, contre 45s pour SPM en utilisant l'Eq. (3).

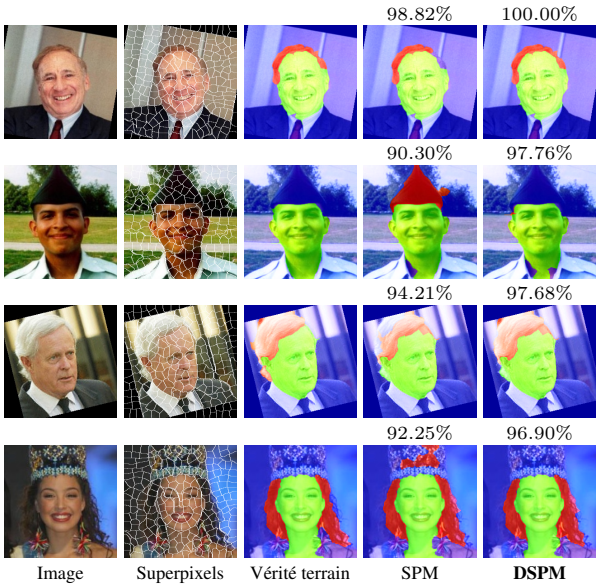


FIGURE 10 – Exemples et précision à l'échelle superpixellique de résultats d'étiquetage obtenus avec DSPM sur la base LFW, et comparés à la méthode initiale SPM.

Notre méthode est donc particulièrement intéressante du fait de sa simplicité d'utilisation, de réglage des paramètres, et d'interprétabilité par rapport aux approches basées sur l'apprentissage, tout en fournissant des résultats plus précis que SPM. Enfin, n'importe quel descripteur peut être directement utilisé dans la méthode, *par ex.*, [31, 30], éventuellement basés sur des architectures d'apprentissage profond préalablement entraînés.

6 Conclusion

Dans ce travail, nous avons répondu aux principales limites des méthodes de correspondance de superpixels existantes, en termes de robustesse et de complexité de calcul. Nous avons introduit le superpatch dual, un nouveau descripteur de voisinage superpixel contenant à la fois des informations intra-région et d'interface, respectivement robuste à l'imprécision des frontières des superpixels et capturant les structures des contours. Nous avons également proposé des distances optimisées et une approche multi-échelles pour rechercher des superpatchs duals similaires dans un ensemble d'images avec des objets de différentes tailles. Nos validations ont montré une amélioration de la précision de la méthode sur l'appariement et les applications d'étiquetage basées exemples. La méthode duale proposée devrait bénéficier aux approches non locales basées superpixels, et les travaux futurs étudieront son application à des ensembles de données hétérogènes en vision par ordinateur ou imagerie médicale.

Remerciements

Ce travail a bénéficié d'un support financier de l'état Français, via l'Agence Nationale de la Recherche (ANR) dans le cadre du projet GOTMI (ANR-16-CE33-0010-01).

Références

[1] R. Achanta et al., SLIC superpixels compared to state-of-the-art superpixel methods, *PAMI*, 2012.

[2] R. Achanta et al., Superpixels and polygons using simple non-iterative clustering, *CVPR*, 2017.

[3] C. Barnes et al., PatchMatch : A randomized correspondence algorithm for structural image editing, *ACM ToG*, 2009.

[4] A. Buades et al., A non-local algorithm for image denoising, *CVPR*, 2005.

[5] M. Calonder et al., Brief : Binary robust independent elementary features, *ECCV*, 2010.

[6] P.-H. Conze et al., Hierarchical multi-scale supervoxel matching using random forests for automatic semi-dense abdominal image registration, *ISBI*, 2017.

[7] N. Dalal et al., Histograms of oriented gradients for human detection, *CVPR*, 2005.

[8] P. Felzenszwalb et al., Object detection with discriminatively trained part based models, *PAMI*, 2010.

[9] R. Giraud et al., SuperPatchMatch : An algorithm for robust correspondences using superpixel patches, *TIP*, 2017.

[10] R. Giraud et al., Evaluation framework of superpixel methods with a global regularity measure, *JEI*, 2017.

[11] R. Giraud et al., Texture Superpixel Clustering from Patch-based Nearest Neighbor Matching, *EUSIPCO*, 2019.

[12] R. Giraud et al., Multi-Scale Superpatch Matching using Dual Superpixel Descriptors, *PRL*, 2020.

[13] S. Gould et al., Multi-class segmentation with relative location prior, *IJCV*, 2008.

[14] G. Huang et al., Unsupervised joint alignment of complex images, *ICCV*, 2007.

[15] G. Huang et al., Labeled faces in the wild : A database for studying face recognition in unconstrained environments, *Technical Report 07-49, Univ. of Massachusetts, Amherst*, 2007.

[16] A. Kae et al., Augmenting CRFs with Boltzmann machine shape priors for image labeling, *CVPR*, 2013.

[17] F. Kanavati et al., Supervoxel classification forests for estimating pairwise image correspondences, *PR*, 2017.

[18] S. Lazebnik et al., Beyond bags of features : Spatial pyramid matching for recognizing natural scene categories, *CVPR*, 2006.

[19] Y. LeCun et al., Deep Learning, *Nature*, 2015.

[20] S. Liu et al., Multi-objective convolutional learning for face labeling, *CVPR*, 2015.

[21] D. Lowe et al., Distinctive image features form scale-invariant keypoints, *IJCV*, 2004.

[22] M. Muja et al., Scalable Nearest Neighbor Algorithms for High Dimensional Data, *PAMI*, 2014.

[23] P. Neubert et al., Benchmarking superpixel descriptors, *EUSIPCO*, 2015.

[24] I. Olonetsky et al., TreeCANN-kd tree coherence approximate nearest neighbor algorithm, *ECCV*, 2012.

[25] S.-C. Pei et al., Saliency detection using superpixel belief propagation, *ICIP*, 2014.

[26] O. Ronneberger et al., U-net : Convolutional networks for biomedical image segmentation, *MICCAI*, 2015.

[27] R. Sawhney et al., GASP : Geometric association with surface patches, *3DV*, 2014.

[28] D. Stutz et al., Superpixels : An evaluation of the state-of-the-art, *CVIU*, 2018.

[29] J. Tighe et al., SuperParsing : Scalable nonparametric image parsing with superpixels, *ECCV*, 2010.

[30] F. Tilquin et al., Robust Supervoxel Matching Combining Mid-Level Spectral and Context-Rich Features, *Patch-MI*, 2018.

[31] Y. Zhang et al., Consistent Correspondence of Cone-Beam CT Images Using Volume Functional Maps *MICCAI*, 2018.