



**HAL**  
open science

## Découverte automatique du recouvrement en recalage direct d'images

Florent Brunet, Adrien Bartoli, Nassir Navab, Rémy Malgouyres

► **To cite this version:**

Florent Brunet, Adrien Bartoli, Nassir Navab, Rémy Malgouyres. Découverte automatique du recouvrement en recalage direct d'images. RFIA'10 - Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle, Jan 2010, Caen, France. pp.1-8. hal-00609669

**HAL Id: hal-00609669**

**<https://hal.science/hal-00609669>**

Submitted on 19 Jul 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Découverte automatique du recouvrement en recalage direct d'images

F. Brunet<sup>1,2</sup>

A. Bartoli<sup>1</sup>

N. Navab<sup>2</sup>

R. Malgouyres<sup>1</sup>

<sup>1</sup> Université d'Auvergne, Clermont-Ferrand, France

<sup>2</sup> CAMP, Technische Universität München, Munich, Allemagne

florent@florentbrunet.com

## Résumé

Le recalage direct d'images consiste à estimer la transformation géométrique (et parfois photométrique) reliant une image source et une image cible. Les méthodes classiques maximisent la ressemblance photométrique des pixels d'une région d'intérêt. Cette dernière, typiquement déterminée par l'utilisateur, est une partie de l'image source nécessairement visible dans l'image cible. Elle devrait idéalement correspondre à la zone de recouvrement entre les deux images, ce qui est rarement le cas car ce recouvrement n'est en général connu qu'après recalage.

Nous proposons une nouvelle approche permettant de découvrir la zone de recouvrement tout en recalant les images. Elle repose sur une vision nouvelle du recalage : nous proposons de maximiser le nombre de pixels de toute l'image source visibles dans l'image cible et photométriquement compatibles avec leur correspondant dans l'image cible. Notre approche est robuste, en ce sens qu'elle gère les pixels corrompus, dus entre autre aux occultations.

Testée avec des homographies et des B-splines comme modèles de déformation, notre approche dépasse les approches classiques en termes de précision et de robustesse, tout en étant capable de détecter précisément la zone de recouvrement dans les deux images.

## Mots Clef

Recalage d'image, approche directe, région d'intérêt.

## Abstract

Image registration consists in estimating the geometric (and possibly photometric) transformation that relates a source and a target images. The direct approach is to maximize the photometric similarity for the pixels of a region of interest. This region, often user-defined, must be a part of the source image seen in the target one. It should ideally equals the actual overlap between the images, which is hard to achieve since the overlap is precisely known only after registration.

We propose a novel approach that discovers the overlap while registering the images. It relies on a new point of view to registration: we propose to maximize the number

of pixels (from all the source image) that are visible in the target image and photometrically compatible with their corresponding pixels in the target image. Making use of a robust framework, our method naturally deals with outliers caused, for instance, by occlusions.

Our method is illustrated with two different types of warps: homographies and B-splines. Experiments on synthetic and real data show that our method outperforms standard approaches in terms of accuracy and robustness while precisely retrieving the overlapping region in the source and in the target images.

## Keywords

Image registration, direct approach, region of interest.

## 1 Introduction

Étant données deux images d'une même scène, l'image source et l'image cible, le problème du recalage consiste à déterminer la transformation qui met en correspondance les pixels des deux images. Ce problème trouve ses applications dans de nombreux sujets de la vision par ordinateur comme les mosaïques d'images [11], la super-résolution [8] ou encore le suivi d'objets [4].

Deux grands types d'approches sont utilisés pour recalage des images : l'approche géométrique et l'approche directe. Dans l'approche géométrique [13], la transformation est déterminée à partir d'éléments saillants mis en correspondance entre les deux images. L'approche directe, objet de cet article, consiste à minimiser la différence photométrique entre l'image cible transformée et l'image source pour un certain ensemble de pixels [9, 1]. Étant donné un modèle de transformation géométrique  $\mathcal{W} : \mathbb{R}^2 \times \mathbb{R}^p \rightarrow \mathbb{R}^2$ , paramétré par un vecteur  $\mathbf{p} \in \mathbb{R}^p$ , l'approche directe est exprimée par un problème de minimisation :

$$\min_{\mathbf{p}} \sum_{\mathbf{q} \in \mathfrak{R}} \mathcal{D}(\mathbf{q}; \mathbf{p})^2, \quad (1)$$

où  $\mathcal{D}(\mathbf{q}; \mathbf{p}) = \mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q}; \mathbf{p}))$  pour une image en niveaux de gris et  $\mathcal{D}(\mathbf{q}; \mathbf{p}) = \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{W}(\mathbf{q}; \mathbf{p}))\|$  pour une image possédant plusieurs canaux.  $\mathcal{S}$  et  $\mathcal{T}$  désignent les images source et cible. Une image  $\mathcal{I}$  est considérée comme une fonction continue de  $\Omega_{\mathcal{I}} \subset \mathbb{R}^2$  à valeur dans  $\mathbb{R}^h$  où

$h$  est le nombre de canaux.  $\Omega_{\mathcal{I}}$  désigne le domaine de l'image  $\mathcal{I}$  qui est généralement un sous-ensemble rectangulaire borné de  $\mathbb{R}^2$ . L'évaluation d'une image pour des coordonnées non-entières est effectuée en utilisant une méthode d'interpolation. L'ensemble  $\mathfrak{R}$  est appelé la région d'intérêt (RDI); il s'agit d'un ensemble de pixels du domaine  $\Omega_S$  de l'image source visibles dans l'image cible. L'approche (1) peut être rendue robuste en utilisant un M-estimateur [6] afin de faire face aux données erronées provenant, entre autre, des occultations :

$$\min_{\mathbf{p}} \sum_{\mathbf{q} \in \mathfrak{R}} \rho(\mathcal{D}(\mathbf{q}; \mathbf{p})). \quad (2)$$

Dans cet article nous ne considérons que des M-estimateurs redescendants comme, par exemple, la fonction bicarrée de Tukey :

$$\rho(x) = \begin{cases} \frac{c^2}{6} \left( 1 - \left( 1 - \frac{x^2}{c^2} \right)^3 \right) & \text{si } |x| \leq c \\ \frac{c^2}{6} & \text{sinon,} \end{cases} \quad (3)$$

où  $c$  est une constante qui détermine la sensibilité du M-estimateur aux données erronées. Dans l'idéal, cette constante devrait être égale à 4.685 fois l'écart-type du bruit entachant les images [6]. En pratique, nous considérons que l'écart-type du bruit ne dépasse pas 20% de la valeur maximale d'un pixel.

Le problème (1) est un problème de moindres carrés non-linéaires optimisable avec un schéma itératif comme l'algorithme de Levenberg-Marquardt ou l'algorithme de Gauss-Newton [2]. Le problème (2), quant à lui, peut être optimisé avec une méthode de moindres carrés itérativement pondérés [7].

Un des avantages de l'approche directe est qu'elle ne nécessite pas les délicates étapes que sont la détection d'éléments saillants et leur mise en correspondance, contrairement aux approches géométriques. Cependant, l'approche directe requiert la donnée d'une RDI  $\mathfrak{R}$ . Déterminer une RDI correcte est un problème délicat mais nécessaire. Elle devrait idéalement être égale au recouvrement réel entre les images source et cible. Or ce recouvrement ne peut être connu précisément qu'après recalage. Notons que la RDI doit rester valide tout au long des itérations de l'algorithme résolvant les problèmes (1) et (2).

En pratique, la RDI est souvent déterminée par l'utilisateur qui spécifie un polygone dans l'image source. Elle peut aussi être définie automatiquement comme une zone rectangulaire obtenue en enlevant au domaine de l'image source une marge de taille fixée *a priori* (voir §2.1). Ces solutions ne sont pas satisfaisantes pour leur manque d'automatisme ou leur manque de fiabilité. En effet, il n'existe aucune garantie sur le fait que les RDI déterminées par ce genre d'approches ne soient constituées que de pixels visibles dans les images source et cible simultanément. Inversement, certaines parties du recouvrement réel entre les images source et cible peuvent ne pas être prises en compte,

privant ainsi les algorithmes d'optimisation de données potentiellement utiles à l'estimation de la transformation. Une autre solution, proposée par [12], consiste à utiliser une RDI *adaptive* (voir §2.2). À tout moment, cette fenêtre contient l'ensemble des pixels de toute l'image source qui, une fois transformés, appartiennent au domaine de l'image cible. Cette approche présente toutefois plusieurs défauts : la fonction de coût possède une infinité de minima locaux ne correspondant pas à la solution recherchée, l'algorithme d'optimisation donné par les auteurs ne minimise pas la fonction de coût proposée et la méthode n'est pas robuste aux données erronées.

Nous proposons une nouvelle solution au problème de la RDI. Notre contribution repose sur une vision nouvelle du recalage direct d'images. Nous montrons en §3 que le problème d'optimisation initial (1) peut être vu comme une relaxation d'un problème visant à estimer la transformation géométrique comme celle qui maximise le nombre de pixels compatibles entre l'image source et l'image cible transformée. Cette relaxation fait intervenir une RDI définie *a priori*. Nous proposons en §4 une relaxation différente, ne faisant pas intervenir de RDI. Notre approche est naturellement robuste aux données erronées.

L'approche que nous proposons est générique dans le sens où tout modèle paramétrique de transformation géométrique peut être utilisé. Nous utilisons principalement deux types de transformation : les homographies et les déformations libres à base de B-Splines (voir §5.1). Des expériences utilisant des données synthétiques et réelles ont été réalisées en §5.2 et §5.3 respectivement. Elles valident l'efficacité de notre approche et démontrent ses avantages par rapport aux approches existantes.

## 2 Région d'intérêt : état de l'art

Choisir une RDI convenable est un problème difficile souvent négligé. Une RDI qui contient des pixels n'appartenant pas au recouvrement des images à recaler entraîne l'échec de l'optimisation puisque la fonction de coût ne peut être calculée pour de tels pixels. De même, si la RDI est un sous-ensemble trop restreint du recouvrement alors l'algorithme d'optimisation sera privé de données potentiellement utiles à l'estimation de la transformation géométrique. Notons que le problème de la RDI ne se pose pas lorsque l'on veut recalage un motif entièrement contenu dans l'image cible.

### 2.1 Région d'intérêt rectangulaire

Une des approches les plus communément utilisées pour définir la RDI consiste à postuler une déformation maximale. La RDI est alors choisie comme une sous-partie rectangulaire du domaine  $\Omega_S$  de l'image source en enlevant au domaine  $\Omega_S$  une marge de largeur déterminée *a priori*. Cette largeur devrait idéalement être proportionnelle à l'amplitude maximale de la transformation mais celle-ci est inconnue avant recalage. Outre les répercussions sur le succès du processus d'optimisation, la taille de

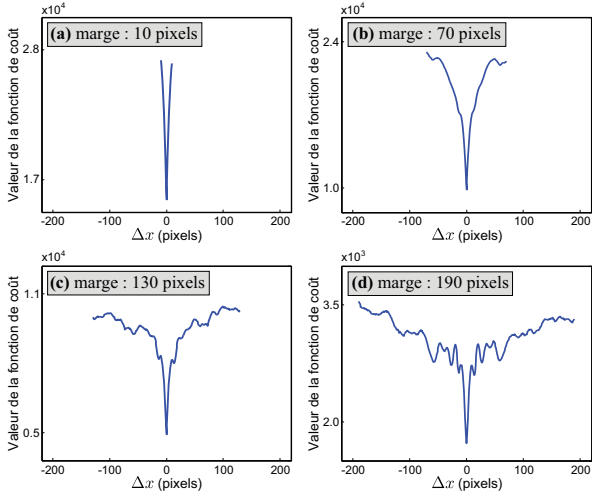


FIGURE 1 – Fonction de coût du problème (1) pour des RDI rectangulaires avec des tailles de marge allant de 10 à 190 pixels (pour des images mesurant  $640 \times 480$ ). Une grande marge étend l'ensemble des transformations admissibles mais accroît le nombre de minima locaux.

la marge influence l'allure de la fonction de coût du problème (1). Une expérience inspirée de [12] permet d'observer ce phénomène. Nous prenons pour modèle de transformation géométrique une translation le long de l'axe des abscisses. La figure 1 montre l'évolution de la fonction de coût du problème (1) en fonction de l'amplitude  $\Delta x$  (en pixels) de la translation. Un bruit gaussien centré d'écart-type égal à 5% de la valeur maximale d'un pixel est ajouté à l'image de la figure 3 afin de générer les images source et cible. Ces images étant identiques, la fonction de coût est minimale pour  $\Delta x = 0$ . Nous observons sur la figure 1 que seule une petite marge donne une fonction de coût bien formée. Cela vient du fait qu'un grand nombre de pixels contribue à la fonction de coût. L'utilisation d'une petite marge a aussi pour effet de réduire l'ensemble des translations admissibles. L'augmentation de la marge fait apparaître de nombreux minima locaux dans la fonction de coût.

## 2.2 Fenêtre adaptative

Une solution au problème de la RDI a été proposée dans [12]. Cette solution consiste à remplacer la RDI  $\mathfrak{R}$ , constante et fixée *a priori*, par une RDI adaptative  $\mathfrak{R}_A(\mathbf{p})$  :

$$\min_{\mathbf{p}} \sum_{\mathbf{q} \in \mathfrak{R}_A(\mathbf{p})} \mathcal{D}(\mathbf{q}; \mathbf{p}). \quad (4)$$

L'ensemble  $\mathfrak{R}_A(\mathbf{p})$  contient tous les pixels<sup>1</sup> du domaine  $\Omega_S$  de l'image source qui, une fois modifiés avec la transformation induite par les paramètres  $\mathbf{p}$ , appartiennent au domaine  $\Omega_T$  de l'image cible. Cela peut être formulé ainsi :

$$\mathfrak{R}_A(\mathbf{p}) = \{\mathbf{q} \in \Omega_S \mid \mathbf{q}' \in \Omega_T\}, \quad (5)$$

1. À l'exception, éventuellement, d'une marge de 1 pixel afin de pouvoir calculer par différences finies le gradient de l'image source.

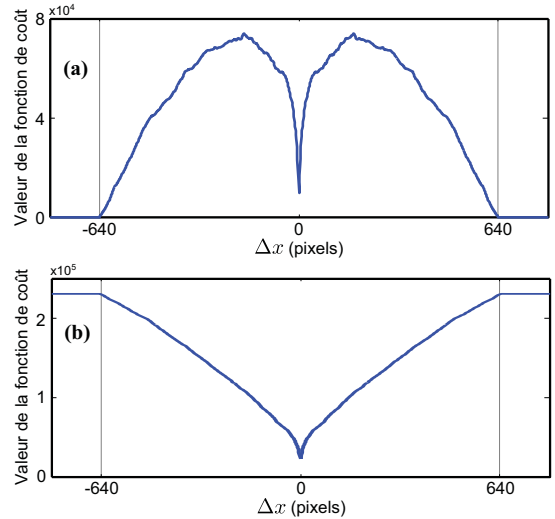


FIGURE 2 – Fonction de coût pour (a) l'approche à RDI adaptative (équation (4)) et pour (b) notre approche (équation (17)). Notre approche est bien posée alors que l'autre présente de « faux » minima dès lors que l'amplitude de la transformation est importante ( $|\Delta x| > 640$ ).

où  $\mathbf{q}' = \mathcal{W}(\mathbf{q}; \mathbf{p})$ . Bien qu'elle permette de s'affranchir de la donnée d'une RDI, cette approche n'est pas complètement satisfaisante. En premier lieu, le problème (4) est mal posé dans le sens où il existe une infinité de minima qui ne correspondent pas à la solution recherchée. Ces minima apparaissent lorsque les domaines de l'image source et de l'image cible transformée sont disjoints. Ce phénomène est illustré par une expérience similaire à celle de la section précédente. Le profil de la fonction de coût du problème (4) est donné par la figure 2(a). Nous constatons sur cette figure que le coût est nul (et donc minimal) dès lors que les domaines sont disjoints ( $|\Delta x| > 640$ ). En second lieu, la dépendance de  $\mathfrak{R}_A$  vis-à-vis de  $\mathbf{p}$  rend le problème (4) difficile à minimiser. Les auteurs de [12] proposent d'optimiser ce problème en omettant la dépendance en  $\mathbf{p}$  de  $\mathfrak{R}_A(\mathbf{p})$  et en alternant l'estimation de  $\mathfrak{R}_A$  et de  $\mathbf{p}$ . Enfin, l'algorithme à RDI adaptative n'est pas robuste aux données erronées et ne peut donc pas gérer convenablement les occultations.

## 3 Maximisation du nombre de pixels compatibles

Dans cette section, nous montrons que la formulation classique (1) peut être vue comme une relaxation d'un autre problème. Ce problème estime la transformation géométrique comme celle qui maximise le nombre de pixels de l'image source qui, une fois transformés, sont compatibles avec leur homologue de l'image cible. Notons  $\mathcal{C}(\mathbf{p})$  l'ensemble des pixels compatibles. Dans un premier temps, nous considérons que l'image cible n'a pas de bords et peut donc être évaluée en tout point de  $\mathbb{R}^2$ . Nous supposons aussi qu'aucun bruit n'entache les données. Dans ces

conditions idéales, l'ensemble  $\mathfrak{C}(\mathbf{p})$  peut être défini ainsi :

$$\mathfrak{C}(\mathbf{p}) = \{\mathbf{q} \in \Omega_S \mid \mathcal{S}(\mathbf{q}) = \mathcal{T}(\mathbf{q}')\}. \quad (6)$$

Dans cet hypothétique monde idéal, le problème du recalage consiste à trouver la transformation telle que tous les pixels de l'image source aient un homologue dans l'image cible, ce qui revient à résoudre le problème suivant :

$$\text{trouver } \mathbf{p} \text{ tel que } \prod_{\mathbf{q} \in \Omega_S} \chi_{\mathfrak{C}(\mathbf{p})}(\mathbf{q}) = 1, \quad (7)$$

où  $\chi_{\mathfrak{C}(\mathbf{p})}$  est la fonction indicatrice de l'ensemble  $\mathfrak{C}(\mathbf{p})$ , c'est-à-dire :

$$\chi_{\mathfrak{C}(\mathbf{p})}(\mathbf{q}) = \begin{cases} 1 & \text{si } \mathbf{q} \in \mathfrak{C}(\mathbf{p}) \\ 0 & \text{sinon.} \end{cases} \quad (8)$$

Bien entendu, la définition (6) et l'équation (7) doivent être relaxées pour tenir compte des contraintes réelles du problème. En premier lieu, il faut prendre en compte le fait que le domaine  $\Omega_T$  de l'image cible est borné. Cela peut être réalisé en introduisant une RDI :

$$\mathfrak{C}^R(\mathbf{p}) = \{\mathbf{q} \in \Omega_S \mid (\mathbf{q} \in \mathfrak{R}) \wedge (\mathcal{S}(\mathbf{q}) = \mathcal{T}(\mathbf{q}'))\}. \quad (9)$$

Il faut aussi prendre en compte le fait que les images sont généralement entachées de bruit. Cela peut être réalisé en considérant  $\mathfrak{C}^R(\mathbf{p})$  comme un ensemble flou [14]. La fonction indicatrice binaire  $\chi_{\mathfrak{C}^R(\mathbf{p})}$  est alors remplacée par une fonction continue  $\mu_{\mathfrak{C}^R(\mathbf{p})} : \mathbb{R}^2 \rightarrow [0, 1]$  qui indique le degré d'appartenance d'un point  $\mathbf{q}$  à l'ensemble  $\mathfrak{C}^R(\mathbf{p})$ . Cette fonction peut être asservie à la probabilité de similarité des pixels  $\mathbf{q}$  et  $\mathbf{q}'$ . En supposant une distribution gaussienne des erreurs,  $\mu_{\mathfrak{C}^R(\mathbf{p})}$  peut être définie par :

$$\mu_{\mathfrak{C}^R(\mathbf{p})}(\mathbf{q}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\mathcal{D}(\mathbf{q}; \mathbf{p})^2}{2\sigma^2}\right). \quad (10)$$

Le problème (7) est alors remplacé par un problème de maximisation de vraisemblance :

$$\max_{\mathbf{p}} \prod_{\mathbf{q} \in \mathfrak{R}} \mu_{\mathfrak{C}^R(\mathbf{p})}(\mathbf{q}). \quad (11)$$

La formulation initiale du recalage d'image (1) est obtenue en prenant l'opposé du logarithme de la fonction de coût dans (11).

En cas d'occultation, les données ne suivent plus une distribution gaussienne. Dans ce cas, les valeurs élevées pour  $\mathcal{D}(\mathbf{q}; \mathbf{p})$ , qui correspondent à des données erronées, ne doivent plus être considérées comme improbables. On leur affecte donc un degré d'appartenance significativement plus grand que zéro :

$$\mu_{\mathfrak{C}^R(\mathbf{p})}^M(\mathbf{q}) = \exp(-\rho(\mathcal{D}(\mathbf{q}; \mathbf{p}))), \quad (12)$$

où  $\rho$  est, par exemple, la fonction bicarrée de Tukey (voir équation (3)). C'est le principe même des M-estimateurs,

comme ils furent introduits par [10]. En remplaçant  $\mu_{\mathfrak{C}^R(\mathbf{p})}$  par  $\mu_{\mathfrak{C}^R(\mathbf{p})}^M$  dans l'équation (11), nous obtenons un problème de maximisation de vraisemblance généralisée.<sup>2</sup> La formulation initiale du recalage d'images robuste (2) est obtenue en prenant l'opposé du logarithme de la fonction de coût dans (11).

## 4 Notre approche

### 4.1 Raisonnement formel

Nous proposons une relaxation différente de la définition (6) et de l'équation (7) permettant de s'affranchir de la donnée d'une RDI. Pour cela, nous considérons que les pixels  $\mathbf{q}$  et  $\mathbf{q}'$  sont compatibles si les valeurs  $\mathcal{S}(\mathbf{q})$  et  $\mathcal{T}(\mathbf{q}')$  sont similaires (lorsque  $\mathbf{q}' \in \Omega_T$ ) et incompatibles sinon. Cette définition correspond à l'ensemble suivant :

$$\mathfrak{C}^T(\mathbf{p}) = \{\mathbf{q} \in \Omega_S \mid (\mathbf{q}' \in \Omega_T) \wedge (\mathcal{S}(\mathbf{q}) = \mathcal{T}(\mathbf{q}'))\}. \quad (13)$$

Nous associons à cet ensemble la fonction indicatrice floue  $\mu^N$ , qui outre la prise en compte du bruit et des données erronées, gère aussi l'appartenance de  $\mathbf{q}'$  à  $\Omega_T$  :

$$\mu_{\mathfrak{C}^T(\mathbf{p})}^N(\mathbf{q}) = \begin{cases} \exp(-\rho(\mathcal{D}(\mathbf{q}; \mathbf{p}))) & \text{si } \mathbf{q}' \in \Omega_T \\ \exp(-\rho(\alpha)) & \text{sinon.} \end{cases} \quad (14)$$

L'appartenance de  $\mathbf{q}'$  à  $\Omega_T$  est gérée en pénalisant les pixels tels que  $\mathbf{q}' \notin \Omega_T$ . La pénalité est choisie de manière à ne pas biaiser le résultat final. Ainsi, la valeur  $\alpha$  est choisie telle qu'un pixel  $\mathbf{q}' \notin \Omega_T$  ait le même score d'appartenance qu'un pixel erroné. Avec le M-estimateur bicarré de Tukey, toute valeur  $\alpha$  telle que  $|\alpha| > c$  est convenable. En combinant les équations (11), (13) et (14), nous obtenons le problème de minimisation suivant :

$$\min_{\mathbf{p}} \sum_{\mathbf{q} \in \Omega_S} \rho(\chi_{\Omega_T}(\mathbf{q}')\mathcal{D}(\mathbf{q}; \mathbf{p}) - (1 - \chi_{\Omega_T}(\mathbf{q}'))\alpha). \quad (15)$$

Cette formulation peut être simplifiée si l'on considère une extension infinie  $\mathcal{T}_\infty$  de l'image cible initiale  $\mathcal{T}$  :

$$\mathcal{T}_\infty(\mathbf{q}) = \begin{cases} \mathcal{T}(\mathbf{q}) & \text{si } \mathbf{q} \in \Omega_T \\ \beta & \text{sinon} \end{cases} \quad \text{et } \Omega_{\mathcal{T}_\infty} = \mathbb{R}^2, \quad (16)$$

où  $\beta$  est une valeur telle que  $|\mathcal{S}(\mathbf{q}) - \mathcal{T}_\infty(\mathbf{q}')| > c, \forall \mathbf{q} \in \Omega_S$ . En pratique, puisque les intensités d'une image numérique sont codées par des valeurs appartenant à un ensemble borné (par exemple,  $[0, 1]$ ), il est facile de déterminer une valeur convenable pour  $\beta$  (par exemple,  $\beta = 2$ ). En notant  $\mathcal{D}_\infty(\mathbf{q}; \mathbf{p}) = \mathcal{S}(\mathbf{q}) - \mathcal{T}_\infty(\mathbf{q}')$ , le problème (15) peut finalement être écrit :

$$\min_{\mathbf{p}} \sum_{\mathbf{q} \in \Omega_S} \rho(\mathcal{D}_\infty(\mathbf{q}; \mathbf{p})). \quad (17)$$

Le profil de la fonction de coût du problème (17) est illustré à la figure 2(b) en utilisant la même expérience que dans

<sup>2</sup>  $\mu_{\mathfrak{C}^R(\mathbf{p})}^M$  ne constitue pas une distribution de probabilité.

les sections précédentes. Cette expérience montre que notre problème est mieux posé que les autres approches dans la mesure où le seul minimum global (atteint en  $\Delta x = 0$ ) correspond à la transformation recherchée. De plus, l'utilisation d'un M-estimateur permet de réduire les oscillations aux petites échelles.

## 4.2 Explication intuitive

La formulation (17) peut être expliquée de manière intuitive. Imaginons que le champ de vue de la caméra cible soit illimité. Une telle caméra produit une image de taille infinie. Un plan avec une ouverture rectangulaire est alors placé entre la « caméra infinie » et la scène, comme illustré à la figure 3(b). La partie de la scène visible à travers l'ouverture correspond à l'image cible réelle,  $\mathcal{T}$ . Le reste de la scène est invisible car il est *occulté* par le plan. Nous pouvons donc proposer un point de vue unifié : les pixels de l'image source qui ne sont pas visibles dans l'image cible, soit parce qu'ils sont occultés par un objet externe, soit parce qu'ils sont hors-champ, peuvent tous être considérés comme des données erronées. La formulation (17) est une mise en œuvre directe de cette vision du problème.

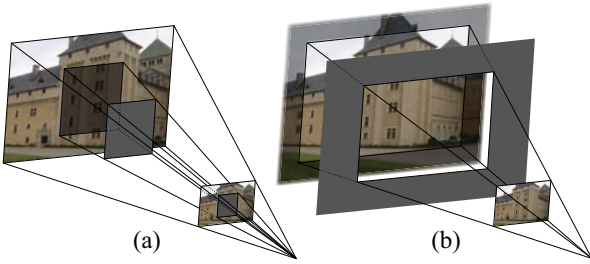


FIGURE 3 – Les objets absents de l'image cibles sont invisibles soit (a) parce qu'ils sont occultés par un objet externe soit (b) parce qu'ils sont hors-champ.

## 4.3 Région d'intérêt et M-estimateur

L'utilisation d'un M-estimateur rend très simple la détection des données erronées. En effet, tout pixel  $\mathbf{q}$  pour lequel  $\rho(\mathcal{D}_\infty(\mathbf{q}; \mathbf{p}))$  est proche de  $\frac{c^2}{6}$  peut être considéré comme un pixel erroné. Par conséquent, le recouvrement entre les images source et cible est égal à l'ensemble des pixels de  $\Omega_S$  qui produisent une réponse strictement inférieure à  $\frac{c^2}{6}$  (en pratique, inférieure à  $\frac{c^2}{6} - \varepsilon$  avec, par exemple,  $\varepsilon = 10^{-4}$ ). Cette propriété intéressante est illustrée par les expériences de la section 5.3.

# 5 Expériences

## 5.1 Modèles de déformation

La méthode que nous proposons dans cet article est générique dans le sens où elle est applicable à n'importe quel type de transformation paramétrique. Nous détaillons ici deux types communs de transformations : les homographies et les déformations libres à base de B-Spline.

**Homographies.** L'homographie est un modèle de déformation global qui explique la relation entre deux images lorsque celles-ci sont prises par une caméra en rotation autour de son centre optique ou lorsque la scène est plane. Une homographie peut être définie de la manière suivante avec  $\mathbf{p} \in \mathbb{R}^8$  et  $\mathbf{q} = (x \ y)^\top$  :

$$\mathcal{W}_H(\mathbf{q}; \mathbf{p}) = \frac{1}{p_7x + p_8y + 1} \begin{pmatrix} p_1x + p_2y + p_3 \\ p_4x + p_5y + p_6 \end{pmatrix}.$$

**Déformations libres.** Les B-Splines sont un type particulier de déformations libres [5] paramétrées par un ensemble de points de contrôle regroupés dans un vecteur  $\mathbf{p} \in \mathbb{R}^{2n_x n_y}$  où  $n_x$  et  $n_y$  sont les nombres de points de contrôle le long de l'axe des abscisses et des ordonnées respectivement. L'expression de cette transformation est donnée par :

$$\mathcal{W}_B(\mathbf{q}; \mathbf{p}) = \sum_{j=1}^{n_y} \sum_{i=1}^{n_x} \begin{pmatrix} p_{2(j(n_x-1)+i)-1} \\ p_{2(j(n_x-1)+i)} \end{pmatrix} N_i(x)N_j(y),$$

où les fonctions  $\{N_i N_j\}_{i=1, \dots, n_x, j=1, \dots, n_y}$  forment la base de l'espace vectoriel des B-Splines [3].

## 5.2 Données synthétiques

**Génération des données.** Les données synthétiques sont générées de la manière suivante. Une transformation géométrique (homographie ou B-Spline) est déterminée en interpolant des correspondances de points générés aléatoirement. La distance entre les points d'une correspondance contrôle la magnitude  $\gamma$  (en pixels) de la transformation. Une proportion  $\alpha$  des images source et cible est ensuite remplacée par des données provenant d'une autre image afin de simuler la présence d'occultations. Enfin, un bruit gaussien centré d'écart-type  $\sigma$  est ajouté aux images source et cible. Nous utilisons des images en couleur dont les intensités sont codées par des valeurs réelles entre 0 et 1. Leur taille est de  $320 \times 240$ .

**Processus expérimental.** Nous étudions l'influence de trois facteurs sur les algorithmes de recalage : l'amplitude  $\gamma$  de la transformation, la quantité de bruit  $\sigma$  et la proportion  $\alpha$  de données erronées. Chaque facteur est étudié indépendamment c'est-à-dire que les facteurs ne varient pas sont fixés à leur valeur par défaut :  $\gamma = 8$  pixels,  $\alpha = 10\%$  et  $\sigma = 0,1$  (soit 10% de la valeur maximale d'un pixel). Différents algorithmes de recalage sont utilisés : RDI rectangulaire (RECT), RDI adaptative de [12] (ADAP) et notre approche (MAXC). Plusieurs variantes de RECT sont considérées : petite (10%) et grande (25%) marge sans M-estimateur (RECTP, RECTG) et avec M-estimateur (RECTPM, RECTGM). Les résultats présentés sont tous obtenus en tant que moyenne sur 100 essais.

**Nombre d'itérations.** La figure 4 reporte le nombre d'itérations que font tous les algorithmes dans différentes situations. Globalement, la convergence est plus rapide avec les homographies qu'avec les B-Splines. Cela est

du au fait que les homographies sont des transformations « globales ». L'apparente rapidité des méthodes à base de RDI rectangulaire est expliquée par le fait que ces approches peuvent s'arrêter prématurément lorsque la RDI n'est plus valide. De manière générale, les résultats obtenus avec notre approche sont meilleurs que ceux obtenus avec l'algorithme ADAP, seule autre méthode ne nécessitant pas la donnée d'une RDI. La vitesse de convergence de notre approche est amortie lorsque l'amplitude de la transformation recherchée est importante. Cela vient de la pénalisation des pixels transformés qui n'appartiennent pas au domaine de l'image cible, ceux-ci étant nombreux pour les transformations de grande amplitude.

**Erreur géométrique.** La figure 5 reporte l'erreur géométrique, c'est-à-dire l'écart en pixels entre la transformation estimée et celle utilisée pour générer les données. Nous observons que, pour tous les algorithmes et toutes les transformations, l'erreur géométrique est peu influencée par la quantité de bruit mais dépend de l'amplitude des transformations et de la proportion de données erronées. Cela est particulièrement vrai pour les approches n'utilisant pas de M-estimateur. Comparée aux autres méthodes, notre approche est celle donnant les meilleurs résultats, atteignant des erreurs de l'ordre d'un pixel (et souvent moins). Ce résultat montre que la pénalisation des pixels transformés qui n'appartiennent pas au domaine de l'image cible ne biaise pas la précision du recalage.

**Erreur photométrique.** L'erreur photométrique moyenne pour la dernière itération des algorithmes est reportée à la figure 6. Notre approche aboutit aux erreurs les plus faibles quel que soit le facteur étudié et quel que soit la transformation géométrique considérée.

### 5.3 Données réelles

**Zones de recouvrement.** La figure 7 montre les zones de recouvrement déterminées lors de la dernière itération des algorithmes RECTPM, RECTGM, ADAP et MAXC. MAXC est la seule méthode estimant correctement l'homographie recherchée. Nous observons aussi que notre méthode est la seule à déterminer précisément la zone de recouvrement, y compris les pixels erronés. RECTPM ne détermine pas le bon résultat car la taille de la marge est inférieure au déplacement. Le faible nombre de pixels utilisés par RECTGM conduit à une estimation imprécise. Le résultat de l'algorithme ADAP est biaisé par les données erronées.

**Panorama maximal.** Une scène est filmée par une caméra en rotation autour de son centre optique et dont le mouvement de gauche à droite est uniforme. La figure 8 montre la mosaïque constituée de la première image de la vidéo et de l'image la plus éloignée de la première pour laquelle les algorithmes étudiés ont correctement estimé l'homographie. Le panorama le plus étendu est obtenu grâce aux algorithmes ADAP et MAXC. Les résultats de ces deux méthodes sont similaires car il y a très peu de données erronées. Les panoramas les moins étendus sont

obtenus avec RECTP et RECTG puisque le déplacement maximal est conditionné par la taille de la marge.

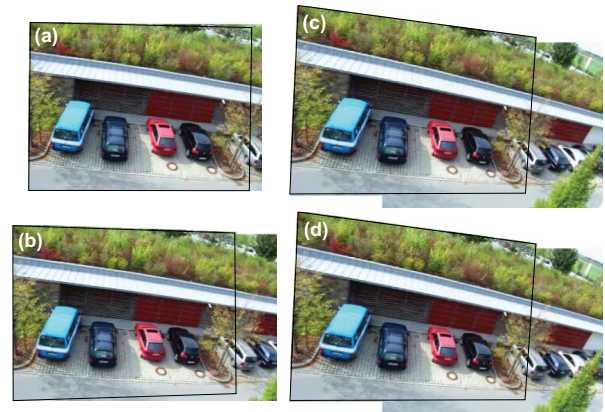


FIGURE 8 – Panorama le plus large pouvant être obtenu avec (a) RECTP, (b) RECTG, (c) ADAP et (d) MAXC.

**Recherche de motif.** Notre méthode permet aussi de déterminer la position d'un motif dans une autre image. La figure 9 illustre cette capacité dans différentes conditions : scène rigide ou déformable et présence de données erronées dues à des occultations ou des spécularités.

**Mosaïque sur scène déformable.** La figure 10 est un exemple de mosaïque d'images en environnement déformable réalisé avec notre méthode.

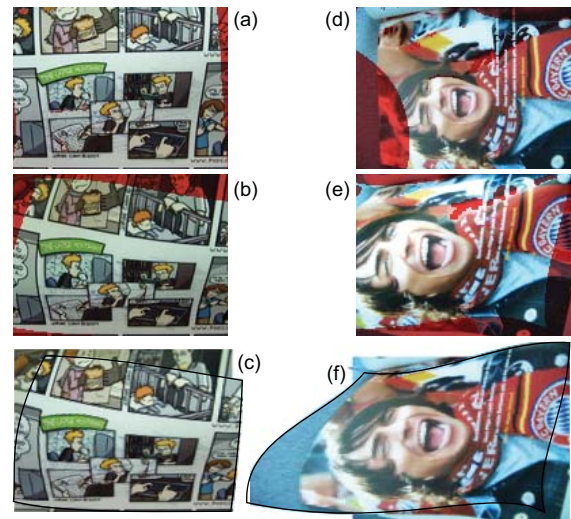


FIGURE 10 – Mosaïques (c,f) en environnement déformable. (a,d) : images source. (b,e) : images cible.

## 6 Conclusion

Nous avons proposé une nouvelle méthode de recalage direct d'images. Tout comme l'approche classique, notre méthode repose sur une relaxation du problème visant à maximiser le nombre de pixels de l'image source compatibles avec leurs homologues de l'image cible. Cependant, notre relaxation a l'avantage de ne pas faire intervenir de région d'intérêt définie *a priori*.

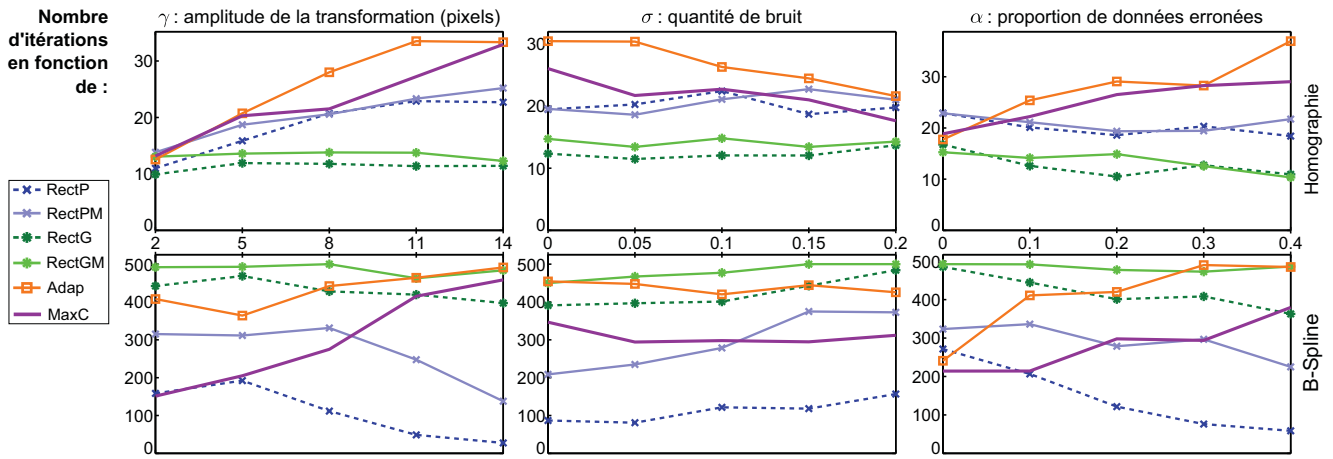


FIGURE 4 – Impact de différents facteurs sur le nombre d’itérations. Le nombre d’itérations des algorithmes à base de RDI rectangulaires est faible car ces méthodes s’arrêtent prématurément lorsque la RDI n’est plus valide.

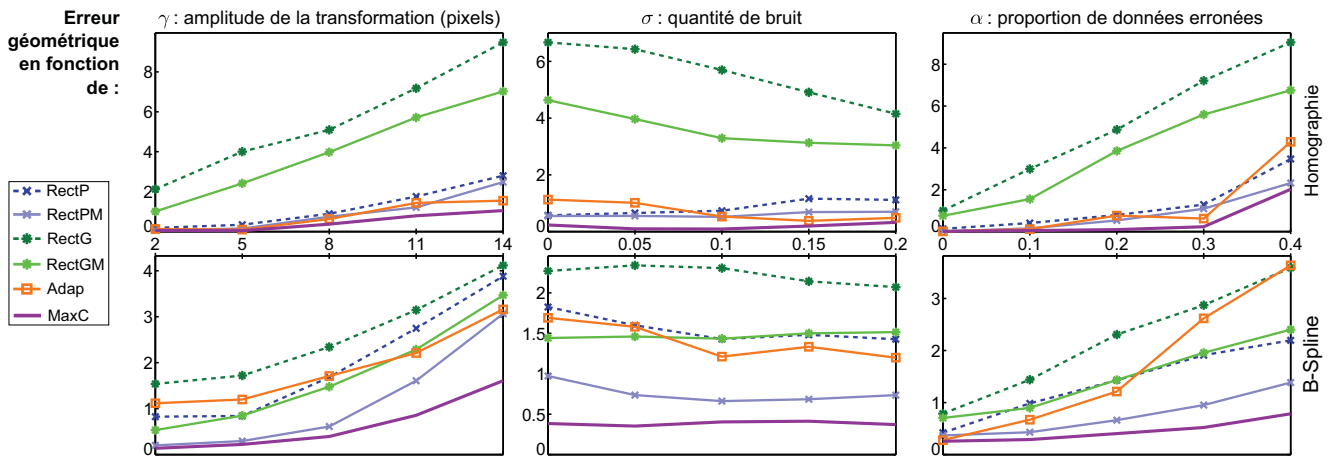


FIGURE 5 – Impact de différents facteurs sur l’erreur géométrique. Notre méthode (MAXC) est celle donnant les meilleurs résultats. Globalement, les approches utilisant un M-estimateur sont les meilleures.

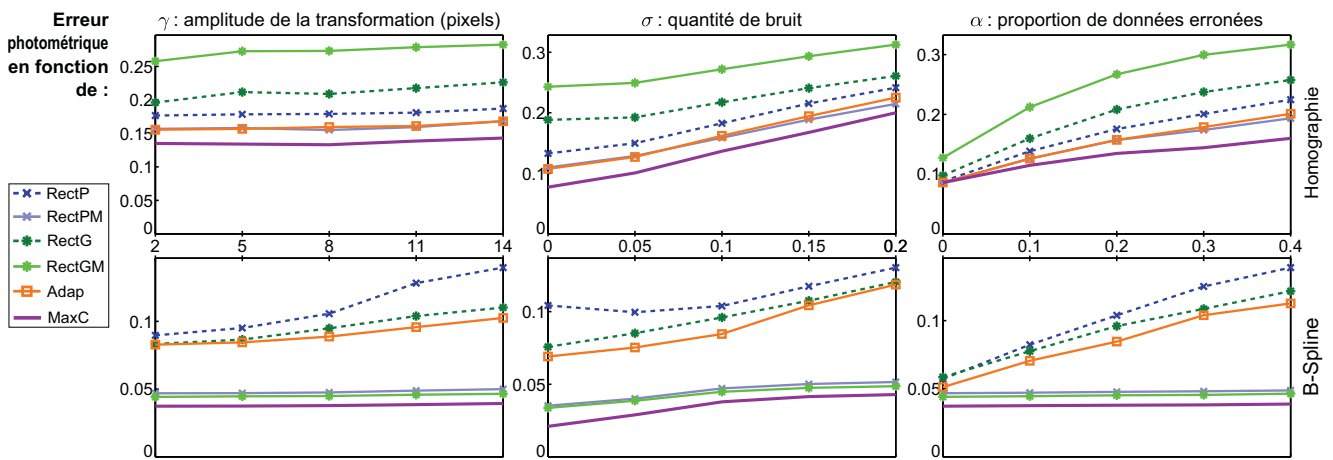


FIGURE 6 – Évolution de l’erreur photométrique moyenne en fonction de différents facteurs.



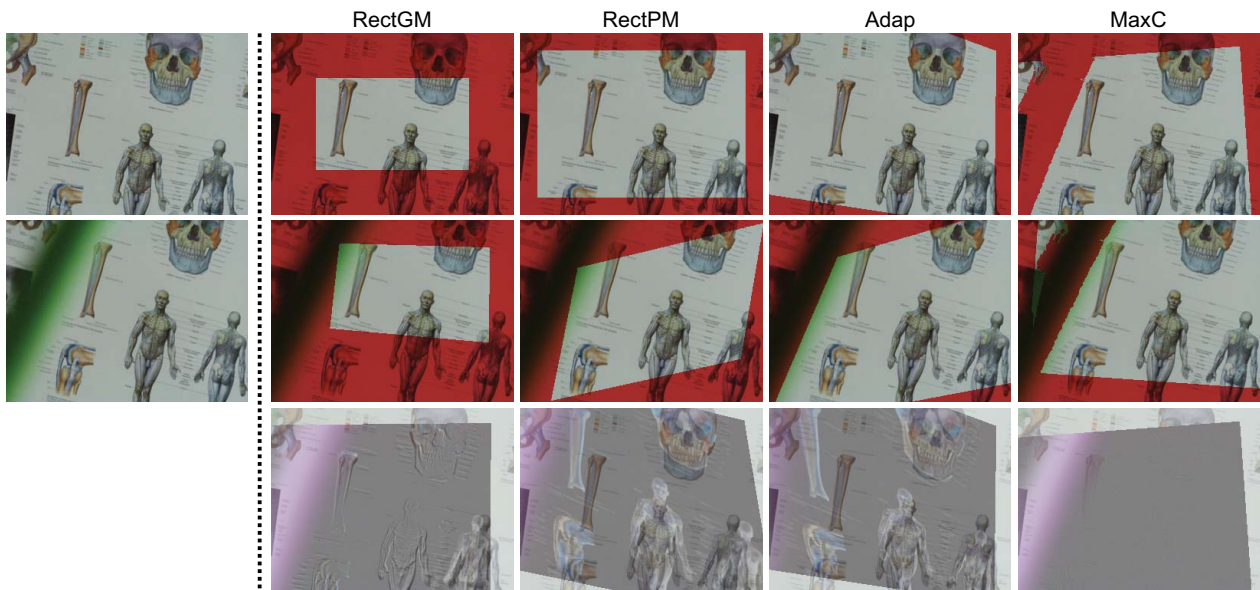


FIGURE 7 – Exemples de recalages pour différents algorithmes. La première et la deuxième lignes correspondent aux images source et cible respectivement. La dernière ligne représente l’image de différence entre l’image source et l’image cible transformée. Les parties rouges représentent les pixels n’appartenant pas aux zones de recouvrement trouvées par les algorithmes.



FIGURE 9 – Recalage d’un motif (a) avec notre algorithme dans différents contextes : scène rigide (b) ou déformable (c) et présence d’une occultation (d) ou d’une spécularité (e). Les pointillés blancs correspondent aux contours du motif recalé. La zone rouge, déterminée par notre algorithme, représente les pixels n’appartenant pas à la zone de recouvrement.

Des expériences ont mis en évidence le fait que notre approche dépasse les approches existantes en terme de précision et de robustesse tout en découvrant automatiquement la zone réelle de recouvrement.

D’éventuelles améliorations pourraient être apportées à notre méthode. Par exemple, il pourrait être intéressant de contraindre la cohérence spatiale de la zone de recouvrement lors de sa découverte.

## References

- [1] A. Bartoli. Groupwise geometric and photometric direct image registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:2098–2108, 2008.
- [2] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [3] C. de Boor. *A Practical Guide to Splines – Revised Edition*. Springer, 2001.
- [4] B. Heigl *et al.* Tracking points in sequences of color images. *Pattern Recognition and Image Understanding*, 1999.
- [5] D. Rueckert *et al.* Nonrigid registration using Free-Form Deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18:712–721, 1999.
- [6] K. V. Arya *et al.* Image registration using robust M-estimators. *Pattern Recognition Letters*, 28:1957–1968, 2007.
- [7] M. Deans *et al.* Terrain model registration for single cycle instrument placement. In *Intelligent Robots and Systems*, 2003.
- [8] R. Hardie *et al.* Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 12:1621–1633, 1997.
- [9] S. Baker *et al.* Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56:221–255, 2004.
- [10] P. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35:73–101, 1964.
- [11] M. Irani and P. Anandan. About direct methods. In *Workshop on Vision Algorithms*, 1999.
- [12] B. Pires and P. Aguiar. Registration of images with small overlap. In *Proceedings of the IEEE sixth workshop on multimedia signal processing*, 2004.
- [13] P. H. S. Torr and A. Zisserman. Feature based methods for structure and motion estimation. In *Vision Algorithms: Theory and Practice, number 1883 in LNCS*, 1999.
- [14] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.