HAL open science

Découverte automatique de la région d'intérêt en recalage d'images direct

Florent Brunet, Adrien Bartoli, Julien Peyras, Nassir Navab, Rémy Malgouyres

To cite this version:

Florent Brunet, Adrien Bartoli, Julien Peyras, Nassir Navab, Rémy Malgouyres. Découverte automatique de la région d'intérêt en recalage d'images direct. ORASIS'09 - Congrès des jeunes chercheurs en vision par ordinateur, 2009, Trégastel, France, France. inria-00404640

HAL Id: inria-00404640

https://inria.hal.science/inria-00404640v1

Submitted on 16 Jul 2009

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 de la région d'intérêt en recalage d'images direct

Direct Image Registration with Automatic Discovery of the Region of Interest

F. Brunet[1,2,3] A. Bartoli[1] J. Peyras[1] N. Navab[2] R. Malgouyres[3]

[1] LASMEA (CNRS / Université Blaise Pascal), Clermont-Ferrand – France
[2] CAMPAR (TU München), Munich – Allemagne
[3] LAIC (Université d'Auvergne) – Clermont-Ferrand – France

florent.brunet@lasmea.univ-bpclermont.fr

Résumé

Le recalage direct et agrégatif d'images d'une séquence vidéo consiste à agréger de manière incrémentale les images en déterminant pour chaque nouvelle image une transformation telle que la différence d'intensité des pixels soit minimisée. Un des défauts majeurs de cette approche directe est la nécessité de disposer d'un a priori sur la zone de chevauchement (ou région d'intérêt) entre l'agrégat courant et l'image à agréger. En pratique, la région d'intérêt est généralement déterminée par l'utilisateur. Dans cet article, nous proposons et comparons différentes approches ne nécessitant pas de connaissances a priori sur les régions d'intérêt.

Mots Clef

Recalage d'image, région d'intérêt, optimisation

Abstract

The goal of the direct aggregative image registration problem is to aggregate in an incremental way a set of images by minimizing the pixel intensity discrepancy between the current aggregate and each new image. One of the bigest drawback of this method is that it relies on an a priori knowledge of the overlapping area (or region of interest) between the aggregate and the new image. In practice, the region of interest is specified by the user. In this article, we propose and compare several new approaches for which an a priori knowledge of the region of interest is not required anymore.

Keywords

Image registration, region of interest, optimization

1 Introduction

Le recalage direct de deux images consiste à appliquer certaines transformations à l'une des deux images de manière à ce qu'elle corresponde le mieux possible à l'autre. Cela peut être vu comme le calcul d'une transformation géométrique, comme une homographie, permettant de déformer la première image jusqu'à ce que la différence d'intensité des pixels avec la deuxième image soit minimale [1, 2, 3, 11, 12]. Lorsque il y a plus de deux images à recaler, comme c'est le cas avec une séquence vidéo, une approche agrégative peut être employée. Cela consiste à recaler successivement chacune des nouvelles images par rapport à l'aggrégat constitué de toutes les images déjà recalées.

Le recalage d'image a constitué un important sujet de recherche de ces dernières décennies. C'est un aspect central dans de nombreuses tâches de vision par ordinateur, d'imagerie médicale ou encore de réalité augmentée. Ce problème trouve des applications dans, par exemple, la construction de mosaïques d'images [11], le suivi d'objets [10, 8] ou encore la super-résolution [9].

D'une manière générale, il existe deux approches pour recaler des images : l'approche géométrique et l'approche directe. L'approche géométrique consiste à déterminer des éléments saillants (points d'intérêts, contours, etc.) dans les deux images à recaler puis à mettre ces éléments en correspondance. La transformation est alors calculée à partir de la différence de localisation des éléments mis en correspondance. L'approche directe, quant à elle, considère l'ensemble des pixels appartenant à la zone de chevauchement (ou région d'intérêt) des deux images. La transformation est alors déterminée de manière à ce que la différence des pixels en intensité entre les deux images soit minimisée.

Bien que les résultats obtenus grâce à l'approche directe soient généralement très bons, cette approche présente un inconvénient majeur. La région d'intérêt, c'est-à-dire la partie de la scène observée qui est visible dans les deux images simultanément, doit être connue *a priori* afin d'obtenir un problème de minimisation bien posé. Cette région d'intérêt est bien souvent déterminée « à la main » par l'utilisateur. Ce travail peut se révéler fastidieux, en particulier lorsque le nombre d'images à traiter est important (ce qui est généralement le cas avec des images issues d'une sé-

quence vidéo).
Nous proposons et comparons trois nouvelles approches permettant de supprimer la nécessité d'une définition *a priori* des régions d'intérêt. En particulier, nous montrons qu'une approche simpliste au problème des régions d'intérêt n'est pas suffisante puisque cela conduit à un problème de minimisation mal posé. Partant de ce constat, nous proposons d'autres méthodes qui résultent en des problèmes d'optimisation mieux posés.

Organisation de l'article. La section 2 présente le cadre général du problème de recalage direct et agrégatif d'images ainsi que l'approche standard permettant de le résoudre. La section 3 détaille nos contributions permettant de s'affranchir des régions d'intérêt. Enfin, un ensemble de résultats expérimentaux sont reportés dans la section 4.

Notations. Les valeurs scalaires sont écrites en caractères italiques (x), les vecteurs en caractères gras ($\mathbf{p}$) et les matrices en caractères sans serif (M). La transposition est notée grâce au symbole $^\mathsf{T}$ ($\mathbf{p}^\mathsf{T}, \mathsf{M}^\mathsf{T}$). Nous utilisons des images en niveaux de gris que nous considérons comme des fonctions de $\mathbb{R}^2$ dans $\mathbb{R}$; elles sont notées en caractères calligraphiques ($\mathcal{I}$). L'évaluation sous-pixélique des images est effectuée par interpolation bilinéaire. Pour une fonction scalaire $\mathcal{F}$ à plusieurs variables, le vecteur colonne $\nabla\mathcal{F}(\mathbf{x})$ est le gradient de $\mathcal{F}$ évalué en $\mathbf{x}$. $\nabla^\mathsf{T}\mathcal{F}(\mathbf{x})$ est aussi le gradient de $\mathcal{F}$ mais disposé en ligne. Étant donné une fonction multivaluée $\mathcal{G}$ à plusieurs variables, la matrice $\mathsf{J}_\mathcal{G}(\mathbf{x})$ est la jacobienne de $\mathcal{G}$ évaluée en $\mathbf{x}$.

2 Énoncé du problème

2.1 Cadre général

Soit $\{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$ un ensemble de n images issues, par exemple, d'une séquence vidéo. Soit $\mathcal{A}_i$ l'agrégat obtenu à la i-ième itération c'est-à-dire l'image construite à partir du recalage successif des images $\mathcal{I}_1$ à $\mathcal{I}_i$ (en prenant $\mathcal{A}_1 = \mathcal{I}_1$ pour l'initialisation). Le recalage de l'image $\mathcal{I}_{i+1}$ par rapport à l'agrégat $\mathcal{A}_i$ consiste à déterminer une transformation géométrique $\mathcal{W}_i$ qui minimise la différence d'intensité entre les pixels de l'agrégat $\mathcal{A}_i$ et la transformée de l'image $\mathcal{I}_{i+1}$. La transformation $\mathcal{W}_i$ établit une correspondance depuis les pixels $\mathbf{q}$ de la région d'intérêt $\mathcal{R}_i$ définie dans $\mathcal{I}_{i+1}$ vers les pixels $\mathcal{W}_i(\mathbf{q};\mathbf{p}_i)$ de l'agrégat $\mathcal{A}_i$. Le vecteur $\mathbf{p}_i$ paramétrise cette transformation. Étant donné un vecteur $\mathbf{p}_i$ approprié, $\mathcal{A}_i(\mathcal{W}_i(\mathbf{q};\mathbf{p}_i))$ doit être proche de $\mathcal{I}_{i+1}(\mathbf{q})$. Formellement, la i-ème étape du recalage agrégatif direct s'écrit comme un problème de moindres carrés non-linéaires :

$$\mathbf{p}_i^\star = \arg\min_{\mathbf{p}} \sum_{\mathbf{q}\in\mathcal{R}_i} \left(\mathcal{A}_i(\mathcal{W}_i(\mathbf{q};\mathbf{p}_i)) - \mathcal{I}_{i+1}(\mathbf{q})\right)^2 \quad (1)$$

La région d'intérêt $\mathcal{R}_i$ est constituée d'un sous-ensemble de pixels qui appartiennent à la zone de chevauchement entre les images $\mathcal{A}_i$ et $\mathcal{I}_{i+1}$. La formulation (1) nécessite une connaissance *a priori* de $\mathcal{R}_i$. En pratique, cette région d'intérêt est donnée par l'utilisateur ce qui constitue un des principaux défauts de l'approche.

L'agrégat $\mathcal{A}_{i+1}$ est calculé à partir de l'agrégat $\mathcal{A}_i$ et de la transformation de l'image $\mathcal{I}_{i+1}$ par $\mathcal{W}_i(\cdot;\mathbf{p}_i^\star)$. Dans cet article, nous ne nous intéressons pas au détail de ce calcul. Notons simplement que de nombreuses approches ont été proposées pour prendre en compte, notamment, les variations d'illumination.

Dans la suite de ce document, les indices correspondant à l'itération sont omis pour des raisons de lisibilité. De même, nous noterons $\mathbf{q}'$ le pixel $\mathcal{W}(\mathbf{q};\mathbf{p})$.

2.2 Processus d'optimisation

Le problème de moindres carrés non-linéaires (1) peut être optimisé en utilisant l'algorithme de Gauss-Newton. Cette section donne le détail de cet algorithme. Appelons $\mathcal{E}_i$ le résidu pour le pixel $\mathbf{q}_i \in \mathcal{R}$:

$$\mathcal{E}_i(\mathbf{p}) = \mathcal{A}(\mathcal{W}(\mathbf{q}_i;\mathbf{p})) - \mathcal{I}(\mathbf{q}_i) \quad (2)$$

et soit $\mathcal{E}$ la fonction définit comme suit :

$$\mathcal{E}(\mathbf{p}) = \begin{bmatrix}\mathcal{E}_1(\mathbf{p}) & \ldots & \mathcal{E}_n(\mathbf{p})\end{bmatrix}^\mathsf{T} \quad (3)$$

où n est le cardinal de l'ensemble $\mathcal{R}$. Avec ces notations, le problème (1) peut être réécrit sous cette forme :

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}} \|\mathcal{E}(\mathbf{p})\|_2^2 \quad (4)$$

L'algorithme de Gauss-Newton consiste à déterminer une approximation de $\mathbf{p}^\star$ de manière itérative. Si on note $\mathbf{p}^{(i)}$ le vecteur de paramètres à la i-ème itération, $\mathbf{p}^{(i+1)}$ est obtenu par :

$$\mathbf{p}^{(i+1)} = \mathbf{p}^{(i)} + \mathbf{d}^{(i)} \quad (5)$$

où $\mathbf{d}^{(i)}$ est le vecteur solution du problème :

$$\mathbf{d}^{(i)} = \arg\min_{\mathbf{d}} \|\mathcal{E}(\mathbf{p}+\mathbf{d})\|_2^2 \quad (6)$$

L'approximation de Taylor au premier ordre de la fonction $\mathcal{E}$ en $\mathbf{p}+\mathbf{d}$ est donnée par :

$$\mathcal{E}(\mathbf{p}+\mathbf{d}) \approx \mathcal{E}(\mathbf{p}) + \mathsf{J}_\mathcal{E}(\mathbf{p})\mathbf{d} \quad (7)$$

En utilisant l'approximation (7), le problème (6) devient un problème de moindres carrés linéaires dont la solution est donnée en utilisant les équations normales [4] :

$$\mathbf{d}^{(i)} = \left(\mathsf{J}_\mathcal{E}^\mathsf{T}\mathsf{J}_\mathcal{E}\right)^{-1}\mathsf{J}_\mathcal{E}^\mathsf{T}\mathcal{E}(\mathbf{p}) \quad (8)$$

2.3 Transformation homographique

Il existe de nombreux modèles paramétriques de transformation géométrique comme, par exemple, les transformations euclidiennes ou les Thin-Plate Spline Warps [5]. Dans cet article, nous considérons que la transformation géométrique $\mathcal{W}$ est une homographie. Ce genre de transformation est typiquement obtenu lorsque la scène observée est planaire ou lorsque le mouvement de la caméra entre deux

prises de vues consiste en une rotation sans translation. La transformation $\mathcal{W}$ s'écrit alors :

$$\mathcal{W}(\mathbf{q};\mathbf{p}) = \frac{1}{p_7 x + p_8 y + 1}\begin{bmatrix} p_1 x + p_2 y + p_3 \\ p_4 x + p_5 y + p_6 \end{bmatrix} \quad (9)$$

où le pixel $\mathbf{q}$ a pour coordonnées (x, y) et où les valeurs p_i pour $i = 1, \ldots, 8$ sont les coefficients du vecteur de paramètres $\mathbf{p}$ servant à paramétrer une matrice d'homographie H :

$$\mathsf{H} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \quad (10)$$

Puisque une expression analytique de la transformation géométrique considérée est connue, il est possible d'expliciter la jacobienne $\mathsf{J}_{\mathcal{E}}(\mathbf{p})$:

$$\mathsf{J}_{\mathcal{E}}(\mathbf{p}) = \begin{bmatrix} \nabla^{\mathsf{T}}\mathcal{E}_1(\mathbf{p}) \\ \vdots \\ \nabla^{\mathsf{T}}\mathcal{E}_n(\mathbf{p}) \end{bmatrix} \quad (11)$$

où $\nabla^{\mathsf{T}}\mathcal{E}_i(\mathbf{p})$ est la transposée du gradient de $\mathcal{E}_i$ évalué en $\mathbf{p}$ et s'écrit comme :

$$\nabla^{\mathsf{T}}\mathcal{E}_i(\mathbf{p}) = (\nabla^{\mathsf{T}}\mathcal{A})(\mathbf{q}'_i)\mathsf{J}_{\mathcal{W}(\mathbf{q}_i;\cdot)}(\mathbf{p}) \quad (12)$$

$\nabla^{\mathsf{T}}\mathcal{A} \in \mathbb{R}^{1\times 2}$ est la transposée du gradient de l'image $\mathcal{A}$; en pratique, ce gradient est calculé numériquement à partir de l'image $\mathcal{A}$ en utilisant un estimateur de dérivées comme celui présenté dans [13]. La matrice $\mathsf{J}_{\mathcal{W}}(\mathbf{p}) \in \mathbb{R}^{2\times 8}$ est la jacobienne de la transformation géométrique évaluée en $\mathbf{p}$. Son expression analytique est donnée par :

$$\mathsf{J}^{\mathsf{T}}_{\mathcal{W}(\mathbf{q}_i;\cdot)}(\mathbf{p}) = \frac{1}{\alpha}\begin{bmatrix} x & 0 \\ y & 0 \\ 1 & 0 \\ 0 & x \\ 0 & y \\ 0 & 1 \\ -\frac{p_1 x^2 + p_2 xy + p_3 x}{\alpha} & -\frac{p_4 x^2 + p_5 xy + p_6 x}{\alpha} \\ -\frac{p_1 xy + p_2 y^2 + p_3 y}{\alpha} & -\frac{p_4 xy + p_5 y^2 + p_6 y}{\alpha} \end{bmatrix}$$

où $\alpha = p_7 x + p_8 y + 1$ et où (x, y) sont les coordonnées du point $\mathbf{q}_i$.

3 Découverte automatique de la région d'intérêt

Dans cette section, nous proposons trois nouvelles approches de recalage d'image permettant de supprimer la nécessité de disposer d'une région d'intérêt *a priori*.

Nous utiliserons les abbréviations données dans la table 1 pour identifier les trois méthodes proposées ainsi que la méthode standard présentée à la section précédente.

3.1 Approche triviale (TRI)

Une approche triviale permettant de s'abstraire de la région d'intérêt dans la formulation (1) consiste à réécrire le problème en ne considérant les différences d'intensité que pour les pixels appartenant à la zone de chevauchement induite par la transformation géométrique :

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}} \sum_{\substack{\mathbf{q}\in\mathcal{I} \text{ t.q.} \\ \mathbf{q}'\in\mathcal{A}}} \big(\mathcal{A}(\mathbf{q}') - \mathcal{I}(\mathbf{q})\big)^2 \quad (13)$$

Dans cette formulation, la région d'intérêt est implicite : elle correspond à l'ensemble des pixels appartenant à l'intersection des domaines de $\mathcal{A}$ et $\mathcal{I}$. Cependant, cette approche triviale présente plusieurs inconvénients. Premièrement, le problème (13) n'est pas bien posé. En effet, lorsque les images $\mathcal{A}$ et $\mathcal{I}$ sont disjointes, le coût est nul. Il existe donc une infinité de minima qui ne correspondent pas à la transformation recherchée. Deuxièmement, la disparition (ou l'apparition) d'un pixel dans la région d'intérêt implicite entraîne une discontinuité dans la fonction de coût. Par conséquent, l'optimisation du problème est délicate. En pratique, nous faisons l'hypothèse que les discontinuités de la fonction de coût sont « petites » et que l'algorithme de Gauss-Newton reste utilisable pour optimiser (13).

3.2 Maximisation du chevauchement (CHM)

L'approche par maximisation du chevauchement (CHM) consiste à introduire des contraintes supplémentaires afin d'éviter les cas dégénérés survenant avec l'approche TRI lorsque les deux images sont disjointes. Plus précisément, une pénalité est ajoutée pour tous les pixels de $\mathcal{I}$ dont l'image par la transformation géométrique n'est pas contenu dans $\mathcal{A}$. Techniquement, l'approche CHM consiste à minimiser (par rapport à $\mathbf{p}$) la fonction de coût suivante :

$$\sum_{\mathbf{q}\in\mathcal{I}} \Big(\mathcal{D}_{\mathcal{A}}(\mathbf{q}')\big(\mathcal{A}(\mathbf{q}') - \mathcal{I}(\mathbf{q})\big) + \big(1 - \mathcal{D}_{\mathcal{A}}(\mathbf{q}')\big)\sigma\Big)^2 \quad (14)$$

où σ est une constante réelle et où $\mathcal{D}_{\mathcal{A}}(\mathbf{q})$ représente le « score d'appartenance » du pixel $\mathbf{q}$ à l'image $\mathcal{A}$. Ce score permet d'obtenir une transition lisse entre les pénalités issues de la différence en intensité des pixels et les pénalités engendrées par la non-appartenance au domaine de $\mathcal{A}$ d'un pixel transformé. $\mathcal{D}_{\mathcal{A}}$ est définit de la manière suivante :

$$\mathcal{D}_{\mathcal{A}}(\mathbf{q}) = \begin{cases} p(d(\mathbf{q},\mathcal{A}); w, c) & \text{si } \mathbf{q}\in\mathcal{A} \\ 0 & \text{sinon} \end{cases} \quad (15)$$

où $d(\mathbf{q},\mathcal{A})$ est la distance euclidienne entre le point $\mathbf{q}$ et le bord de l'image $\mathcal{A}$ le plus proche de $\mathbf{q}$. p est une fonction définit par :

$$p(x; w, c) = \begin{cases} \overline{p}(x) & \text{si } |x - c| \leq w \\ 1 & \text{si } x > c + w \\ 0 & \text{si } x < c - w \end{cases} \quad (16)$$

Abbréviation	Descriptif	Section
STD	Approche directe standard	2
TRI	Approche triviale	3.1
CHM	Approche par maximisation du chevauchement	3.2
NRM	Approche par normalisation du résidu	3.3

TABLE 1 – Abbréviations des méthodes de recalage d'images présentées dans cet article.

où $\overline{p}$ est l'unique polynôme de degrés 6 tel que :

$$\begin{array}{ll} \overline{p}(c-w)=0 & \overline{p}(c+w)=1 \\ \overline{p}'(c-w)=0 & \overline{p}'(c+w)=0 \\ \overline{p}''(c-w)=0 & \overline{p}''(c+w)=0 \end{array}$$

L'approche CHM présente plusieurs avantages. Comparée à l'approche STD, elle permet de supprimer la connaissance *a priori* de la région d'intérêt. De plus, le problème d'optimisation de l'approche CHM est mieux posé que celui de l'approche TRI puisque le cas des minima dégénérés est éliminé. En outre, la fonction de coût (14) est dérivable ce qui la rend plus facile à minimiser. Cependant, l'approche CHM présente l'inconvénient d'être légèrement biaisée puisque les transformations géométriques maximisant le chevauchement sont favorisées.

3.3 Normalisation du résidu (NRM)

La dernière approche que nous proposons (NRM) est une adaptation de l'approche TRI où la fonction de coût est normalisée. Cela est effectuée en divisant le coût du problème (13) par la surface de l'intersection entre les domaines de $\mathcal{A}$ et de $\mathcal{I}$. Ainsi, le coût a tendance à augmenter lorsque la zone de chevauchement se réduit. Prendre simplement la surface de l'intersection n'est pas suffisant. Par exemple, si la transformation géométrique consiste en un fort agrandissement de l'image $\mathcal{I}$, il est possible qu'un seul pixel de l'image $\mathcal{I}$ transformée recouvre entièrement l'image $\mathcal{A}$. La surface de l'intersection serait alors maximale (puisque égale à la superficie totale de l'image $\mathcal{A}$) et le fait que la majeure partie de la transformée de $\mathcal{I}$ ne chevauche pas $\mathcal{A}$ ne serait donc pas pénalisé. Pour remédier à ce genre de problème, la surface de l'intersection est considérée de manière relative à la surface totale de l'image $\mathcal{I}$ une fois transformée.

La fonction de coût du problème de minimisation de l'approche NRM est le suivant :

$$\frac{s(\mathcal{W}(\mathcal{I};\mathbf{p}))}{\max\left(s(\mathcal{A}\cap\mathcal{W}(\mathcal{I};\mathbf{p})),\varepsilon\right)}\sum_{\substack{\mathbf{q}\in\mathcal{I}\text{ t.q.}\\ \mathbf{q}'\in\mathcal{A}}}\left(\mathcal{A}(\mathbf{q}')-\mathcal{I}(\mathbf{q})\right)^2 \quad (17)$$

où $\mathcal{W}(\mathcal{I};\mathbf{p})$ représente la transformée de l'image $\mathcal{I}$ selon la transformation géométrique paramétrée par $\mathbf{p}$. La fonction s, appliquée à une image $\mathcal{J}$, renvoie la superficie du domaine de définition de $\mathcal{J}$.

4 Résultats expérimentaux

4.1 Processus expérimental

Génération d'images synthétiques. Afin de pouvoir évaluer la qualité des approches étudiées, nous générons synthétiquement des couples d'images de manière à ce que l'homographie les reliant soit connue (homographie de référence). Un couple d'images $\mathcal{I} \leftrightarrow \mathcal{I}'$ est construit de la manière suivante :

1. Nous considérons une image quelconque $\mathcal{I}$ (par exemple, une photographie numérique) dont les quatres coins du domaine sont $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ et $\mathbf{c}_4$.
2. Les points $\mathbf{c}_i$ sont transformés de la manière suivante : $\mathbf{c}'_i = \mathbf{c}_i + \mathbf{v}_i$ où $\mathbf{v}_i$ est un vecteur déterminé de manière aléatoire sous la contrainte que sa norme euclidienne soit égale à a pour les quatre points (a est appelé amplitude de la transformation et est exprimé en pixels).
3. Les points $\mathbf{c}'_i$ définissent le quadrilatère représentant le domaine de la deuxième image $\mathcal{I}'$ une fois transformée.
4. L'homographie de référence entre les images $\mathcal{I}$ et $\mathcal{I}'$ est calculée à partir des 4 correspondances $\mathbf{c}_i \leftrightarrow \mathbf{c}'_i$. On note $\underline{\mathsf{H}}$ cette homographie.
5. Pour finir, un bruit additif gaussien centré est éventuellement ajouté aux images $\mathcal{I}$ et $\mathcal{I}'$.

Paramètres par défaut. Sauf mention contraire explicite, il est considéré dans les expériences suivantes que les images utilisées ont été obtenues grâce à la méthode décrite au paragraphe précédent avec une amplitude de 5 pixels et un bruit gaussien dont l'écart type est égal à 10% de la plage de valeurs possibles pour un pixel (typiquement 25,5 pour des images en niveaux de gris classiques).

Puisque les fonctions de coût de toutes les méthodes sont minimisées en utilisant l'algorithme de Gauss-Newton, il est nécessaire de disposer d'une solution initiale. Dans les expériences suivantes, nous prenons toujours le jeu de paramètre $\mathbf{p}$ qui correspond à la matrice identité.

4.2 Exemple de recalage agrégatif

La figure 1 montre un exemple de recalage agrégatif des images d'une séquence vidéo. Pour cet exemple, nous avons utilisé l'algorithme NRM. Les images de la séquence vidéo sont reliées par des homographies dans la mesure où la scène observée est planaire. De plus, nous avons négligé les distortions radiales et tangentielles car la séquence vidéo a été enregistrée un utilisant un niveau de zoom élevé.

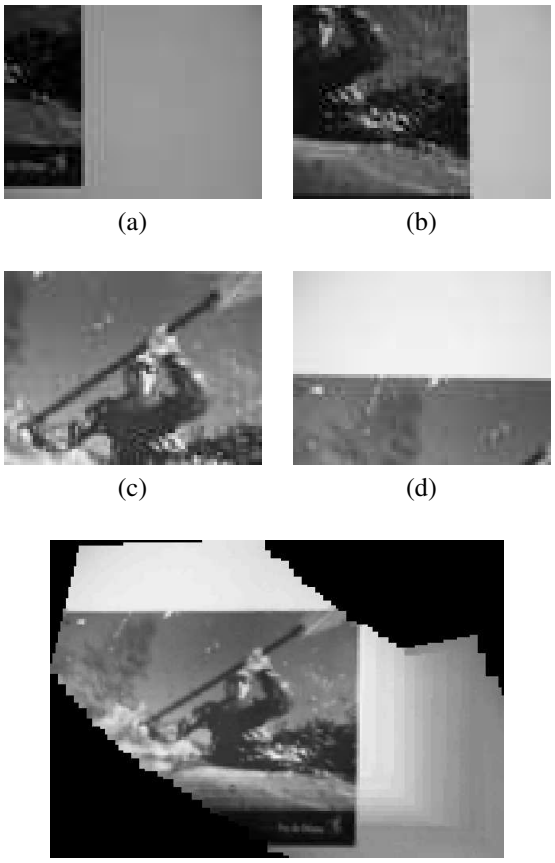

FIGURE 1 – Exemple de recalage agrégatif des images d'une séquence vidéo. Les vignettes (a), (b), (c) et (d) sont quatre des images extraites de la séquence vidéo (parmi environ 120 images au total). La vignette (e) constitue le résultat du recalage.

4.3 Biais des méthodes

Dans cette expérience, nous étudions le biais des quatre méthodes présentées dans les sections précédentes en fonction de l'intensité du bruit ajouté aux images. Soient $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ et $\mathbf{c}_4$ les coins du domaine de l'image $\mathcal{I}'$. En utilisant l'homographie de référence, le domaine de l'image $\mathcal{I}'$ une fois transformée est décrit par les points $\underline{\mathsf{H}}\mathbf{c}_i$ pour $i = 1, \ldots, 4$. Soit H l'homographie estimée avec une des méthode. La qualité de l'homographie H est mesurée en prenant la moyenne des distances euclidiennes entre les points $\underline{\mathsf{H}}\mathbf{c}_i$ et $\mathsf{H}\mathbf{c}_i$. On appelle DMC (Distance Moyenne des Coins) cette mesure. Notons que la DMC est exprimée en pixels. Les résultats de cette expérience sont reportés à la figure 2. Nous constatons que le biais des méthodes STD, TRI et NRM est similaire alors que le biais de la méthode CHM est plus important. Toutefois, notons que pour un bruit avec un écart type inférieur à environ 12 unités de valeur des pixels, l'ensemble des quatre méthodes sont équivalentes.

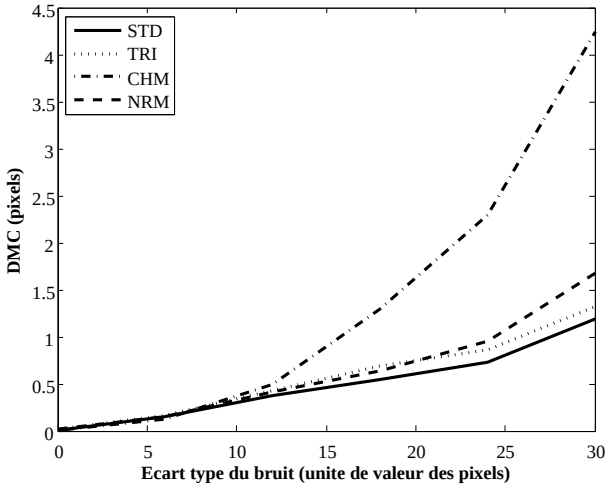

FIGURE 2 – Biais des méthodes en fonction de l'écart type du bruit ajouté aux images de test.

4.4 Vitesse de convergence

Nous étudions dans cette expérience la vitesse de convergence des quatre méthodes décrites dans cette article. Dans la mesure où les problèmes d'optimisation des méthodes présentées sont des problèmes de moindres carrés non-linéaires, nous utilisons l'algorithme de Gauss-Newton pour les résoudre. La vitesse de convergence est évaluée grâce au nombre d'itérations requis pour que l'algorithme converge. Les résultats reportés ont été obtenus en prenant la moyenne des résultats obtenus pour 50 couples d'image différents.

La figure 3 montre le résidu (c'est-à-dire l'écart en intensité des pixels entre l'image de base et l'image à recaler) en fonction de l'itération. Nous constatons que la vitesse de convergence des quatre méthodes est globalement similaire. Il existe cependant quelques différences mineures. La

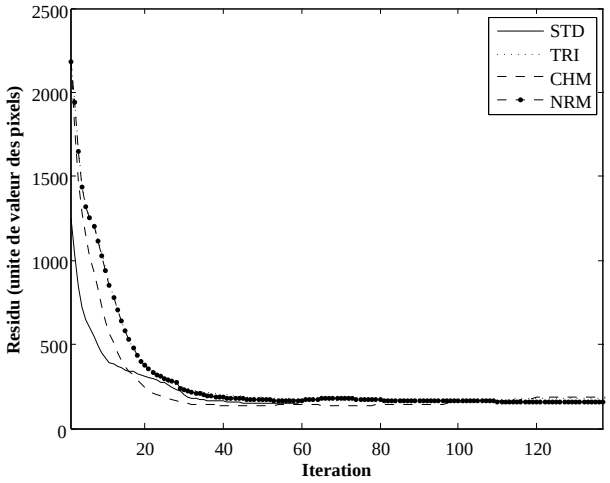

FIGURE 3 – Vitesse de convergence

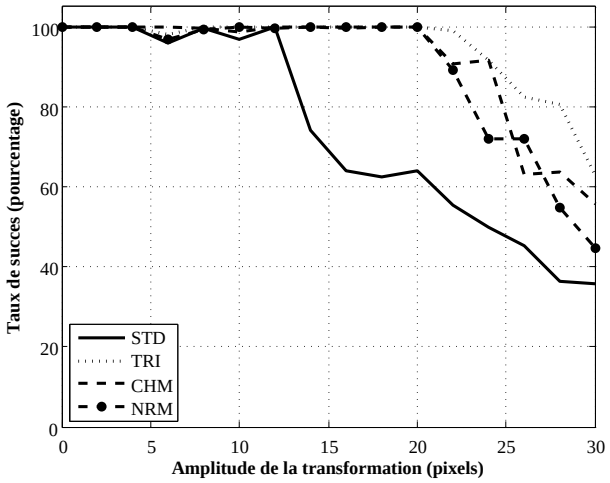

FIGURE 4 – Taux de succés de la convergence des méthodes en fonction de l'amplitude des transformations.

meilleure vitesse de convergence est obtenue pour l'algorithme STD. Cela peut être expliqué par le fait que l'ensemble des pixels utilisés pour le processus d'optimisation sont à la fois présents dans l'image de base et dans l'image à recaler ce qui n'est pas forcément le cas pour les autres. La vitesse de convergence de la méthode CHM est aussi légèrement meilleure que les approches TRI et NRM. Cela provient du fait que la méthode CHM tend à favoriser la transformation identité qui est assez proche de la solution recherchée puisque l'amplitude des transformations considérées est relativement faible (5 pixels). On remarque aussi que les approches TRI et NRM ont des vitesses de convergence identiques. Cela est dû au fait que l'approche NRM est « juste » une normalisation de l'approche TRI.

4.5 Taux de succès

Dans cette expérience, nous montrons la fréquence avec laquelle chacune des quatre méthodes converge vers la solution correcte en fonction de l'amplitude de la transformation géométrique. Il existe plusieurs cas faisant qu'une méthode échoue :

- Une méthode peut échouer lorsqu'elle diverge complètement c'est-à-dire lorsque la transformation estimée implique qu'il n'existe plus aucun pixel commun entre l'image de base et l'image à recaler.
- Une méthode peut aussi échouer lorsque le processus d'optimisation aboutit à une solution correspondant à un minimum local de la fonction de coût.
- Enfin, dans le cas de l'approche STD uniquement, il y a échec dès lors que la région d'intérêt n'est plus entièrement contenue dans les images après transformation.

Les résultats reportés à la figure 4 ont été obtenus à partir de 50 images synthétiques différentes.

Naturellement, le taux de succès est maximal pour une transformation d'amplitude nulle. Pour de réelles transformations, nous constatons sur la figure 4 que les taux de succés de la convergence pour les méthodes TRI, CHM et NRM ne sont pas significativement différents. En revanche, ce taux de succés est plus bas pour la méthode STD. Deux raisons sont à l'origine des mauvais résultats obtenus pour la méthode STD. D'une part, la méthode STD est plus encline à faillir puisque il y a échec dès que la région d'intérêt n'est plus entièrement contenue dans les deux images. D'autre part, utiliser une région d'intérêt élimine de la fonction de coût certains pixels qui font pourtant parti du champ de vue commun aux deux images. Ainsi, moins d'information est utilisée par la méthode STD que par les trois autres approches.

5 Conclusion

Nous avons présenté trois approches permettant de supprimer la nécessité d'une définition *a priori* des régions d'intérêt dans le problème du recalage direct d'images. Outre l'aspect pratique de ces méthodes, les expérimentations montrent que supprimer la région d'intérêt donne de meilleurs résultats que l'approche classique. Cela est principalement dû au fait que les approches sans région d'intérêt exploitent toute l'information contenue dans les images.

Parmi les trois approches proposées, les expérimentations mettent en évidence un très léger avantage pour la méthode TRI. Cependant, il faut bien noter que cet avantage n'est pas tellement significatif. De plus, l'aspect théorique de l'approche TRI n'est pas satisfaisant puisque la fonction de coût optimisée est mal posée : ses minima ne correspondent pas obligatoirement à la solution recherchée. Il en va de même avec l'approche NRM. En effet, la fonction de coût avec cette approche est indéfinie dès lors que la transformation géométrique est telle que les deux images n'ont plus de parties communes. De plus, l'approche NRM présente une vitesse de convergence plus faible que celle des autres méthodes. La meilleure approche est probablement l'approche CHM. En effet, même si l'expérimentation met en évidence que cette approche est légèrement biaisée, il s'agit de la seule méthode qui conduise à une fonction de

coût correcte. Le biais de cette méthode pourrait être réduit en se plaçant dans un contexte robuste. En affectant aux pixels qui sortent de l'image cible une pénalité similaire à celle des outliers et en utilisant une fonction de coût robuste, l'impact des points responsables du biais serait limité. De plus, l'utilisation d'une fonction de coût robuste (à base, par exemple, de M-estimateur) permettrait de prendre en compte les occultations et les erreurs d'acquisition.

Références

[1] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on : A unifying framework : Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Pittsburgh, PA, November 2003.

[2] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on : A unifying framework. *International Journal of Computer Vision*, 56 :221–255, 2004.

[3] A. Bartoli. Groupwise geometric and photometric direct image registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12) :2098–2108, December 2008.

[4] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, 1996.

[5] F.L. Bookstein. Principal warps : Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6) :567–585, 1989.

[6] V. Gay-Bellile, A. Bartoli, and P. Sayd. Direct estimation of non-rigid registrations with image-based self-occlusion reasoning. In *Proceedings of the Eleventh IEEE International Conference on Computer Vision*, 2007.

[7] V. Gay-Bellile, M. Perriollat, A. Bartoli, and P. Sayd. Image registration by combining Thin-Plate Splines with a 3D morphable model. In *Proceedings of the International Conference on Image Processing*, 2006.

[8] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, and A. Trémeau. Feature point tracking : Robustness to specular highlights and lighting variations. In *European Conference on Computer Vision*, 2006.

[9] R. Hardie, K. Barnard, and E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 12 :1621–1633, 1997.

[10] B. Heigl, D. Paulus, and H. Niemann. Tracking points in sequences of color images. *Pattern Recognition and Image Understanding*, 1999.

[11] M. Irani and P. Anandan. About direct methods. In *Workshop on Vision Algorithms*, pages 267–277, 1999.

[12] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, 1981.

[13] R. Malgouyres, F. Brunet, and S. Fourey. Binomial convolutions and derivatives estimation from noisy discretizations. In *Discrete Geometry for Computer Imagery*, 2008.

[14] H. S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 1999.