

Fusion de données spatiales (3D) et temporelles (2D+t) pour la reconstruction de scènes 3D+t : deux approches

Wafa Rekik, Dominique Béréziat, Séverine Dubuisson

► **To cite this version:**

Wafa Rekik, Dominique Béréziat, Séverine Dubuisson. Fusion de données spatiales (3D) et temporelles (2D+t) pour la reconstruction de scènes 3D+t : deux approches. [Rapport de recherche] 2008. inria-00583277

HAL Id: inria-00583277

<https://hal.inria.fr/inria-00583277>

Submitted on 1 Jun 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.

Rapport de recherche LIP6

**Fusion de données spatiales (3D) et temporelles
(2D+t) pour la reconstruction de scènes 3D+t : deux
approches.**

**Merging 3D Spatial Data and 2D+t Temporal
Sequences for 3D+t Scene Reconstruction : Two
Approaches.**

Wafa Rekik¹, Dominique Béréziat^{1,2}, Séverine Dubuisson¹.

¹ Université Pierre et Marie Curie

² INRIA

Correspondance : Dominique.Bereziat@inria.fr

Mai 2008



Résumé

L'objet de cette étude est la reconstruction de scènes Temps/Volume, ou, $3D+t$ à partir de données statiques 3D encadrant des séquences temporelles $2D+t$. La combinaison de ces données complémentaires fournit une séquence décrivant la dynamique 3D des structures d'intérêt en reconstruisant le volume original à partir de chaque image 2D de la séquence temporelle. Le problème de reconstruction inverse 2D-3D est complexe et mal posé à cause du manque d'information et de la distorsion spatiale induite par la vue projective dans chaque image 2D. Afin de compenser ce manque d'information critique, nous modélisons le processus de formation des images 2D à partir du volume 3D original et nous mettons en correspondance chaque image 2D avec le volume 3D connu à l'instant précédent. Cette mise en correspondance est établie/schématisée par un champ de vecteurs de déplacements quantifiant les variations de position de chaque voxel entre deux instants successifs. Pour ce faire, nous utilisons l'hypothèse de conservation de la luminosité. Nous proposons deux approches pour estimer le mouvement des voxels et nous reconstruisons la scène $3D+t$ de proche en proche estimé par compensation du mouvement. Des applications potentielles de cette étude concernent la biologie microscopique cellulaire et l'imagerie médicale.

Mots Clef

Reconstruction 3D, compensation du mouvement, contrainte du flot optique.

Abstract

The main focus of this paper is $3D+t$ shape recovery from 3D spatial data framing $2D+t$ temporal sequences. Merging these two kind of data yields a temporal sequence describing 3D dynamic of structures of interest by reconstructing the original volume from each 2D frame. 2D to 3D inverse reconstruction is a complex and ill posed problem due to the great deal of distortion and in-depth information loss generated by the projective view in each 2D frame. To handle this critical lack of information, we model relevantly the 3D-to-2D transformation that reduces each 3D image into a 2D one and we match each frame with the 3D image known at a previous time. Assuming that structures evolve scarcely between a couple successive moments, these images are matched by a displacement vector field measuring voxel position variation meanwhile. We embed the luminosity conservation constraint and we propose two approaches to estimate voxel movement

over time. We reconstruct gradually the 3D+t sequence by motion compensation. Some applications observing in different modalities 2D+t and 3D data can benefit from this study such as microscopical cell biology or medical imaging.

Keywords

3D reconstruction, motion compensation, optical flow constraint.

1 Introduction

La reconstruction d'une scène tridimensionnelle à partir d'indices visuels bidimensionnels ou d'images est l'une des problématiques les plus étudiées de nos jours. Elle trouve un intérêt en vision par ordinateur, en imagerie médicale et biologique ou dans des applications plus récentes comme le contrôle de qualité en industrie, *etc.* En particulier, nous nous intéressons au cas générique suivant : nous disposons d'images volumiques (3D) et de séquences d'images 2D (2D+t).

1. Les données **3D** définissent la géométrie en trois dimensions de la scène observée. Elles comportent des informations purement spatiales.
2. Les séquences **2D+t** sont porteuses d'informations temporelles et partiellement spatiales puisqu'elles décrivent en deux dimensions la dynamique des objets en mouvement.

En général, les données 3D sont contiguës aux séquences temporelles 2D+t (voir la figure 1). Pour exploiter d'une manière exhaustive l'information contenue dans les données 2D+t, il est intéressant de reconstruire une scène temps/volume ou 3D+t. Cette reconstruction est réalisée en restaurant une structure 3D originale à partir de chaque image 2D de la séquence temporelle intermédiaire. Les données 3D dont nous disposons peuvent alors servir de conditions initiales aux procédés de reconstruction 3D.

La situation décrite dans la figure 1 existe en imagerie biologique où le dispositif d'acquisition est en mesure de fournir une image 3D – en faisant varier la mise au point de ce dispositif et en utilisant un procédé de déconvolution calibrée – ou une séquence d'images mais focalisée à un plan précis dans la scène. On a donc un compromis entre précision temporelle et précision spatiale. Il existe un second cas de figure, en imagerie médicale. En acquisition par rayons X, on peut obtenir des images 3D par un scanner mais au prix d'une exposition plus longue et moins ciblée du patient. Mais on peut également se contenter d'une simple acquisition X projective, comme un angiogramme, qui permet d'obtenir la précision temporelle souhaitée. Le compromis se situe plutôt en terme de minimisation d'irradiation du patient.

Malheureusement, nous ne disposons pas, pour cette étude, de données réelles. Nous nous plaçons donc dans un contexte méthodologique dans lequel les données utilisées sont synthétiques ce qui a pour avantage de quantifier la pertinence des méthodes.

Dans cet article, nous proposons une extension à nos travaux de reconstructions 3D+t [13] qui fait maintenant abstraction de la forme des objets observés.

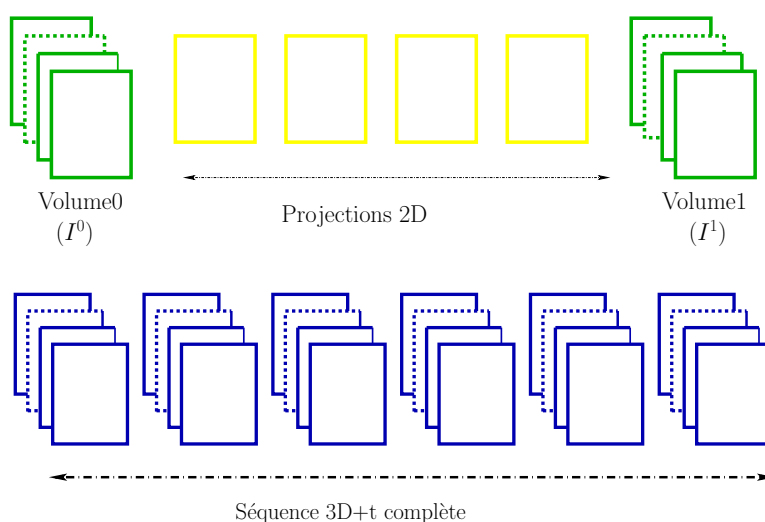


FIG. 1 – Reconstruction 3D+t complète à partir de volumes 3D ponctuels et d’acquisitions temporelles 2D+t.

Nous posons toutefois deux hypothèses. La première hypothèse concerne le procédé d’acquisition, en particulier, nous supposons connu le procédé qui transforme une scène 3D en une image 2D. Cette transformation est modélisée comme une transformée linéaire et stationnaire. La seconde hypothèse concerne la dynamique d’évolution des images 3D à reconstruire : dans cette séquence nous supposons que tous les points suivent une trajectoire de même valeur de niveaux de gris, autrement dit, nous utilisons une hypothèse de flot optique. L’hypothèse de transport de la luminosité doit se retrouver dans la séquence 2D+t et c’est ce qui nous permet de reconstruire la séquence 3D+t. Évidemment, l’utilisation d’une contrainte de type flot optique est un problème mal posé. Aussi nous proposons, à des fins de comparaison, deux méthodes de résolution. La première méthode est variationnelle : elle minimise une énergie qui mesure la discrédance du flot optique dans le domaine 2D en utilisant la transformée 3D-2D et la régularité spatiale du champ de déplacements. La seconde méthode élimine le critère de régularité spatiale du flot optique en calculant un champ constant par morceaux : le flot optique est alors calculé par moindres carrés à partir d’un système linéaire surcontraint.

2 Travaux précédents

Un grand nombre de méthodes, qui fournissent un modèle ou une image 3D à partir de vues projectives 2D, ont été proposées ces dernières années. Ces méthodes dépendent complètement du contexte d'étude : système d'acquisition et modalité d'image, type de traitement d'image souhaitée, nature de la scène observée, *etc.* Dans cette brève revue méthodologique, nous distinguons deux grandes classes de méthode.

La première classe regroupe toutes les approches proposant une reconstruction 3D effective. Cette reconstruction peut être complète en fournissant une image volumique complète : la scène 3D est décrite par une collection dense de voxels. Dans ce type de reconstruction, on privilégie l'information textuelle au détriment de la localisation des structures. La reconstruction peut aussi être incomplète en fournissant, par exemple, des reconstructions 3D de surfaces, généralement des objets ciblés dans la scène 3D. Dans ce type de reconstruction, on met l'accent sur l'information topologique ou la visualisation. Ces reconstructions sont obtenues à partir d'une série d'acquisition 2D. Nous pouvons alors identifier deux sous-groupes de méthodes selon que les acquisitions sont réalisées à partir d'un même angle de vue ou d'angles de vue différents. Dans le premier cas, observant la scène toujours d'un même point de vue, on peut focaliser l'acquisition à un plan de profondeur précis. Selon le type de capteur (focal, confocal) la focalisation est plus ou moins précise mais elle passe toujours par une étape de déconvolution [3] qui permet d'éliminer l'information parasite qui provient de plans de focalisation adjacents. Cette étape de traitement est difficile, elle nécessite des conditions d'acquisition calibrées qui ne sont pas toujours possibles. La calibration permet de rendre réalisable la déconvolution qui est un problème, rappelons-le, mal posé. Dans le second cas, où l'on observe donc la scène à différents points de vue, la situation est plus favorable. À partir d'un grand nombre d'acquisitions, on peut utiliser des techniques de reconstruction tomographique [11] pour obtenir une image 3D complète. Le principe de la tomographie repose sur l'inversion de la transformée de radon, qui est un problème bien posé. Cette technique est largement utilisée en imagerie médicale (CT-X, IRM). Avec deux ou trois caméras, on peut aussi reconstruire une image 3D par stéréoscopie [7]. Le procédé de reconstruction est purement géométrique et bien posé. Toutefois, la stéréoscopie nécessite un bon alignement des images, donc une phase de recalage qui peut être plus ou moins difficile selon la nature des images et des scènes. La stéréovision est utilisée en vision robotique mais aussi en imagerie médicale (images ultrasons 3D).

La seconde classe regroupe les méthodes de mise en correspondance 2D-3D.

Elles ont pour objectif de reconstruire des primitives 3D à partir d'une seule vue projective. Son rôle est de plus en plus important en imagerie médicale où elle est connue sous le terme de recalage 2D-3D [10]. Ce recalage est défini par la recherche de la meilleure transformation permettant de replacer un ensemble de données multimodales dans un espace tridimensionnel commun. En utilisant des images issues de diverses modalités, il permet de faire apparaître l'information qui manque à une modalité dans l'autre modalité. La majorité des applications pour ce genre de recalage se trouve en radiologie. Elles concernent l'alignement d'images scanner (CT) à une ou plusieurs images de projection à rayons X (images de simulateurs, images portiques, images générées par la fluoroscopie, *etc.*). Ces méthodes d'alignement dépendent bien-sûr du contexte applicatif et des modalités utilisées. Nous pouvons distinguer essentiellement les méthodes basées sur des primitives caractéristiques comme notamment les contours [1] et les méthodes basées directement sur l'intensité de l'image [5, 8, 14].

Pour notre part, nous avons proposé une méthode de reconstruction 2D+t vers 3D+t utilisant des informations *a priori* sur la structure topologique des objets observés [13]. Cette méthode fournit d'abord une reconstruction de données géométriquement correctes mais souffrant de l'ambiguïté de positionnement dans la direction z , *i.e.* en profondeur. Cette ambiguïté est ensuite levée par une régularisation spatiale adéquate. C'est en continuité à ces travaux antérieurs que se situe notre étude. Notamment, les méthodes proposées ici permettent de s'affranchir des hypothèses faites sur la nature topologique des objets observés dans les acquisitions. Nous nous sommes placés dans un contexte générique où les acquisitions sont réalisées selon un point de vue unique, comme c'est le cas en imagerie biologique. Nous supposons connue la transformation de la scène 3D en une image 2D.

3 Modélisation et hypothèses utilisées

Nous menons la reconstruction 3D+t à partir d'acquisitions temporelles 2D+t, entrecoupées par des acquisitions volumiques 3D. Nous adoptons dorénavant les notations suivantes :

- \mathbf{X} un vecteur position 3D, *i.e.* $\mathbf{X} = (x, y, z)$;
- \mathbf{x} un vecteur position 2D, *i.e.* $\mathbf{x} = (x, y)$;
- \mathbf{W} champ de vecteurs vitesse 3D, *i.e.* $\mathbf{W} = (u, v, w)$;
- $I(\mathbf{X}, t)$ une image 3D+t ;
- $I_{2D}(\mathbf{x}, t)$ une image 2D+t ;

– $\hat{I}(\mathbf{X}, t)$ l'estimée d'une image 3D+t ;

Concrètement, nous disposons de deux images 3D acquises à deux instants distincts : $I^0 = I(\mathbf{X}; 0)$ et $I^1 = I(\mathbf{X}; 1)$. Ces deux images bordent une séquence temporelle $I_{2D}(\mathbf{x}; t)$ décrivant les projections successives des volumes $I(\mathbf{X}; t)$ entre l'instant initial $t = 0$ et l'instant final $t = 1$ (voir figure 1). Notre étude méthodologique repose sur l'existence d'une fonctionnelle p qui réduit une scène réelle tri-dimensionnelle en une image bi-dimensionnelle. Ainsi, chaque image 2D de la séquence vidéo $I_{2D}(\mathbf{x}; t)$ est une projection de la structure 3D réelle correspondante. Cette fonctionnelle dépend uniquement du processus d'acquisition des données. Nous modélisons p par une transformation linéaire et stationnaire :

$$p : \quad E \quad \longmapsto \quad F$$

$$p(I)(x, y) \quad \longmapsto \quad \int_{\mathbb{R}} I(x, y, z) \underbrace{a(z)}_{d\mu(z)} dz, \quad (1)$$

E correspond au domaine des images 2D et F à celui des images 3D et, la densité de mesure $a(z)$ dépend du dispositif d'acquisition et en particulier du plan de focalisation et $d\mu(z)$ est une mesure de Lebesgue. D'après cette formulation, cette projection intègre la contribution de tous les plans d'une structure 3D pour former une image 2D, la contribution de chaque plan est pondérée par la densité de mesure $a(z)$. La position de ce plan dépend des caractéristiques du dispositif d'acquisition et principalement de la distance focale. La transformation induit par conséquent une perte d'information importante dans la direction z . L'objectif est de réaliser une reconstruction 3D+t (temps/volume) complète de la scène observée à partir de deux échantillons I^0 et I^1 et d'une séquence vidéo $I_{2D}(\mathbf{x}, t)$. Les données 3D sont statiques, alors que la séquence $I_{2D}(\mathbf{x}, t)$, partiellement spatiale, présente une résolution temporelle intéressante. Nous souhaitons restituer une structure 3D à partir de chaque projection $I_{2D}(\mathbf{x}; t)$ existante. Évidemment, ce problème est mal posé puisqu'il existe une infinité d'images 3D correspondant à la projection 2D. Toutefois, nous pouvons contraindre l'espace des solutions en utilisant les valeurs des bords données par les images I^0 et I^1 et en imposant une contrainte de régularité. $\hat{I}(\mathbf{X}, t)$ désigne l'estimée tri-dimensionnelle déduite de $I_{2D}(\mathbf{x})$ à l'instant t . Une condition sur $\hat{I}(\mathbf{X}, t)$ découle directement du principe de reconstruction adopté. En effet, il s'agit de maximiser la ressemblance entre la projection de l'estimée 3D à l'instant t , *i.e.* $p(\hat{I})(\mathbf{X}, t)$, et la donnée $I_{2D}(\mathbf{x}, t)$. Cela nous fournit une première équation d'attache aux données qui contraint $\hat{I}(\mathbf{X}, t)$:

$$p(\hat{I})(\mathbf{x}, t) - I_{2D}(\mathbf{x}, t) = 0, \quad \forall t \in]0, 1[\quad (2)$$

Puisque les instants t et $t - 1$ sont supposés très proches et que les objets d'intérêt évoluent lentement pendant cette durée, nous pouvons mettre en correspondance $\hat{I}(\mathbf{X}, t)$ avec la structure 3D $I(\mathbf{X}, t - 1)$ connue à l'instant $t - 1$. Cette mise en correspondance est quantifiée en terme d'un champ de déplacement \mathbf{W} entre les deux images. Il exprime l'hypothèse qu'un point en mouvement garde la même valeur de niveau de gris au cours du temps ce qui implique la contrainte de mouvement suivante :

$$\hat{I}(\mathbf{X}, t) = I(\mathbf{X} + \mathbf{W}, t - 1) \quad (3)$$

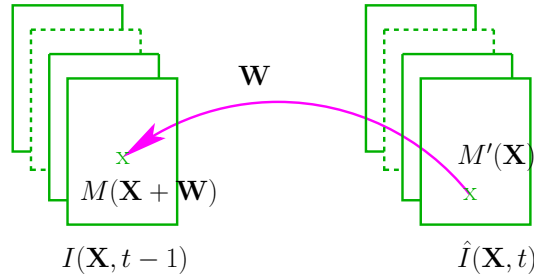


FIG. 2 – Mise en correspondance entre 2 structures 3D successives.

Nous estimons un champ rétrograde (voir figure 2) car la reconstruction des images par compensation de mouvement rétrograde simplifie le processus d'interpolation des voxels manquant dans les images estimées. Néanmoins, nous visualisons, dans la suite de l'article le champ opposé, *i.e.* $-\mathbf{W}$, traduisant le mouvement réel des pixels de $t - 1$ à t . Puisque la distance temporelle entre deux acquisitions vidéo est infiniment petite, nous pouvons étendre le second membre de l'égalité précédente à un développement limité d'ordre 1. L'équation précédente s'écrit alors :

$$\hat{I}(\mathbf{X}, t) = \nabla I(\mathbf{X}, t - 1) \cdot \mathbf{W} + I(\mathbf{X}, t - 1), \quad (4)$$

où $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$ est l'opérateur gradient. Puisque p est linéaire, l'injection de l'équation (4) dans la contrainte d'attache aux données (2) produit une contrainte de données et de mouvement relative à \mathbf{W} :

$$p(\nabla I(\mathbf{X}, t - 1) \cdot \mathbf{W}) + p(I(\mathbf{X}, t - 1))(\mathbf{x}, t) - I_{2D}(\mathbf{x}, t) = 0 \quad (5)$$

Notre but est de résoudre l'équation (5) afin d'estimer un champ de déplacements \mathbf{W} à chaque instant t . Nous présentons deux types d'approches d'estimation du

champ de vitesse de proche en proche. La première fournit un champ de vitesse C^0 par morceaux et la seconde estime un champ dense de vecteurs et régularisé.

En notant $I'_{2D}(\mathbf{x}, t)$ la différence d'images $p(I(\mathbf{X}, t - 1)) - I_{2D}(\mathbf{x}, t)$, l'équation (5) s'écrit plus simplement :

$$p(\nabla I(\mathbf{X}, t - 1) \cdot \mathbf{W}) + I'_{2D}(\mathbf{x}, t) = 0 \quad (6)$$

Cette équation est de type flot optique dans laquelle le gradient temporel de l'image est remplacé par une différence d'image projetée.

Notre objectif est de résoudre cette équation pour chaque structure 3D, $I(\mathbf{X}, t - 1)$, et chaque différence d'images, $I'_{2D}(\mathbf{x}, t)$, afin de reconstruire la séquence de proche en proche.

4 Résolution directe par approximation constante par morceaux

Le principe consiste à résoudre l'équation (6) sans recourir à l'ajout d'une contrainte de régularisation supplémentaire sur le champ de vitesse recherché. Le modèle ajuste, dans chaque voisinage spatial Ω_l du point considéré (voir figure 3), les contraintes locales du premier ordre à une valeur fixe de \mathbf{W}_l . Cela revient à résoudre l'équation (6) en tout point de l'image volumique en supposant que le champ soit constant en tout point d'un voisinage donné centré autour de ce point [9]. Ainsi, pour estimer le champ de vitesse en une position donnée de la structure 3D, il suffit de définir un voisinage Ω_l cubique de côté n et de résoudre localement l'équation (6) sous l'hypothèse que le champ de vecteur vitesse est égal à une valeur fixe \mathbf{W}_l sur Ω_l . Cela se traduit par l'équation suivante valable sur le sous-domaine Ω_l :

$$E(\mathbf{W}_l) = \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} \|p_k(\nabla I \cdot \mathbf{W}_l)(i, j) + I'_{2D}(i, j)\|^2 \quad (7)$$

où \mathbf{W}_l est un vecteur constant dans Ω_l , J_k la restriction de $I(\mathbf{X}, t - 1)$ au voisinage considéré de côté n et p_k la restriction de la projection p aux plans de ce voisinage, définie par :

$$p_k(I)(i, j) = \sum_{m=k-\frac{n}{2}}^{m=k+\frac{n}{2}} a_m J_k(i, j, m), \quad (8)$$

Les coefficients a_m proviennent de la formulation discrète de la projection p donnée, dans le cas continu, par l'équation (1) à la section 3. Ainsi on peut écrire matriciellement E sous la forme :

$$E(W) = \|PAW - B\|^2, \quad (9)$$

Quatre éléments agissent dans l'énergie E : P , A , W et B se rapportant respectivement à la projection, la fenêtre cubique des gradients spatiaux, le vecteur vitesse et la composante temporelle. $P(n \times n, n \times n \times n)$ est une matrice bande (voir équation (10) à la page suivante) dont les éléments dépendent des coefficients a_k de la restriction de la projection p_k au voisinage considéré.

$$P = \begin{pmatrix} a_{k-\frac{n}{2}} \dots a_{k+\frac{n}{2}} & 0 & \dots & 0 \\ 0 \dots \dots \dots 0 & a_{k-\frac{n}{2}} \dots a_{k+\frac{n}{2}} & 0 & \dots & 0 \\ \vdots & 0 \dots \dots \dots 0 & \ddots & & 0 \\ \vdots & \dots & \dots & 0 & a_{k-\frac{n}{2}} \dots a_{k+\frac{n}{2}} \end{pmatrix} \quad (10)$$

Pour un voisinage centré au pixel P de coordonnée (i, j, k) , chaque colonne de A (voir équation (12)) représente une suite ordonnée de pixels obtenue en parcourant d'abord les restrictions des gradients spatiaux I_x , I_y et I_z à ce voisinage en profondeur puis en hauteur et finalement en largeur. Le voisinage Ω_l est représenté schématiquement à la figure 3. Rappelons que :

$$\nabla I = (I_x, I_y, I_z)^T \quad (11)$$

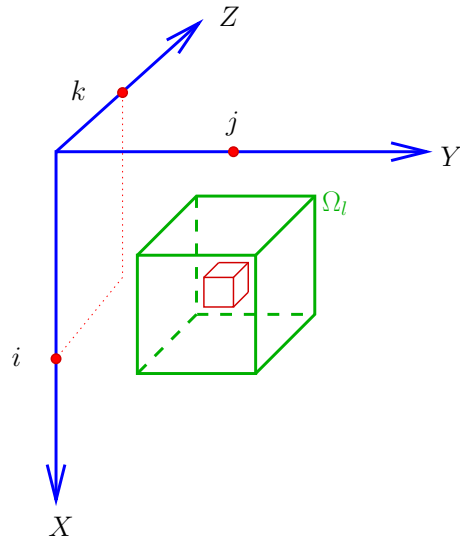


FIG. 3 – Représentation schématique du voisinage Ω_l .

$$A(n \times n \times n, 3) = \begin{pmatrix} \nabla I(i - \frac{n}{2}, j - \frac{n}{2}, k - \frac{n}{2}) \\ \nabla I(i - \frac{n}{2}, j - \frac{n}{2}, k - \frac{n}{2} + 1) \\ \vdots \\ \nabla I(i - \frac{n}{2}, j - \frac{n}{2}, k + \frac{n}{2}) \\ \nabla I(i - \frac{n}{2} + 1, j - \frac{n}{2}, k - \frac{n}{2}) \\ \vdots \\ \nabla I(i - \frac{n}{2} + 1, j - \frac{n}{2}, k + \frac{n}{2}) \\ \vdots \\ \vdots \\ \nabla I(i + \frac{n}{2}, j - \frac{n}{2}, k + \frac{n}{2}) \\ \nabla I(i + \frac{n}{2}, j - \frac{n}{2} + 1, k - \frac{n}{2}) \\ \vdots \\ \nabla I(i + \frac{n}{2}, j - \frac{n}{2} + 1, k + \frac{n}{2}) \\ \vdots \\ \vdots \\ \vdots \\ \nabla I(i + \frac{n}{2}, j + \frac{n}{2}, k + \frac{n}{2}) \end{pmatrix} \quad (12)$$

$W_l(3, 1)$ est un vecteur constant à trois composantes, soit :

$$W_l = \begin{pmatrix} u_l \\ v_l \\ w_l \end{pmatrix}$$

Et $B(n \times n)$ est un vecteur représentant une suite ordonnée des pixels composant le voisinage carré de côté n considéré dans la différence d'images I'_{2D} :

$$B = \begin{pmatrix} I'_{2D}(i - \frac{n}{2}, j - \frac{n}{2}) \\ I'_{2D}(i - \frac{n}{2}, j - \frac{n}{2} + 1) \\ \vdots \\ I'_{2D}(i - \frac{n}{2}, j + \frac{n}{2}) \\ I'_{2D}(i - \frac{n}{2} + 1, j - \frac{n}{2}) \\ \vdots \\ I'_{2D}(i - \frac{n}{2} + 1, j + \frac{n}{2}) \\ \vdots \\ \vdots \\ I'_{2D}(i + \frac{n}{2}, j + \frac{n}{2}) \end{pmatrix}$$

Le problème de la minimisation de (9) correspond donc simplement à une régression linéaire. Dans ce cas, la solution est directe et donnée par la méthode des moindres carrés [2] :

$$W_l = [A'^T A']^{-1} A'^T B, \quad (13)$$

Avec $A' = PA$ et A'^T la transposée de A' . La matrice produit $A'(n \times n, 3)$ a la forme analytique suivante :

$$A' = \begin{pmatrix} \sum_{m=k-\frac{n}{2}}^{m=k+\frac{n}{2}} a_m \nabla I(i - \frac{n}{2}, j - \frac{n}{2}, m) \\ \sum_{m=k-\frac{n}{2}}^{m=k+\frac{n}{2}} a_m \nabla I(i - \frac{n}{2}, j - \frac{n}{2} + 1, m) \\ \vdots \\ \sum_{m=k-\frac{n}{2}}^{l=k+\frac{n}{2}} a_m \nabla I(i - \frac{n}{2}, j + \frac{n}{2}, m) \\ \vdots \\ \sum_{m=k-\frac{n}{2}}^{m=k+\frac{n}{2}} a_m \nabla I(i + \frac{n}{2}, j + \frac{n}{2}, m) \end{pmatrix}$$

L'équation (9) peut être résolue si la matrice $A'^T A'$ n'est pas singulière, ce qui n'est *a priori* jamais le cas sauf pour les zones où la distribution de l'intensité est constante. $A'^T A'$ est la matrice de taille 3×3 symétrique définie par l'expression donnée par l'équation (14) à la page suivante.

$$A'^T A' = \begin{pmatrix} \sum_{n \times n} p_k(I_x)^2 & \sum_{n \times n} p_k(I_x)p_k(I_y) & \sum_{n \times n} p_k(I_x)p_k(I_z) \\ \sum_{n \times n} p_k(I_x)p_k(I_y) & \sum_{n \times n} p_k(I_y)^2 & \sum_{n \times n} p_k(I_y)p_k(I_z) \\ \sum_{n \times n} p_k(I_x)p_k(I_z) & \sum_{n \times n} p_k(I_y)p_k(I_z) & \sum_{n \times n} p_k(I_z)^2 \end{pmatrix} \quad (14)$$

Cette méthode permet en principe de fournir un champ dense de vecteur vitesse puisque en chaque point l de l'image, on peut calculer un vecteur vitesse par inversion de l'équation du flot optique sur le voisinage Ω_l . En pratique, on ne procède pas ainsi car les coûts algorithmiques seraient trop importants. Il est préférable de calculer le champ de vitesse sur une grille sous-échantillonnée. On peut alors

faire se chevaucher les voisinages Ω_l pour contrôler dans une certaine mesure la régularité spatiale du champ des vitesses.

5 Approche variationnelle

Une seconde méthode proposée est d'estimer un champ dense par formulation variationnelle [6]. Afin de résoudre l'équation (6), nous construisons une énergie dont le minimum fournit une estimation de la solution. Le terme de coût lié à cette contrainte d'attache aux données est :

$$E_1(\mathbf{W}) = \iint_{\mathbb{R}^2} (p(\nabla I(\mathbf{X}, t - 1) \cdot \mathbf{W}) + I'_{2D}(\mathbf{x}, t))^2 dx dy$$

En introduisant les gradients spatiaux de $I(\mathbf{X}, t - 1)$ dans les trois directions (I_x , I_y et I_z) (11), cette équation s'écrit :

$$E_1(\mathbf{W}) = \iint_{\mathbb{R}^2} (p(I_x u + I_y v + I_z w) + I'_{2D}(\mathbf{x}, t))^2 dx dy \quad (15)$$

La solution du problème de la minimisation de (15) est alors donnée par un zéro des équations d'Euler-Lagrange associées, c'est-à-dire $\frac{\partial E_1}{\partial \mathbf{W}} = 0$. La quantité $\frac{\partial E_1}{\partial \mathbf{W}}$ est la différentielle de la fonctionnelle E_1 . Cette fonctionnelle est convexe, donc les équations d'Euler-Lagrange associées ont une solution unique qui correspond à un minimum global de (15). Malheureusement l'équation (6) est sous dimensionnée car le vecteur \mathbf{W} possède 3 composantes. Nous ajoutons, dans la fonctionnelle, un terme régularisant qui contraint les variations spatiales de \mathbf{W} :

$$E_2(\mathbf{W}) = \iiint_{\mathbb{R}^3} \left(\|\nabla u\|^2 + \|\nabla v\|^2 + \|\nabla w\|^2 \right) dx dy dz$$

Finalement, nous minimisons l'énergie totale en \mathbf{W} :

$$E(\mathbf{W}) = E_1(\mathbf{W}) + \alpha E_2(\mathbf{W}), \quad (16)$$

où α est un coefficient pondérant les effets de la régularisation. Nous avons maintenant un problème bien posé mais la nouvelle fonctionnelle n'est plus convexe. Cela implique que la solution obtenue par calcul d'un zéro de la différentielle de E correspond à un minimum local.

Nous devons maintenant calculer explicitement la différentielle de E . Pour cela nous utilisons la dérivée directionnelle définie par la formule de Gâteaux :

Définition 1 (Dérivée de Gâteau) La dérivée d'une fonctionnelle E en un point u dans la direction g est donnée par :

$$\lim_{\lambda \rightarrow 0} \frac{F(u + \lambda g) - F(u)}{\lambda}$$

Si cette quantité existe, c'est une forme linéaire continue en g et nous la notons $\frac{\partial F}{\partial u}(g)$.

Propriété 1 Dans ce cas, le théorème de représentation de Riesz établit que :

$$\frac{\partial F}{\partial u}(g) = \langle \frac{\partial F}{\partial u}, g \rangle_{L^1}, \quad (17)$$

où $\langle \cdot, \cdot \rangle_{L^1}$ est le produit scalaire associé à l'espace L^1 des fonctions intégrables.

Le terme $\frac{\partial F}{\partial u}$ est la différentielle recherchée. Nous devons donc écrire les dérivées directionnelles sous la forme d'un produit scalaire. Comme \mathbf{W} est une fonction à valeurs dans \mathbb{R}^3 , nous effectuons d'abord le calcul des dérivées de E_1 par rapport à u , v et w fixes, suivant une direction g quelconque. En d'autres termes, il s'agit de calculer :

$$\lim_{\lambda \rightarrow 0} \frac{E_1(u + \lambda g) - E_1(u)}{\lambda} \quad (18)$$

Nous développons alors l'expression $E_1(u + \lambda g, v, w) - E_1(u, v, w)$ et obtenons après simplification :

$$\begin{aligned} E_1(u + \lambda g, v, w) - E_1(u, v, w) &= \\ \lambda^2 \iint p(I_x g)^2 dx dy &+ \\ 2\lambda \iint p(I_x g)[p(I_x u + I_y v + I_z w) + I'_{2D}] dx dy & \end{aligned} \quad (19)$$

En divisant par λ et en passant à la limite, nous obtenons :

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \frac{E_1(u + \lambda g) - E_1(u)}{\lambda} &= \\ 2 \iint p(I_x g)[p(I_x u + I_y v + I_z w) + I'_{2D}] dx dy & \end{aligned} \quad (20)$$

À cause de la présence de la fonction de projection p , le second membre de l'égalité (20) n'est pas sous la forme du produit scalaire défini par l'expression (17).

Nous remplaçons, dans l'équation (20), p par son expression générale dans le cas continu 1. Injecté dans le premier membre de l'égalité (20), nous obtenons :

$$\iint p(I_x g)(p(I_x u + I_y v + I_z w) + I'_{2D}) dx dy = \iint \left[\int_z a(z) I_x g dz \right] (p(I_x u + I_y v + I_z w) + I'_{2D}) dx dy \quad (21)$$

La fonction $\psi(x, y) = p(I_x u + I_y v + I_z w) + I'_{2D}$ dépend uniquement des variables x et y . En d'autres termes, à cause de la présence de la fonction de projection p , cette fonction est indépendante de la variable z . D'après le théorème de Fubini, nous avons :

$$\iint p(I_x g) \psi(x, y) dx dy = \iiint a(z) I_x g(x, y, z) \cdot \psi(x, y) dz dx dy \quad (22)$$

En posant le changement de mesure : $a_z dz = d\mu(z)$, l'égalité (22) devient :

$$\iint p(I_x g) \psi(x, y) dx dy = \iiint I_x g(x, y, z) \psi(x, y) dx dy d\mu(z) \quad (23)$$

Donc :

$$\lim_{\lambda \rightarrow 0} \frac{E_1(u + \lambda g) - E_1(u)}{\lambda} = 2 \iiint I_x g [p(I_x u + I_y v + I_z w) + I'_{2D}] dx dy d\mu(z) \quad (24)$$

Nous remarquons alors que cette dernière expression représente bien le produit scalaire entre deux fonctions données définies en (17). Ainsi :

$$\lim_{\lambda \rightarrow 0} \frac{E_1(u + \lambda g) - E_1(u)}{\lambda} = \langle g, 2I_x [p(I_x u + I_y v + I_z w) + I'_{2D}] \rangle \quad (25)$$

D'après la propriété 1, la dérivée de Gâteau par rapport à u est donc :

$$\frac{\partial E_1}{\partial u} = 2I_x [p(I_x u + I_y v + I_z w) + I'_{2D}] \quad (26)$$

D'une manière analogue, les différentielles de E_1 en v et w sont :

$$\begin{aligned}\frac{\partial E_1}{\partial v} &= 2I_y [p(I_x u + I_y v + I_z w) + I'_{2D}] \\ \frac{\partial E_1}{\partial w} &= 2I_z [p(I_x u + I_y v + I_z w) + I'_{2D}]\end{aligned}$$

Nous ne donnons pas les détails du calcul des gradients du terme de régularisation E_2 qui est classique, le lecteur pourra consulter [4, 12]. Ce calcul donne le système d'équations suivant :

$$\begin{aligned}\frac{\partial E_2}{\partial u} &= -2\alpha\Delta u \\ \frac{\partial E_2}{\partial v} &= -2\alpha\Delta v \\ \frac{\partial E_2}{\partial w} &= -2\alpha\Delta w\end{aligned}$$

Nous écrivons ensuite que les trois composantes de $\frac{\partial E}{\partial \mathbf{W}}$ sont nulles pour aboutir au système d'équations suivant :

$$\begin{aligned}I_x [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta u &= 0 \\ I_y [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta v &= 0 \\ I_z [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta w &= 0\end{aligned}\tag{27}$$

Afin de converger vers un minimum local, nous utilisons les équations d'évolution associées à celles du système (27). Nous sommes toutefois assurés d'atteindre le minimum global que si la condition initiale est suffisamment proche de la solution. Les équations d'évolution sont alors équivalentes à une méthode de descente dans la direction du gradient :

$$\begin{aligned}\frac{du}{dt} + I_x [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta u &= 0 \\ \frac{dv}{dt} + I_y [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta v &= 0 \\ \frac{dw}{dt} + I_z [p(I_x u + I_y v + I_z w) + I'_{2D}] - \alpha\Delta w &= 0\end{aligned}$$

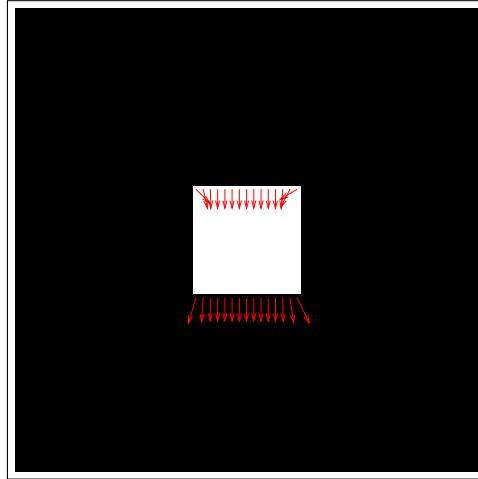
6 Expérimentation

Afin de valider l'étude théorique concernant les approches ne tenant pas compte d'un modèle de données bien déterminé, nous construisons un lot de données volumiques. Nous considérons une séquence 3D+t simple contenant un cube de couleur homogène qui se déplace vers le bas et en profondeur, vers l'intérieur. Nous simulons la projection perspective produisant la séquence vidéo 2D+t intermédiaire. Considérons maintenant la première image 3D I^0 obtenue et la première projection $I_{2D}(\mathbf{x}, t_1)$. Ces deux structures nous permettent de calculer le champ de déplacements entre les deux structures et de restituer la donnée 3D à l'instant t_1 . Nous estimons ce champ de déplacement à l'aide des deux méthodes présentées dans cet article. Afin de visualiser le résultat de calcul des champs de déplacements, nous représentons une section de l'image volumique avec la restriction des champs calculés aux composantes u et v (voir figure 4). Nous rappelons que les composantes u et v contiennent des informations relatives aux déplacements dans la direction respectivement de X et de Y . Enfin, les gradients spatio-temporels qui constituent nos données d'entrée sont calculés par convolution sur un noyau gaussien 3D dérivé respectivement dans les trois directions. La variance du noyau est fixée à 1 ce qui correspond à un lissage modéré.

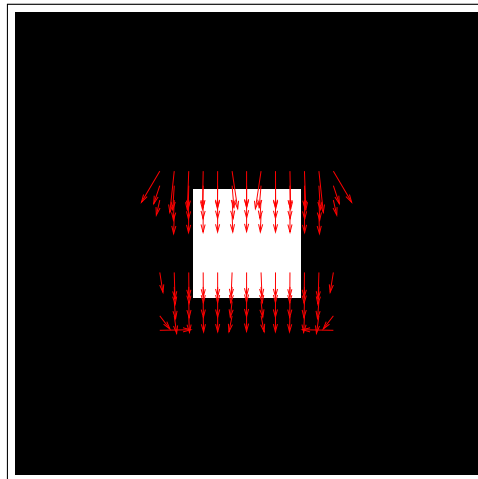
D'après la figure 4, le champ de déplacements généré par la méthode différentielle globale est lisse. De plus, les déplacements sont localisés sur les frontières du cube en mouvement. Néanmoins, les déplacements estimés par la méthode locale s'étalent sur une région plus large. Cela est dû au fait que l'estimation locale du mouvement sur un voisinage de taille donnée produit un champ continu par morceaux. Plus la taille du voisinage considéré est grande, meilleure est la qualité de reconstruction, mais plus lente est la résolution. Nous abordons, par ailleurs, un cas d'occultation d'objets en profondeur. Nous considérons une image 3D contenant deux cubes qui se recouvrent lors de la projection 2D. D'après la figure 5, nos deux méthodes estiment des champs corrects de point de vue amplitude et directions.

7 Conclusion

Nous avons présenté dans cette étude deux approches possibles pour résoudre la problématique de la reconstruction 3D+t n'impliquant pas de connaissances *a priori* sur le modèle de données. Le principe est de résoudre une équation de transport de la luminosité en chaque voxel, en établissant des hypothèses sur la



(a) Champ de vitesses estimé à l'aide de la méthode différentielle globale, $\alpha = 100$, le nombre d'itérations est de 50.



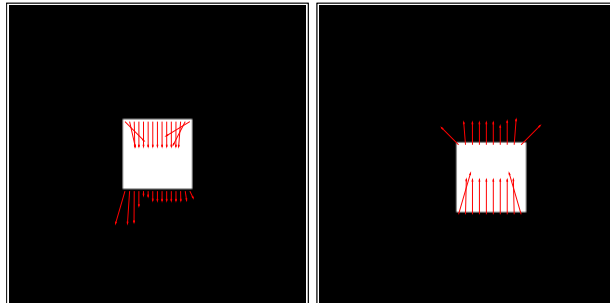
(b) Champ de vitesses estimé à l'aide de la méthode locale, testé sur un voisinage cubique de côté 5 pixels.

FIG. 4 – Représentation des composantes u et v des champs de vitesse estimés, de gauche à droite, par la méthode variationnelle (a) et la méthode directe (b) sur une section d'une image volumique comprenant un cube qui se déplace de haut vers le bas et vers l'intérieur.

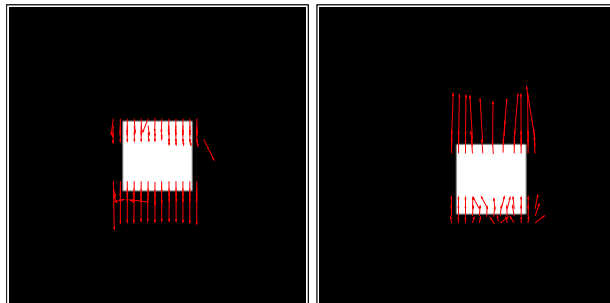


(a) Coupe de l'image 3D I^0 à deux profondeurs montrant les deux carrés en déplacement.

(b) Image 2D $I_{2D}(\mathbf{x}, t_1)$, projection à l'instant t_1 .



(c) Champ de vitesse estimé à l'aide de la méthode différentielle, $\alpha = 100$, le nombre d'itérations est de 50.



(d) Champ de vitesse estimé à l'aide de la méthode directe, testé sur un voisinage cubique de côté 5 pixels.

FIG. 5 – Représentation des composantes u et v des champs de vitesse estimés, de haut en bas, par la méthode variationnelle et la méthode directe sur deux sections d'une image volumique comprenant deux cubes qui se déplacent respectivement vers le bas et vers le haut.

nature du mouvement à calculer. La méthode différentielle consiste à contraindre spatialement le champ de déplacement estimé de proche en proche. Cette méthode génère un champ dense ayant un aspect visuel lisse satisfaisant. La méthode directe consiste à résoudre l'équation de conservation de la luminosité en supposant que le champ de vitesse soit constant sur un voisinage de taille donnée. Cette méthode fournit un champ continu par morceaux. Plus la taille du voisinage considéré pour la résolution de cette équation est élevée, meilleure est la qualité de reconstruction, mais aussi, plus le temps d'exécution de l'algorithme associé est important. Ce dernier, comparé à la résolution itérative mise en œuvre dans la méthode différentielle, présente l'avantage d'être plus simple. En évitant le lissage généré par la régularisation, pouvant être une contrainte trop forte sur les données, il peut s'avérer plus adapté au contexte des données biologiques. Néanmoins, l'exécution de cet algorithme peut induire des erreurs d'estimation des déplacements car elle repose sur l'inversion de matrices dépendant des gradients de l'image. En particulier la méthode ne fonctionne pas sur des zones homogènes. Cette difficulté peut être levée en excluant du calcul les zones de pixels de faible norme du gradient spatial. Nous avons montré que la reconstruction 3D est possible par calcul du flot optique. Évidemment, ces approches souffrent des limitations connues de la formulation relative à la contrainte du flot optique. En l'occurrence, l'estimation des larges déplacements n'est pas adapté. Cette formulation a aussi une faible robustesse vis à vis des gradients spatio-temporels servant de données d'entrée. De plus, la régularisation L^2 mise en œuvre dans l'approche variationnelle peut être très contraignante. Néanmoins, il est parfaitement possible d'adapter toutes les variantes qui existent autour de la résolution du flot optique qui permettent de résoudre les problèmes que nous venons de citer.

Références

- [1] P.J. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE PAMI*, 14(2) :239–256, Mar. 1992.
- [2] R. Burden and L. Fairies. *Numerical Analysis*. Prindle, Weber & Schmidt, 1985.
- [3] G. Demoment. Image reconstruction and resotration : overview of common estimation structures and problems. *IEEE Trans. on ASSP*, 37(12) :2024–2036, Dec. 1989.

- [4] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Series in computational physics, New York, springer edition, 1984.
- [5] A. Gueziec, P. Kazanzides, B. Williamson, and R.H. Taylor. Anatomy-based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot. *IEEE Trans Med Imaging*, 17 :715–728, 1998.
- [6] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 :185–203, 1981.
- [7] L. Huei-Yung. *Computer vision techniques for complete 3D model reconstruction*. PhD thesis, State university of New York at Stony Brook, 2002.
- [8] L. Lemieux, R. Jagoe, D.R. Fish, N.D. Kitchen, and D.G. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Med Phys*, 21 :1749–1760, 1994.
- [9] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–697, Canada, 1981.
- [10] J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1), 1998.
- [11] F. Natterer. *The mathematics of computerized tomography*. Wiley, New York, 1986.
- [12] W. Rekik. *Fusion de données temporelles, ou 2D+t, et spatiales, ou 3D, pour la reconstruction de scènes 3D+t et traitement d’images sphériques. Applications à la biologie cellulaire*. PhD thesis, Université Pierre Et Marie Curie, Paris 6, Département Calcul Scientifique, Laboratoire d’Informatique de Paris 6 (LIP6), février 2007.
- [13] W. Rekik, D. Béréziat, and S. Dubuisson. 3D+t reconstruction in the context of locally spheric shaped data observation. In *12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria, 2007. Accepted for publication.
- [14] P. Viola and W. Wells. alignment by maximization of mutual information. *International Journal of Computer Vision*, 24 :137–154, 1997.