HAL
open science

# Flot de scène à partir d'images couleur et de cartes de profondeur

Antoine Letouzey, Benjamin Petit, Edmond Boyer

## To cite this version:

Antoine Letouzey, Benjamin Petit, Edmond Boyer. Flot de scène à partir d'images couleur et de cartes de profondeur. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656484v1

**HAL Id: hal-00656484**

**https://hal.science/hal-00656484v1**

Submitted on 17 Jan 2012 (v1), last revised 19 Jan 2012 (v2)

**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.

# Flot de scène à partir d'images couleur et de cartes de profondeur

Antoine Letouzey Benjamin Petit Edmond Boyer

INRIA Grenoble Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot S$^{t}$-Ismier
prenom.nom@inria.fr

## Résumé

*Dans cet article nous traitons le problème de l'estimation du flot de scène 3D à partir de plusieurs caméras. En particulier, nous considérons le cas où une caméra de profondeur est associée à une ou plusieurs caméras couleur. Il s'agit d'un cas de figure commun avec l'avènement des capteurs hybrides tels que la caméra Kinect. Dans ce cas, les informations géométriques provenant des cartes de profondeur peuvent être combinées aux variations d'intensité dans les images couleur pour l'estimation d'un champ de déplacement 3D dense et régulier. Nous proposons pour cela un cadre unifié capable de gérer des déplacements arbitrairement grands tout en conservant une précision sous-pixélique. L'estimation est formulée sous la forme d'un problème d'estimation linéaire qui peut être résolu de manière très efficace. La contribution principale réside dans l'utilisation conjointe d'une carte de profondeur et d'images couleur pour estimer le flot de scène. La carte de profondeur s'avère très utile dans ce cas puisqu'elle définie une surface 3D sur laquelle les contraintes photométriques peuvent être intégrées de manière cohérente. Les expérimentations menées à la fois sur des données de synthèse et des données réelles fournissent des résultats qualitatifs et quantitatifs qui démontrent l'intérêt et la faisabilité de l'approche proposée.*

## Mots Clef

Surface, carte de profondeur, flot de scène, déplacement 3D.

## Abstract

*In this paper we consider the problem of estimating a 3D motion field using multiple cameras. In particular, we focus on the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The estimation is cast as a linear optimization problem that can be solved very efficiently. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrate the interest of the approach.*

## Keywords

Surface, depth map, scene flow, 3D motion.

# 1 Introduction

Le déplacement est une source d'information importante lors de l'analyse et de l'interprétation de scènes dynamiques. Il fournit une information riche et discriminante sur les objets qui composent la scène et est utilisé, par exemple, dans les systèmes de vision humaine et artificielle pour suivre et délimiter ces objets. L'intérêt apparaît surtout dans le cas d'applications interactives, telles que les jeux vidéos ou les environnements intelligents, pour lesquels le mouvement est une source d'information primordiale dans la boucle perception-action. Ces applications utilisent souvent des caméras de profondeur, par exemple des caméras à temps de vol ou à lumière structurée, qui fournissent directement une information 3D, sans recourir à un traitement multi-vue additionnel. Dans cet article nous proposons de calculer un flot de scène dense à partir de telles caméras.

Les capteurs actifs ou les systèmes de vision basés marqueurs peuvent fournir directement un ensemble épars d'informations de déplacement sur des scènes en mouvement. Les méthodes non-invasives considèrent traditionnellement des images d'intensité et leurs variations dans le temps pour obtenir une estimation dense du champ de déplacement. Dans le cas monoculaire, les projections en 2D de tels champs de déplacement sont estimés via le *flot optique* [6, 10]. Quand plusieurs caméras sont disponibles, l'intégration sur plusieurs points de vue permet l'estimation d'un champ de déplacement 3D, le

*flot de scène* [15, 11]. Dans les deux cas, 2D et 3D, les variations d'intensité seules ne suffisent cependant pas pour estimer le déplacement, et de nouvelles contraintes doivent être introduites, dans la plupart des cas il s'agit d'une contrainte de régularité. De ce point de vue, les caméras de profondeur ont l'avantage de fournir une information géométrique très utile à partir de laquelle des contraintes cohérentes de régularité 3D peuvent être déduites. Cependant peu de travaux ont été menés dans le but d'intégrer des caméras de profondeur dans l'estimation de champs de déplacement. C'est notre objectif dans cet article. Plus particulièrement, nous étudions une manière de combiner des informations de variation d'intensité avec des informations de profondeur pour en déduire une estimation des déplacements 3D d'une scène. Dans ce but, nous étendons les travaux récemment proposés par [12] pour le cas d'un système multi-caméra couleur à celui d'un système disposant de caméras de profondeur. La méthode proposée permet l'estimation des déplacements sans nécessiter de correspondances spatiales ou temporelles comme c'est le cas pour le suivi d'objet, même si de telles correspondances peuvent être introduites pour améliorer l'estimation. Comme démontré dans cet article, cela aboutit à un schéma simple et efficace pour obtenir une information de mouvement 3D dense et instantanée sur des scènes dynamiques en utilisant des caméras de profondeur.

L'article est organisé comme suit : dans la section 2 nous présentons un état de l'art. Nous introduisons la méthode proposée dans la section 3 ainsi que son implémentation dans la section 4. La section 5 contient les évaluations quantitatives et qualitatives de notre approche. Enfin nous concluons et discutons notre travail dans la section 6.

# 2 État de l'art

Un grand nombre de travaux ont été menés dans le but d'estimer des champs de déplacement en utilisant des informations photométriques. Les premiers travaux dans ce domaine se concentraient sur le champ de déplacement entre deux images consécutives. L'estimation du flot optique par [6, 1] fait appel aux contraintes de flot normal dérivées des variations d'intensité dans les images. Lorsque l'information vient d'images stéréo, le champ de déplacement 3D peut être calculé. Cette estimation peut être faite en utilisant une information de disparité connue [15] ou en combinaison avec l'estimation de cette disparité [7, 8]. De plus, on peut faire l'hypothèse de cohérence temporelle [14]. Dans le cas de flux d'images multiples, il est possible d'estimer à la fois la structure et le mouvement [13, 2].

Plus proche de la configuration considérée dans cet article, certains travaux existant considèrent en données d'entrée la structure 3D de la scène ainsi que les flux d'images d'une ou plusieurs caméras. Le flot de scène peut ainsi être obtenu en reprojetant le flot optique calculé dans les images sur la structure 3D [15] ou en intégrant directement les contraintes de flot normal sur la forme [11]. Nous nous basons sur le travail présenté dans [12] qui fournit un cadre où plusieurs types de contraintes photométriques, flot de normal et points d'intérêt, peuvent être combinés avec un modèle de déformation de surface favorisant la rigidité locale. Nous étendons ces travaux dans le cas où l'on dispose d'une caméra de profondeur, fournissant ainsi une information directe sur la structure de la scène. À notre connaissance, il s'agit des premiers travaux qui tentent d'estimer un champ de déplacement en combinant les informations d'une caméra de profondeur avec des variations temporelles d'intensité dans les images de caméras couleur.

# 3 La méthode

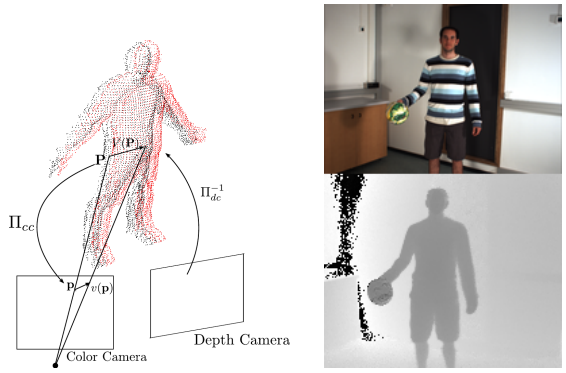
L'approche proposée estime directement un champ de mouvement 3D sur la surface en utilisant des contraintes photométriques 2D. Pour cela, elle prend en entrée des flux d'images couleur et de profondeur venant d'un ensemble de caméras pré-étalonnées et synchronisées. La configuration prise en compte se compose d'une caméra de profondeur et d'une ou plusieurs caméras couleur. Dans la suite, pour des raisons de simplicité, nous ne détaillons que le cas disposant d'une caméra couleur. Néanmoins l'extension à plusieurs caméras couleur est directe et sera expliquée plus loin dans l'article. Contrairement aux travaux précédents sur le flot de scène [12] notre méthode ne nécessite pas la reconstruction d'un modèle de la surface, mais utilise plutôt la carte de profondeur et peut donc fonctionner sur une configuration beaucoup plus simple ne disposant que d'un capteur couleur et un capteur de profondeur, par exemple la caméra Kinect.

## 3.1 Notations

À chaque instant de temps $t$, une image couleur $I_t$ et une carte de profondeur $D_t$ sont acquises. On suppose que les caméras sont étalonnées et que les matrices de projection $\Pi_{cc}$ et $\Pi_{dc} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ font correspondre des points 3D $\mathbf{P} = (x, y, z)^T \in \mathbb{R}^3$ à leur correspondant 2D $\mathbf{p} = (u, v)^T \in \mathbb{R}^2$ dans les images couleur et de profondeur, respectivement. La carte de profondeur fournit un nuage de points 3D à partir duquel une surface maillée $S^t$ peut être construite en utilisant la connectivité dans l'image de profondeur. C'est à dire que chaque pixel devient un sommet du maillage en 3D, connecté à ses voisins dans l'image.

À partir de ces images, la méthode proposée estime un champ de déplacement 3D dense $V^t$ entre les instants $t$ et $t + 1$. Cela correspond à un ensemble de vecteurs de déplacement instantané $V^t(\mathbf{P})$ pour chaque point de $S^t$ (voir figure 1). Ainsi, le flot optique $v^t$ est la projection du flot de scène estimé $V^t$ sur les plans images des caméras couleur.

Pour estimer le flot 3D $V^t(\mathbf{P})$, le problème est formulé sous la forme d'une optimisation où un terme d'attache aux données renforçant les contraintes photométriques est associé à un terme de lissage favorisant un champ de dé-




(a) (b)

FIG. 1: (a) Projection de la surface sur le plan image, et (b) images de couleur et de profondeur.

placement régulier :

$$\mathbf{E} = \mathbf{E}_{data} + \mathbf{E}_{smooth}. \tag{1}$$

Le terme d'attache aux données contrôle à la fois les grands et petits déplacements tandis que le terme de lissage impose un modèle de déformation avec des contraintes de rigidité locale.

## 3.2 Contraintes photométriques

Comme suggéré par Xu *et al.* dans leur travaux sur le flot optique [17], l'utilisation de deux types d'information photométrique distincts permet de gérer correctement les grands et les faibles déplacements. Chacune de ces contraintes ajoute un nouvel élément dans le terme d'attache aux données de la fonctionnelle d'énergie (1).

**Flot normal dense.** Dans la lignée des méthodes traditionnelles d'estimation du flot optique, nous faisons l'hypothèse que l'illumination reste constante entre deux trames successives. C'est à dire que deux projections successives d'un même point de la scène possèdent la même intensité lumineuse dans les images. Bien que discutable, cette hypothèse semble raisonnable dans notre cas où les surfaces sont souvent Lambertienne et où l'illumination est approximativement constante pour de faibles déplacements.
Pour de petits déplacements nous pouvons écrire l'équation de flot normal suivante [1] pour chaque pixel de $I^t$ où la surface $S^t$ est visible :

$$\nabla I^t.v^t + \frac{\mathrm{d}I^t}{\mathrm{d}t} = 0 \ ,$$

ou dans le cas de déplacements 3D :

$$\nabla I^t.\left[J_\Pi V^t\right] + \frac{\mathrm{d}I^t}{\mathrm{d}t} = 0 \tag{2}$$

où le déplacement 3D associé à un point de la scène est relié au déplacement de sa projection par la matrice Jacobienne $2 \times 3$, $J_\Pi(\mathbf{p}) = \frac{\partial \mathbf{p}}{\partial \mathbf{P}}$. L'équation (2) fournit le terme d'attache aux données suivant dans l'équation (1) :

$$\mathbf{E}_{flow} = \int_{I^t} \|\nabla I^t.\left[J_\Pi V^t\right] + \frac{\mathrm{d}I^t}{\mathrm{d}t}\|^2 \, \mathrm{d}\mathbf{p}. \tag{3}$$

Le critère ci-dessus seul n'est pas suffisant puisque les contraintes reposent uniquement sur le déplacement 2D dans la direction normale au gradient dans les images. Le *problème de l'ouverture* en 2D se transpose en 3D [15]. De plus le domaine de validité de la Jacobienne ainsi que les différentes hypothèses imposent de traiter uniquement des petits déplacements. Ces limitations seront traitées dans la suite de l'article.

**Correspondances éparses de points d'intérêt.** Nous considérons un second type d'information photométrique pour rendre notre méthode robuste aux larges mouvements. Nous avons choisi d'utiliser un ensemble épars de points d'intérêt mis en correspondance entre deux images couleur successives, $I^t$ et $I^{t+1}$, pour guider l'estimation du champ de déplacement. De telles correspondances peuvent facilement être extraites en utilisant une des nombreuses méthodes de détection et d'appariement de points caractéristiques. On peut citer par exemple SIFT [9] ou SURF [3]. Ces points d'intérêt agissent comme des ancres pour l'estimation du déplacement. La confiance que nous accordons à ces ancres est critique puisque les erreurs introduites par de mauvais appariements sont propagées par l'étape de régularisation. Notre implémentation utilise le descripteur SIFT associé à un seuil très sélectif sur le score d'appariement (identique pour toutes les séquences traitées). Grâce à cette stratégie conservatrice, tous les faux appariements ont été supprimés automatiquement lors de nos expérimentations. Ces correspondances éparses contraignent directement le flot optique ; elles peuvent aussi bien contraindre le flot de scène en les re-projetant sur les surfaces 3D. Ainsi, chaque couple de points appariés donne une correspondance 3D-3D par reprojection sur les deux surfaces successives associées, et par conséquent une contrainte directe sur le champ de déplacement 3D, $V_f$. Le terme d'attache aux données associé s'écrit :

$$\mathbf{E}_{3D} = \sum_{F^t} \|V^t - V_f^t\|^2 \ , \tag{4}$$

où $F^t$ est l'ensemble des points 3D contraints correspondant aux projections des points d'intérêt 2D appariés.

## 3.3 Régularisation du champ de déplacement

Comme mentionné précédemment, les contraintes du flot normal ne sont pas suffisantes pour estimer les déplacements. En dépit d'apporter une information supplémentaire, l'utilisation de points d'intérêt 2D ne résulte qu'en l'ajout de contraintes locales. Dans le but d'obtenir une estimation dense du déplacement présent dans une scène, il est nécessaire d'introduire une notion de régularisation. Nous avons choisi de faire l'hypothèse d'un champ de déplacement régulier. Dans le cas du flot optique, il existe

deux stratégies principales de régularisation : locale ou globale. Leurs deux extensions au cas 3D pourraient être considérées mais nous avons choisi d'utiliser une méthode de régularisation globale, et plus précisément, l'extension de la méthode de Horn et Schunck [6]. La motivation de ce choix vient du fait que, bien que souvent fausse dans des images en 2D, surtout aux frontières des objets, une hypothèse de régularité a plus de sens en 3D. Ainsi nous définissons le terme de lissage suivant pour nous assurer d'une régularité globale du champ de déplacement estimé :

$$\mathbf{E}_{smooth} = \int_{S^t} \|\nabla V^t\|^2 \mathrm{d}\mathbf{P}\,. \tag{5}$$

Le terme de lissage Laplacien ci-dessus favorise une rigidité locale du champ de déplacement estimé. La contrainte de lissage peut être pondérée en utilisant les informations géométriques fournies par la carte de profondeur. Nous expliquerons plus loin comment cela permet de gérer convenablement les discontinuités de la surface.

# 4 Formulation et résolution

En regroupant tous les termes précédemment définis, notre fonctionelle d'énergie se réécrit comme ceci :

$$\mathbf{E} = \left[\lambda_{3D}^2 \mathbf{E}_{3D} + \lambda_{flow}^2 \mathbf{E}_{flow} + \lambda_{smooth}^2 \mathbf{E}_{smooth}\right]\,, \tag{6}$$

où les paramètres lambdas sont des valeurs scalaires servant à pondérer l'influence des différents termes. Minimiser cette équation peut se formuler de la manière suivante :

$$\begin{aligned} \arg\min_{V^t} \lambda_{flow}^2 \delta_{\overline{F^t}} \|\nabla I^t.\left[J_\Pi V^t\right] + \frac{\mathrm{d}I^t}{\mathrm{d}t}\|^2 \\ + \lambda_{3D}^2 \delta_{F^t} \|V^t - V_f^t\|^2 \\ + \lambda_{smooth}^2 \|\nabla V^t\|^2\,, \end{aligned} \tag{7}$$

où $\delta$ est le symbole de Kronecker indiquant que ce terme ne s'applique qu'à un sous ensemble de points et $\overline{F^t}$ indique les points de la surface pour lesquels aucune information venant des points d'intérêt n'est disponible.
En dérivant l'équation (7), nous obtenons pour chaque point $\mathbf{P}$ de la surface, l'équation d'Euler-Lagrange discrète, de la forme :

$$\mathbf{A_P} V_\mathbf{P} + \mathbf{b_P} - \Delta V_\mathbf{P} = 0\,, \tag{8}$$

où $\Delta$ est l'opérateur de Laplace-Beltrami normalisé sur la surface.

## 4.1 Système linéaire

Étant donné que l'équation (8) met en jeux un ensemble de contraintes linéaires pour chaque point 3D de la surface, une solution est donnée par la résolution du système linéaire suivant :

$$\begin{bmatrix}\mathbf{L}\\ \mathbf{A}\end{bmatrix} V^t + \begin{bmatrix}\mathbf{0}\\ \mathbf{b}\end{bmatrix} = \mathbf{0}\,, \tag{9}$$

où $\mathbf{L}$ est la matrice Laplacienne du maillage de la surface construit de telle manière que $\mathbf{L}(i,j)$ pondère la relation entre les points $i$ et $j$ (les poids du Laplacien sont discutés dans la section 4.3). $\mathbf{A}$ et $\mathbf{b}$ stockent toutes les contraintes visuelles sur le déplacement venant des termes d'attache aux données. Ce système linéaire est creux et peut être résolu en utilisant un solveur adapté tel que *Taucs*. L'équation (7) peut aussi être résolue de manière itérative en appliquant la méthode de Jacobi. De cette manière on résout le système indépendamment en chaque point en utilisant la solution courante du voisinage comme présenté dans la section suivante.

## 4.2 Résolution itérative

Inspiré par les travaux de Horn et Schunck [6], nous dérivons de l'équation (8) la résolution itérative suivante en chaque point de la scène :

$$\begin{aligned} v_x^{k+1} &= \overline{v}_x^k + A_x^x v_x^k + A_y^x v_y^k + A_z^x v_z^k - b_x \\ v_y^{k+1} &= \overline{v}_y^k + A_x^y v_x^k + A_y^y v_y^k + A_z^y v_z^k - b_y \\ v_z^{k+1} &= \overline{v}_z^k + A_x^z v_x^k + A_y^z v_y^k + A_z^z v_z^k - b_z \end{aligned} \tag{10}$$

où $(v_x, v_y, v_z)$ et $(\overline{v}_x, \overline{v}_y, \overline{v}_z)$ représentent respectivement le déplacement et le déplacement local moyen d'un point, l'indice $k$ représente l'itération courante et les $A_i^j$ et $b_i$ sont les éléments de la matrice $\mathbf{A}$ et du vecteur $\mathbf{b}$ de l'équation (8).
Nous pouvons remarquer que les équations (10) sont indépendantes, à une itération donnée, pour chaque point de la surface. Ainsi l'implémentation de la résolution peut être massivement parallélisée. Dans ces équations, le déplacement local moyen pour un point 3D est donné par le voisinage de ce point en respectant la connectivité de la surface discrétisée et est pondéré exponentiellement en utilisant la taille des arêtes. Ainsi nous renforçons la relation qui lie deux points proches tout en empêchant les points aux frontières des objets d'être perturbés par des points lointains.
Cette formulation permet une approximation rapide du champ de déplacement. La rapidité et la précision de la résolution dépend fortement d'une bonne initialisation. Comme mentionné dans [6], une bonne solution initiale peut être donnée par l'estimation obtenue à la trame précédente.

## 4.3 Détails d'implémentation

Cette section présente en détail les choix importants faits au moment de l'implémentation. Premièrement, nous discutons les poids qui déterminent, lors de la régularisation, l'influence du voisinage de la surface. Ensuite, nous présentons comment les grands et petits déplacements sont gérés séparément par le biais d'un algorithme en deux passes.

**Poids Laplaciens.** Dans le terme de lissage de l'équation (7), l'opérateur de Laplace-Beltrami $\nabla^2$ défini sur la surface de manière continue est approché par la matrice Laplacienne du graphe du maillage $\mathbf{L}$, c'est à

dire $\nabla^2 V^t = \mathbf{L}V^t$, où :

$$\mathbf{L}(i,j) = \begin{cases} \deg(P_i) & \text{si } i = j, \\ -w_{ij} & \text{si } i \neq j \text{ et } P_i \text{ est adjacent à } P_j, \\ 0 & \text{sinon,} \end{cases}$$

où les $w_{ij}$ correspondent aux poids des arêtes et $\deg(P_i) = \sum_{j \neq i} w_{ij}$. La matrice $\mathbf{L}$ peut être purement combinatoire, c'est à dire $w_{ij} \in \{0, 1\}$, ou contenir des poids $w_{ij} \geq 0$, par exemple poids cotangents, souvent utilisés en graphisme [16] avec un échantillonage uniforme. Dans le cadre de ce travail, la connectivité du maillage vient de celle de l'image de profondeur, c'est à dire que les points voisins dans l'image sont reliés sur la surface maillée par une arête. Cela aboutit à la construction d'un maillage cohérent, c'est à dire sans auto-intersection, mais avec potentiellement des arêtes de très grande taille correspondant aux discontinuités de la carte de profondeur. Pour gérer correctement ces discontinuités lors de la régularisation, nous proposons l'utilisation des poids suivants :

$$w_{ij} = -G(|P_i - P_j|, \sigma),$$

où $G$ est un noyau Gaussien, $|\cdot|$ est la distance Euclidienne et $\sigma$ l'écart type. En plus de fortement limiter la diffusion le long des grandes arêtes, les noyaux Gaussiens sont aussi préconisés par Belkin *et al.* [4] pour leur propriété de convergence vers le cas continu de l'opérateur de Laplace-Beltrami lorsque la résolution du maillage augmente.

**Algorithme en deux passes.** Comme mis en évidence précédemment, les hypothèses du flot normal sont uniquement valides à l'échelle des pixels et les contraintes qui en découlent ne sont donc plus utilisables pour de grands déplacements. Dans le cas du flot optique, la majorité des méthodes existantes reposent sur une approche multi-échelle pour répondre à cette limitation. Néanmoins, une telle stratégie n'est pas bien adaptée dans notre contexte, en effet elle nécessite de construire une pyramide d'images, conduisant à un lissage dans les images et, par là, à une régularisation en 2D que nous souhaitons justement éviter. La mise en place de l'algorithme en deux passes décrit ci-dessous permet de résoudre ce problème.

- Dans un premier temps nous ne prenons en compte que les correspondances éparses de points d'intérêt et réalisons une première estimation de $V^t$ en résolvant l'équation (7) avec $\lambda_{flow} = 0$. Cette étape nous permet de créer une version *déplacée* de la surface, $S^{t\prime}$, en appliquant la première estimation de $V^t$ à $S^t$. Ensuite nous effectuons une reprojection de la surface $S^{t\prime}$ dans la caméra couleur en la texturant à l'aide de $I^t$.
- Nous disposons maintenant de deux images couleur, celle de la surface reprojetée et $I^{t+1}$. Elle doivent à ce point être suffisamment similaires pour pouvoir utiliser les contraintes du flot normal. Les informations sur les grands et petits déplacements peuvent maintenant être combinées lors d'une seconde résolution de l'équation (7), avec cette fois-ci des valeurs identiques pour chaque lambda. Dans cette seconde phase, les contraintes éparses jouent un rôle d'ancres au lieu de diriger la déformation.

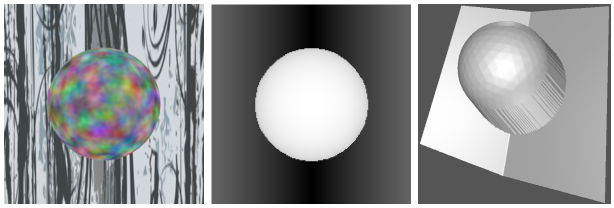
Le champ de déplacement final est ensuite calculé en combinant ces deux estimations. En pratique, les expérimentations montrent que si les grands déplacements sont correctement estimés par la première passe, alors la seconde permet de retrouver les petits détails locaux du déplacement.

# 5 Évaluations

Pour procéder à l'évaluation de la méthode proposée, nous avons utilisé plusieurs séquences dans différentes conditions. Pour commencer, nous avons créé des données de synthèse pour permettre une évaluation quantitative. Dans un second temps nous avons procédé à l'acquisition et au traitement de données réelles à l'aide de deux configurations différentes, avec une ou plusieurs caméras couleur. Les différentes configurations ainsi que les résultats obtenus sont détaillés dans cette section.

## 5.1 Données de synthèse

Les données de synthèse représentent une sphère en mouvement devant deux plans également en déplacement. Cette scène est ensuite projetée dans deux caméras de synthèse de 1 Mpixels. Nous utilisons le *depth buffer* d'une de ces deux caméras virtuelles pour obtenir la carte de profondeur de la scène (voir figure 2-(a)-(b)). Cette carte de profondeur a été rééchantillonée à une résolution de $200 \times 200$ et utilisée pour créer un maillage (voir figure 2-(c)). Dans la séquence générée, la sphère se déplace en s'éloignant de la caméra, tandis que l'un des plans se déplace vers le haut et l'autre vers le bas. Il est important de noter que l'extension de la méthode proposée à $N > 1$ caméras couleur ne change rien à la formulation, cela ne fait qu'empiler plus de contraintes dans l'équation (9).



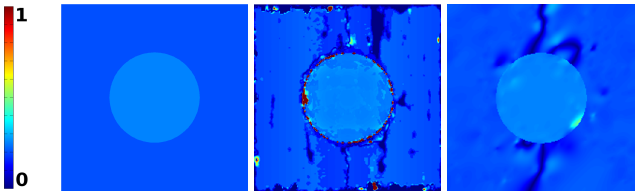
(a) (b) (c)

FIG. 2: Données de synthèse : image couleur (a), carte de profondeur (b) et le maillage inféré (c).

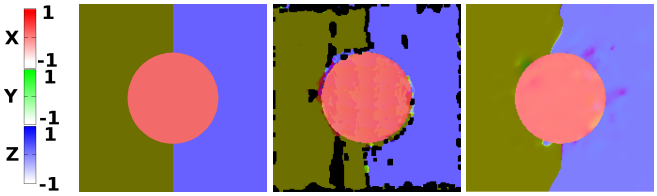
Nous avons comparé notre approche à la méthode de référence présentée dans [15] dans le cas *"Single camera, known scene geometry"*. Cette méthode requiert les même données en entrée que la nôtre et est aussi extensible au cas multi-caméra.

Les résultats obtenus sont présentés dans la figure 3 où les normes et orientations des déplacements 3D sont représentés depuis le point de vue de la caméra avec un code couleur. La figure 4 montre l'erreur en chaque point de la sur-

face maillée, tandis que la table 1 présente une comparaison numérique.

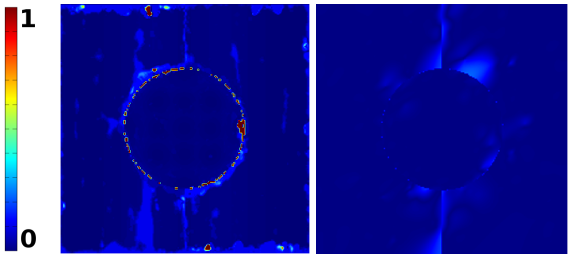



(a) Norme du déplacement 3D (code couleur)

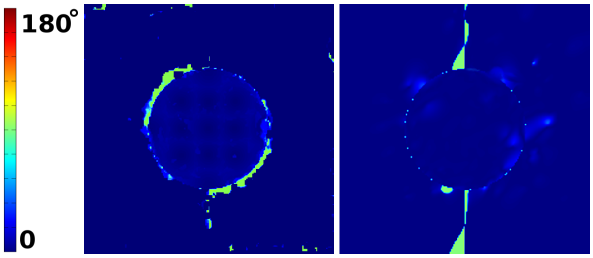



(b) Direction du déplacement 3D (code couleur)

FIG. 3: Résultats sur des données de synthèse et comparaison entre vérité terrain (gauche), la méthode de Vedula [15] (centre) et notre méthode (droite).




(a) Erreur sur la norme du déplacement 3D




(b) Erreur sur l'orientation du déplacement 3D

FIG. 4: Erreur sur des données de synthèse avec comparaison entre la méthode de Vedula [15] (gauche) et la méthode proposée (droite).

Les résultats obtenus montrent que la méthode proposée est capable de gérer correctement les discontinuités de la carte de profondeur entre la sphère et les plans. Néanmoins, à l'endroit où les deux plans se croisent, il y a une ambiguïté qui conduit à supposer que les deux plans sont connectés sur la surface maillée. Ainsi la régularisation a du mal à évaluer correctement les déplacements dans cette zone.
Les expérimentations menées montrent qu'avec de bonnes textures et des données de synthèse, les contraintes du flot

| | Vedula [15] | | Notre méthode | |
|---|---|---|---|---|
| Erreur | Moyenne | Médiane | Moyenne | Médiane |
| Norme | 33% | 7.27% | 8.68% | 2.33% |
| Angle | 8.6° | 0.10° | 2.7° | 0.12° |

TAB. 1: Erreur numérique sur des données de synthèse avec comparaison entre la méthode de Vedula [15] et la méthode proposée.

normal n'aident pas réellement à améliorer les résultats puisque les déformations sont strictement rigides et beaucoup de points d'intérêt sont détectés et appariés correctement, ce qui suffit à retrouver les mouvements dans le cas de scènes basiques. L'ajout de caméras couleur supplémentaires n'améliore pas significativement l'estimation des déplacements puisque les points d'intérêt détectés tendent à être les mêmes d'une image à l'autre dans le cas d'un faible parallaxe.

## 5.2 Données réelles

Nous avons aussi procédé à des expérimentations sur des données réelles acquises avec deux systèmes différents : (1) un système multi-caméra composé d'une caméra temps de vol Swiss Ranger SR4000 de résolution $176 \times 144$ accompagnée de deux caméras couleur de 2 Mpixels, et (2) une caméra Kinect de Microsoft capable de fournir un flux d'images couleur, chacune alignée sur une carte de profondeur de résolution $640 \times 480$. Le système multi-caméra avec la caméra temps de vol a été calibré en utilisant les travaux présentés dans [5].
Pour tester efficacement notre méthode nous avons acquis avec les deux systèmes une scène identique dans laquelle un homme se tient debout dans une pièce et joue avec une balle, la faisant sauter d'une main à l'autre. Cette scène présente à la fois de grandes discontinuités dans la carte de profondeur et des déplacements larges et rapides.

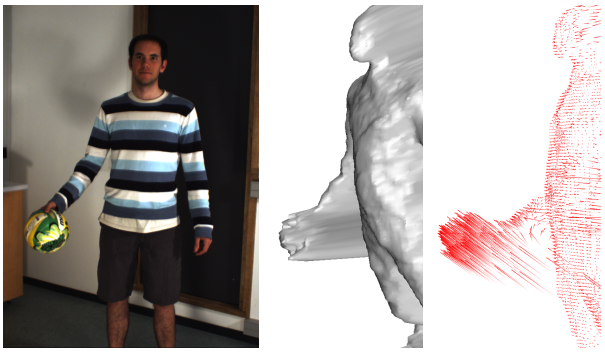


FIG. 5: Données en entrée : une des deux images couleur (gauche) et la surface calculée (centre). Résultat : le champ de déplacement 3D obtenu sur les données de la caméra temps de vol (droite).

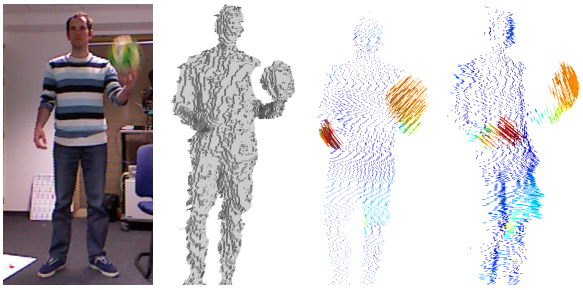


FIG. 6: Données en entrée : l'image couleur (gauche) et la surface calculée (centre). Résultat : le champ de déplacement 3D obtenu sur les données de la caméra Kinect (droite), la couleur encode la norme des vecteurs.

Les résultats présentés sur les figures 5 et 6 démontrent l'intérêt ainsi que la faisabilité de notre méthode sur des données réelles. Le champ de déplacement estimé est cohérent avec les actions exécutées par la personne. L'utilisation de deux caméras couleur dans le cas de la caméra temps de vol permet d'obtenir un résultat satisfaisant bien que les données géométriques soient très bruitées. En ce qui concerne la caméra Kinect, la résolution des données acquises induit un maillage de haute densité qui accentue la complexité du système linéaire. Dans ce cas, une implémentation parallèle peut permettre de balancer la complexité des données.

# 6 Discussion

Nous avons présenté une méthode précise et efficace pour l'estimation du flot de scène 3D permettant de fusionner des données venant de systèmes multi-caméra hybrides (par exemple couleur + profondeur) récents. La principale contribution de notre travail est de permettre de fusionner directement en 3D différents types d'informations photométriques. L'utilisation conjointe d'informations denses et éparses venant des images couleur permet à notre méthode de retrouver correctement les grands mouvements présents dans une scène tout en préservant les détails du champ de déplacement.

Une amélioration directe de ce travail serait de faire son extension au cas de $N$ caméras de profondeur. Cette étape nécessite de pouvoir fusionner de manière cohérente, en une seule surface, les différentes cartes de profondeur. À plus long terme, les champs de déplacement obtenus via notre méthode peuvent être utilisés en entrée d'un traitement de plus haut niveau sur la scène, tels que la reconnaissance d'actions ou pour une application interactive.

# Références


[1] J.-L. Barron, D.-J. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 1994.

[2] T. Basha, Y. Moses, and N. Kiryati. Mutli-View Scene Flow Estimation : A View Centered Variational Approach. In *Computer Vision and Pattern Recognition*, 2010.

[3] H. Bay, T. Tuytelaars, and L. Van Gool. SURF : Speeded Up Robust Features. In *European Conference on Computer Vision*, 2006.

[4] M. Belkin, J. Sun, and Y. Wang. Discrete Laplace Operator on Meshed Surfaces. In *Proceedings of the Symposium on Computational Geometry*, 2008.

[5] M. Hansard, R. Horaud, M. Amat, and S. Lee. Projective Alignment of Range and Parallax Data. In *Computer Vision and Pattern Recognition*, 2011.

[6] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 1981.

[7] F. Huguet and F. Devernay. A Variational Method for Scene Flow Estimation From Stereo Sequences. In *International Conference on Computer Vision*, 2007.

[8] R. Li and S. Sclaroff. Multi-scale 3D Scene Flow from Binocular Stereo Sequences. *Computer Vision and Image Understanding*, 2008.

[9] D. Lowe. Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision*, 2004.

[10] B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, 1981.

[11] J. Neumann and Y. Aloimonos. Spatio-Temporal Stereo Using Multi-Resolution Subdivision Surfaces. *International Journal of Computer Vision*, 2002.

[12] B. Petit, A. Letouzey, E. Boyer, and J. Franco. Surface Flow from Visual Cues. In *Vision, Modeling and Visualization Workshop, to appear*, 2011.

[13] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view Stereo Reconstruction and Scene Flow Estimation with a Global Image-based Matching Score. *International Journal of Computer Vision*, 2007.

[14] C. Rabe, T. Müller, A. Wedel, and U. Franke. Dense, Robust, and Accurate Motion Field Estimation from Stereo Image Sequences in Real-Time. In *European Conference on Computer Vision*, 2010.

[15] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-Dimensional Scene Flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[16] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun. Discrete Laplace Operators : No free lunch. In *Eurographics Symposium on Geometry Processing*, 2007.

[17] L. Xu, J. Jia, and Y. Matsushita. Motion Detail Preserving Optical Flow Estimation. In *Computer Vision and Pattern Recognition*, 2010.