



Suivi multiplans 3D utilisant la somme des variances conditionnelles

Bertrand Delabarre, Eric Marchand

► **To cite this version:**

Bertrand Delabarre, Eric Marchand. Suivi multiplans 3D utilisant la somme des variances conditionnelles. Congrès francophone des jeunes chercheurs en vision par ordinateur, ORASIS'13, Jun 2013, Cluny, France. pp.1-6, 2013. <hal-00829378>

HAL Id: hal-00829378

<https://hal.archives-ouvertes.fr/hal-00829378>

Submitted on 5 Jun 2013

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.

Suivi multiplans 3D utilisant la somme des variances conditionnelles

Bertrand Delabarre

Eric Marchand

Université Rennes 1, IRISA, INRIA Rennes Bretagne Atlantique, Equipe LAGADIC

bertrand.delabarre@irisa.fr

eric.marchand@irisa.fr

Résumé

Dans cet article nous nous intéressons à la réalisation d'un suivi dense robuste aux variations d'illuminations utilisant des informations 3D afin d'assurer une cohérence spatio-temporelle. La méthode proposée se base sur un algorithme de suivi de plans 3D qui minimise l'erreur entre une image courante et une image de référence en estimant les paramètres de la transformation 3D induite par le mouvement de la caméra tout en intégrant des contraintes entre les positions des différents plans de la scène. L'inconvénient majeur de cette approche est lié à la fonction de similarité utilisée pour effectuer cette minimisation (la somme des différences au carré ou SSD). Cette dernière est très peu résistante face aux variations de luminosité. Pour cela, un autre critère a été choisi : la somme des variances conditionnelles (SCV). Celle-ci est invariante aux variations globales de luminosité, ce qui confère à l'algorithme proposé une robustesse vis à vis de ces conditions. L'approche choisie, après avoir été décrite, est testée face à diverses conditions d'illumination avant d'être utilisée pour effectuer du suivi en extérieur.

Mots Clef

Suivi visuel, somme des variances conditionnelles, suivi visuel dense, suivi visuel 3D.

Abstract

This paper deals with dense visual tracking robust towards illumination variations using 3D information to provide a space-time coherency. The proposed method is based on a piecewise-planar scenes visual tracking algorithm which aims to minimize an error between an observed image and a reference template by estimating the parameters of a rigid 3D transformation taking into account the relative positions of the planes in the scene. The major drawback of that approach stems from the similarity function used to perform the minimization (the sum of squared differences or SSD) as it is very poorly robust towards illumination variations. To prevent that, another similarity function has been chosen : the sum of conditional variance (SCV). As it is invariant to global illumination variations, the proposed algorithm is robust with relation to those conditions.

The proposed approach, after being described, is tested in different varying illumination conditions before being used for an outdoor visual tracking task.

Keywords

Visual tracking, sum of conditional variance, dense visual tracking, 3D visual tracking.

1 Introduction

Le suivi visuel est une étape fondamentale de la majeure partie des applications de vision par ordinateur. Son domaine d'application s'étend, entre autres, de l'asservissement visuel [5] à la réalité augmentée [6]. Il existe plusieurs familles de suivi visuel. Il est par exemple possible de distinguer les méthodes basées sur des primitives visuelles telles que des points [10, 14], des lignes ou encore des moments et d'autre part les méthodes dites denses ou basées template qui se reposent sur un patch extrait d'une image référence. Cet article se place dans la seconde catégorie. Le but de ce type de tracking est de minimiser une fonction représentant la différence entre un patch de référence et un patch extrait de l'image courante. Plusieurs travaux se sont basés sur différentes mesures de similarité. Parmi celles-ci figurent la SSD [1] très simple à calculer mais assez peu robuste, l'information mutuelle [7] très robuste vis à vis des perturbations de la scène observée mais assez complexe et coûteuse, la somme des variances conditionnelles (SCV) [15] ou encore la cross corrélation normalisée (NCC) [16], des "entre-deux" qui permettent une robustesse vis à vis des conditions d'illumination globale en gardant un faible coût de calcul. Ces approches permettent alors un suivi visuel optimisant pour la plupart des paramètres de mouvement 2D comme dans [1, 7, 15] mais peuvent aussi se baser sur des paramètres de mouvement 3D rigide comme dans [2] ou directement de pose comme dans [4].

L'objectif de cet article est d'introduire un suivi visuel robuste aux conditions d'illuminations intégrant les contraintes Euclidiennes et gardant ainsi une cohérence spatiale. Pour ce faire, nous avons choisi de nous baser, dans cet article, sur les approches proposées par [2] et [15]. En effet, l'utilisation de la somme des variances condi-

tionnelles permet une invariance vis à vis des changements d'éclairage globaux de la scène, chose qui arrive fréquemment dans des scènes réelles, notamment en extérieur. De plus, cette mesure n'est que peu coûteuse en temps de calcul et permet une convergence rapide. L'approche proposée se repose également sur une optimisation de mouvement 3D (SE(3)), ce qui permettra de garder une cohérence spatiale de la scène et évitera une couche algorithmique nécessaire au passage du mouvement 2D vers le mouvement 3D afin d'évaluer le déplacement de la scène. La principale contribution de cet article est l'utilisation de la SCV dans le cadre d'un suivi visuel multiplans optimisant des paramètres de pose 3D contrairement à [2] où le suivi se basait sur la SSD et [15] dans lequel la SCV était utilisée pour optimiser les paramètres d'un mouvement 2D dans l'image. L'article est organisé de la manière suivante. Dans un premier temps, les principes du suivi visuel différentiel sont rappelés puis la somme des variances conditionnelles est définie et son utilisation en tant que fonction de similarité dans le cadre d'un suivi visuel est analysée. Les travaux présentés par [2] permettant un suivi 3D sont alors rappelés et adaptés à la SCV. L'algorithme proposé est enfin comparé avec les approches de [2] et [15] à travers différents résultats expérimentaux.

2 Suivi visuel différentiel

Le suivi visuel différentiel dense est une famille d'approches basées sur l'optimisation d'une fonction de similarité [1]. Le but est d'estimer les paramètres \mathbf{x} de la transformation $\mathbf{T}(\mathbf{x})$ d'un patch de référence I^* dans une séquence d'images. Lors d'un suivi image par image, la référence est habituellement extraite de la première image \mathbf{I}^* de la séquence. En considérant une fonction de dissimilarité f , le problème peut être exprimé par :

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} f(I^*, \mathbf{w}(I_t, \mathbf{T}(\mathbf{x}))) \quad (1)$$

dont le but est de trouver les paramètres $\hat{\mathbf{x}}_t$ du mouvement $\mathbf{T}(\mathbf{x})$ qui minimise la différence entre la référence I^* et le patch courant I_t warpé. L'espace des paramètres \mathbf{x} à estimer est variable puisque $\mathbf{T}(\mathbf{x})$ peut représenter plusieurs types de transformations. Ainsi, l'approche introduite par [11] a commencé par considérer de simples translations, conduisant à $\mathbf{x} \in \mathbb{R}^2$. Par la suite, d'autres modèles de transformation ont été étudiés comme la transformation affine [8] qui amène à $\mathbf{x} \in \mathbb{R}^6$. D'autres approches ont quant à elles considéré des homographies comme dans [1] donnant alors $\mathbf{x} \in \text{SL}(3)$ ou [2] conduisant à $\mathbf{x} \in \text{SE}(3)$.

3 Somme des variances conditionnelles

Le suivi visuel différentiel optimise une fonction de similarité f . Dans cet article, la fonction choisie est la somme des variances conditionnelles ou SCV. Cette section commencera donc par définir cette fonction avant de vérifier

si cette dernière est adaptée à un usage dans le cadre d'un suivi visuel.

3.1 Définition de la fonction de similarité

Les auteurs de [15] ont proposé une méthode se basant sur la somme des variances conditionnelles (SCV). Le but est de ne pas utiliser simplement une image de référence constante lors du suivi mais d'adapter cette dernière aux conditions d'illumination de la scène observée. Ainsi, à chaque itération du suivi, l'image référence est adaptée à l'image courante et l'optimisation est effectuée par rapport aux paramètres d'un warp 2D. Cette image calculée sera notée $\hat{\mathbf{I}}$:

$$\hat{\mathbf{I}}(\mathbf{q}) = \mathcal{E}(\mathbf{I}(\mathbf{q}) \mid \mathbf{I}^*(\mathbf{q})) \quad (2)$$

où \mathcal{E} est un opérateur de prédiction qui permet de prédire l'allure de \mathbf{I}^* une fois adaptée aux conditions d'éclairage subies par \mathbf{I} . La mesure de similarité sera alors simplement la somme des différences au carré entre le patch dans l'image courante \mathbf{I} et celui dans l'image attendue $\hat{\mathbf{I}}$:

$$SCV = \sum_{i=1}^{n \times m} [\mathbf{I}(\mathbf{q}_i) - \hat{\mathbf{I}}(\mathbf{q}_i)]^2. \quad (3)$$

L'image prédite est calculée à partir des fonctions de probabilité d'apparition des niveaux de gris dans les images \mathbf{I} et \mathbf{I}^* . Ces probabilités sont calculées à partir des analyses empiriques de \mathbf{I} et \mathbf{I}^* . Ainsi, pour chaque couple de niveaux de gris i dans \mathbf{I} et j dans \mathbf{I}^* , la probabilité jointe de i et j est donnée par :

$$\begin{aligned} P_{II^*}(i, j) &= P(\mathbf{I}(\mathbf{q}) = i, \mathbf{I}^*(\mathbf{q}) = j) \\ &= \frac{1}{N} \sum_{\mathbf{q}} \alpha(\mathbf{I}(\mathbf{q}) - i) \alpha(\mathbf{I}^*(\mathbf{q}) - j) \end{aligned} \quad (4)$$

où $N = n \times m$ et $\alpha(u) = 1$ si et seulement si $u = 0$. $P_{II^*}(i, j)$ représente alors la probabilité qu'un pixel \mathbf{q} prenne la valeur i dans \mathbf{I} et j dans \mathbf{I}^* . À partir de ces probabilités jointes, la fonction de probabilité de \mathbf{I}^* est donnée par :

$$P_{I^*}(j) = \sum_i P_{II^*}(i, j). \quad (5)$$

Pour chaque niveau de gris j dans \mathbf{I}^* , son équivalent dans l'image prédite $\hat{\mathbf{I}}$ est finalement donné par :

$$\hat{\mathbf{I}}(j) = \sum_i i \frac{P_{II^*}(i, j)}{P_{I^*}(j)}. \quad (6)$$

3.2 Analyse de la fonction de similarité

Afin de s'assurer que la SCV est plus adaptée aux variations d'illumination de la scène observée que la SSD, ces deux mesures ont été comparées. Ce paragraphe définira dans un premier temps le protocole de test suivi puis décrira les résultats observés.

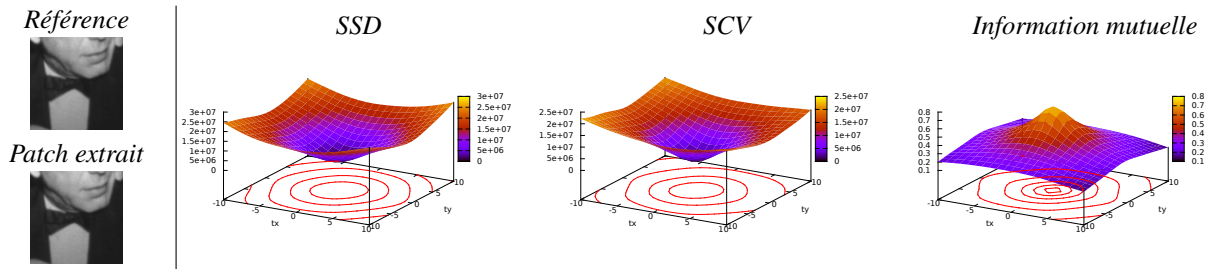


FIGURE 1 – Fonctions de coût dans le cas nominal. Le patch affiché est celui qui correspond à un déplacement de la fenêtre de (0,0).

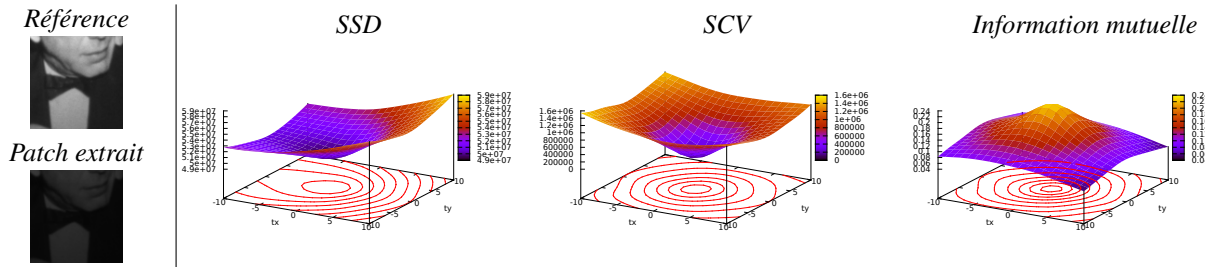


FIGURE 2 – Fonctions de coût dans le cas où la luminosité a été diminuée de manière significative. Le patch affiché est celui qui correspond à un déplacement de la fenêtre de (0,0).

Protocole de test

Tout d’abord, un patch de taille (100×100) est extrait depuis une position connue dans une image référence. Par la suite, une fenêtre de taille (100×100) est déplacée autour de la position du patch de référence et à chaque étape le contenu de la fenêtre est extrait dans l’image à comparer. Les deux fonctions de coût sont alors calculées en considérant ces deux patches et les résultats sont reportés dans un graphe. Pour ces tests la fenêtre commence avec une position relative de (-10, -10) par rapport à la position du patch et finit à une position relative de (10,10).

Résultats

Les courbes obtenues sont reportées dans les figures 1 et 2. En plus de la SSD et de la SCV, les courbes associées à l’information mutuelle [17] ont été ajoutées. En effet, cette mesure est notoirement connue pour sa robustesse vis à vis de tout type de perturbations [18, 7] et est en conséquence un bon témoin. Pour le calcul de la SCV, le nombre de niveaux de quantification est de 64 et dans le cas de l’information mutuelle il est de 8. La figure 1 représente l’allure de la courbe dans un cas nominal. Il apparaît que les trois fonctions présentent un optimum global au point (0,0), ce qui signifie que les trois sont adaptées à une optimisation dans ces conditions. Une fois ce témoin établi, l’image comparée a subi une diminution de luminosité (la référence reste inchangée) afin de simuler une variation brutale de luminosité de la scène suivie (voir figure 2). Cette fois, la courbe représentant la SSD est très perturbée. Le minimum n’est plus marqué et est même légèrement déplacé. Aussi bien la SCV que l’information mutuelle ne sont elles que très peu impactées, seules les dérivées autour des optima étant légèrement atténuées. Il en découle que dans ces conditions la SCV est toujours adaptée alors que la SSD ne l’est plus.

4 Suivi dense multiplans

La fonction de similarité ayant été définie, cette section a pour but de présenter l’algorithme de suivi utilisé pour effectuer le suivi dense. Elle rappellera donc dans un premier temps sur les notations utilisées avant de décrire l’algorithme de suivi visuel.

4.1 Notations et principe

Notons \mathbf{T} la transformation permettant de passer d’un repère à un autre dans le monde 3D. Elle peut être représentée par une matrice homogène appartenant à $SE(3)$:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \quad (7)$$

où \mathbf{R} est une matrice de rotation ($\mathbf{R} \in SO(3)$) et \mathbf{t} est un vecteur correspondant à la translation entre les deux repères ($\mathbf{t} \in \mathbb{R}^3$). Notons ensuite \mathbf{H} l’homographie reliant deux plans. Cette dernière peut être exprimée comme :

$$\mathbf{H}(\mathbf{T}) = \mathbf{R} + \mathbf{t}\mathbf{n}_d^{*\top} \quad (8)$$

où $\mathbf{n}_d^* = \frac{\mathbf{n}}{d^*}$ est le rapport entre la normale unitaire au plan \mathbf{n}^* et la distance d^* entre le plan et l’origine du repère considéré. Notons que $\mathbf{H}(\mathbf{T})$ peut être divisée par $\sqrt[3]{1 + \mathbf{t}^\top \mathbf{R} \mathbf{n}_d^{*\top}}$ à des fins de normalisation. Cette homographie peut alors être exprimée dans le plan image de la caméra par :

$$\mathbf{G}(\mathbf{T}) = \mathbf{K}\mathbf{H}(\mathbf{T})\mathbf{K}^{-1} \quad (9)$$

où \mathbf{K} est la matrice de projection liée à la caméra :

$$\mathbf{K} = \begin{bmatrix} P_x & 0 & u_0 \\ 0 & P_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (10)$$

Cette transformation dans le plan image sera effectuée grâce à un warp $\mathbf{w}(\mathbf{G})$ qui pour tout point $\mathbf{q} = (x, y, 1)^\top$ donne :

$$\mathbf{r} = \mathbf{w}(\mathbf{G})(\mathbf{q}) = \mathbf{G}(\mathbf{T})\mathbf{q}. \quad (11)$$

Il est également possible de retrouver \mathbf{q} à partir de \mathbf{r} :

$$\mathbf{q} = \mathbf{w}^{-1}(\mathbf{G})(\mathbf{r}) = \mathbf{w}(\mathbf{G}^{-1})(\mathbf{r}) = \mathbf{G}(\mathbf{T})^{-1}\mathbf{r}. \quad (12)$$

4.2 Algorithme de suivi

Les auteurs de [2] proposent une méthode permettant d'intégrer les propriétés euclidiennes à la tâche de suivi. Notons $\hat{\mathbf{I}}$ la prédiction d'une image de référence \mathbf{I}^* basée sur une image courante \mathbf{I} comme présenté en équation (2). Considérons une région de taille $(n \times m)$ de cette nouvelle référence $\hat{\mathbf{I}}$ représentant la projection d'un plan. Cette région sera appelée patch de référence. Le but de la tâche de suivi de ce patch de référence est de trouver la transformation $\bar{\mathbf{T}} \in \text{SE}(3)$ qui relie les pixels de la référence dans $\hat{\mathbf{I}}$ à leur équivalent dans l'image considérée \mathbf{I} , soit $\forall i \in \text{ref}$:

$$\mathbf{I}(\mathbf{r}_i) = \mathbf{I}(\mathbf{w}(\bar{\mathbf{T}})(\mathbf{q}_i)) = \hat{\mathbf{I}}(\mathbf{q}_i). \quad (13)$$

L'homographie concernée change selon le plan auquel appartient le point. Pour plus de simplicité, le reste de l'article ne considérera qu'un plan lors des explications des différentes étapes liées aux algorithmes décrits. En partant d'une approximation $\mathbf{T}(\mathbf{x})$ de $\bar{\mathbf{T}}$, le but de l'algorithme est de trouver une transformation incrémentale $\hat{\mathbf{T}}$ qui minimise la somme des différences au carré (SSD) entre la région dans l'image courante \mathbf{I} de la région warpée par la transformation $\hat{\mathbf{T}}\mathbf{T}(\mathbf{x})$ et la région de référence, soit les 6 paramètres \mathbf{x} de $\mathbf{T}(\mathbf{x})$ minimisant :

$$\phi(\mathbf{x}) = \sum_{i=1}^{n \times m} \left[\mathbf{I}(\mathbf{w}(\hat{\mathbf{T}}\mathbf{T}(\mathbf{x}))(\mathbf{q}_i)) - \hat{\mathbf{I}}(\mathbf{q}_i) \right]^2. \quad (14)$$

Le minimum global est alors atteint en $\mathbf{x} = \mathbf{x}_0$ et \mathbf{x}_0 vérifie :

$$\hat{\mathbf{T}}^{-1}\bar{\mathbf{T}} = \mathbf{T}(\mathbf{x}_0). \quad (15)$$

Pour atteindre ce minimum, une minimisation itérative est effectuée et à chaque nouvelle itération l'approximation $\hat{\mathbf{T}}$ de $\bar{\mathbf{T}}$ est mise à jour à partir de la valeur de \mathbf{x} calculée à l'itération précédente :

$$\hat{\mathbf{T}} \leftarrow \hat{\mathbf{T}}\mathbf{T}(\mathbf{x}). \quad (16)$$

Pour effectuer cette minimisation, les auteurs de [2] utilisent un algorithme de type ESM ([3]) qui permet d'approcher la matrice Hessienne associée à la fonction à optimiser en se basant sur sa matrice Jacobienne. Pour cela, le calcul se base sur la série de Taylor du second ordre associée à la fonction (14) comme dans [12]. En $\mathbf{x} = \mathbf{0}$, la série de $\phi(\mathbf{x})$ donne :

$$\phi(\mathbf{x}) = \phi(0) + \mathbf{J}(0)\mathbf{x} + \frac{1}{2}\mathbf{H}(0, \mathbf{x})\mathbf{x} + O(\|\mathbf{x}\|^3) \quad (17)$$

où \mathbf{J} est la Jacobienne de $\phi(\mathbf{x})$, \mathbf{H} sa Hessienne et $O(\|\mathbf{x}\|^3)$ le reste non quantifié d'ordre 3. La Jacobienne peut elle aussi être développée :

$$\mathbf{J}(\mathbf{x}) = \mathbf{J}(0) + \mathbf{H}(0, \mathbf{x}) + O(\|\mathbf{x}\|^2). \quad (18)$$

L'injection de (18) dans (17) donne alors :

$$\phi(\mathbf{x}) = \phi(0) + \frac{1}{2}(\mathbf{J}(0) + \mathbf{J}(\mathbf{x}))\mathbf{x} + O(\|\mathbf{x}\|^3). \quad (19)$$

À chaque itération de la minimisation, les nouveaux paramètres de la transformation sont alors calculés grâce à la décomposition de la Jacobienne en cinq termes (voir détails en [2]) :

$$\mathbf{x}_0 = - \left(\left(\frac{\mathbf{J}_\mathbf{I} + \mathbf{J}_{\hat{\mathbf{I}}}}{2} \right) \mathbf{J}_\mathbf{w} \mathbf{J}_\mathbf{K} \mathbf{J}_{\hat{\mathbf{T}}} \mathbf{J}_\mathbf{X}(0) \right)^+ \phi(0). \quad (20)$$

5 Expériences

Différentes expérimentations ont été réalisées afin de valider l'approche proposée. Pour initialiser les trackers, une estimation de la pose est calculée dans l'image de référence grâce à la mise en correspondance de quatre points. Cette pose est par la suite facilement mise à jour suivant les paramètres de déplacement estimés par l'algorithme pour chaque image. Chaque plan est alors projeté suivant cette pose estimée ce qui permet d'initialiser le tracker. Le suivi peut alors être lancé et se dérouler sur la séquence choisie. Pour le calcul de la SCV, une quantification de l'image est effectuée sur 64 niveaux de gris afin de lisser la fonction de coût tout en évitant de perdre trop de détails du patch de référence. De plus, les implémentations sont effectuées sur des pyramides d'images et utilisent des M-estimateurs [9, 13] afin d'en augmenter la robustesse.

5.1 Analyse du domaine de convergence

Une première expérience a été réalisée afin de mettre en évidence le domaine de convergence de chaque approche dans différentes conditions. Pour cela, le tracker est dans un premier temps initialisé par rapport à la référence, donnant une vérité terrain. Les paramètres de pose correspondant sont alors perturbés grâce à l'ajout d'un bruit gaussien de paramètre σ variable et le suivi est lancé depuis cette pose perturbée. Une fois le suivi achevé, la pose obtenue est comparée à la vérité terrain et si l'erreur est négligeable le suivi est considéré comme réussi. Ce processus est réitéré 500 fois pour chaque valeur de σ . Les résultats sont reportés sur la figure 5.1. Le bruit choisi sur les degrés de liberté en rotation est fixé à $\sigma_R = 0.01$ rad comme leur impact est grand sur la position de départ (voir figure 4.2 pour des exemples de positions de départ bruitées). Le bruit appliqué sur les degrés de liberté en rotation σ_t varie quant à lui de 0.002 à 0.05 m. Les résultats montrent que dans un cas nominal, le domaine de convergence de la méthode SCV est légèrement plus étroit que celui de la SSD mais que dans le cas de variations de luminosité de la scène (ici via l'extinction des lumières de la pièce) ce domaine de convergence n'est pas impacté au contraire de celui de la SSD qui n'est plus du tout adaptée.

5.2 Conditions d'illuminations changeantes

Une deuxième expérience a été réalisée afin de comparer l'algorithme proposé à celui décrit dans [2] (noté SSD) dans des conditions d'illuminations changeantes (voir figure 5). Lors de la séquence, l'algorithme présenté arrive à

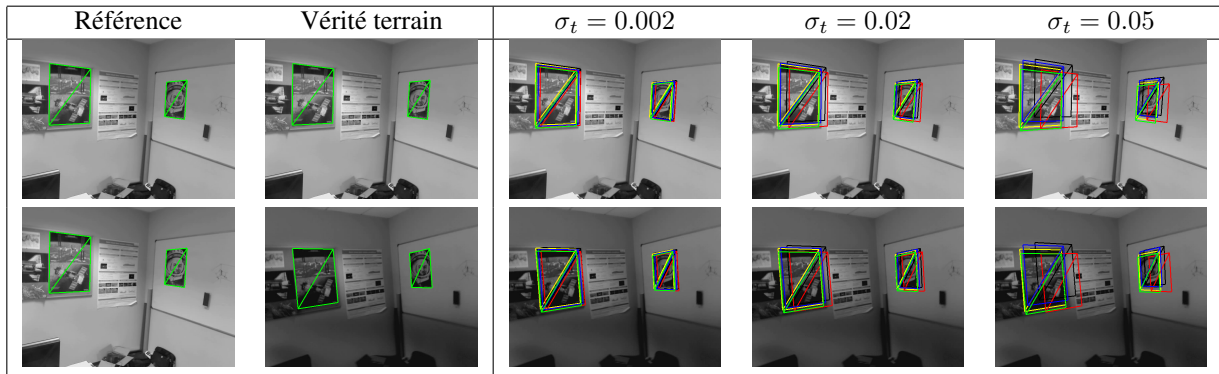


FIGURE 3 – Exemples de positions de départs suivant σ_t . Pour toutes les expériences $\sigma_R = 0.1$.

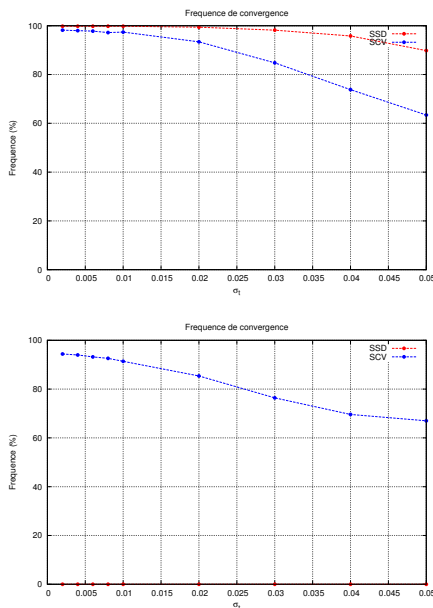


FIGURE 4 – Au dessus : dans des conditions nominales, le domaine de convergence de la SCV est légèrement moins grand que celui de la SSD. En dessous : dans des conditions de variation de luminosité, la SSD diverge alors que la SCV n'est pas impactée.

suivre les patches de référence appartenant à deux plans perpendiculaires malgré des conditions de flou et de changement d'illumination brutales. La figure montre que l'algorithme SSD échoue rapidement à la première variation brutale d'illumination à l'image 12. Cela confirme que l'adaptation du patch de référence effectuée par l'approche proposée permet de s'affranchir de l'hypothèse de constance de la luminance de la scène nécessaire à l'utilisation de la SSD. Cette expérience permet aussi de mettre en avant la grande résistance de la méthode proposée au flou.

5.3 Suivi de bâtiments

Une dernière expérience a été réalisée dans des conditions extérieures. Elle a pour but de réaliser le suivi de bâtiments. Dans un premier temps deux trackers simples, un pour chaque patch suivi dans la scène, ont été lancés sur la séquence considérée. La figure 6 montre les résultats obtenus. Elle montre que cette approche a échoué au moment

où un patch revenait dans l'image après en avoir été partiellement sorti. Ce patch a alors été totalement perdu ce qui a provoqué l'échec de la tâche de suivi. Grâce à la cohérence spatiale apportée par l'approche proposée l'algorithme suivant les deux patches au sein du même tracker a permis d'éviter ce problème et de continuer le suivi sans difficulté.

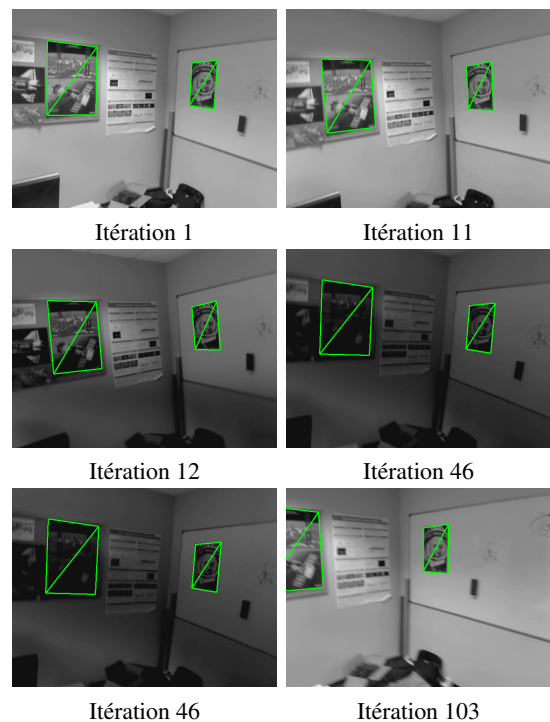


FIGURE 5 – Algorithmes de suivi dans des conditions variables. L'algorithme basé sur la SSD est perdu au premier changement brutal d'illumination à l'itération 12. L'algorithme proposé suit toujours le patch malgré les diminutions et augmentations brutales de luminosité (itérations 12, 34 et 46) ainsi que par de fortes conditions de flou (itération 103).

6 Conclusion

Dans cet article nous avons proposé une méthode de suivi dense, permettant de garder une cohérence spatiale de la scène observée et robuste aux changements de conditions

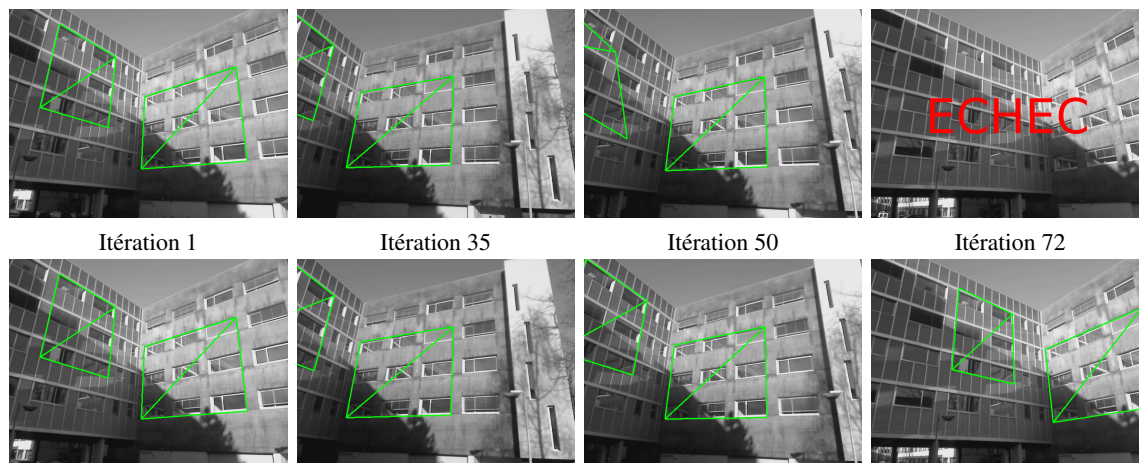


FIGURE 6 – Algorithmes de suivi dans des conditions variables. Au dessus : deux trackers monoplans. En dessous : un tracker multiplans. L’algorithme basé sur un suivi de plan simple est perdu à l’itération 50.

d’illumination de la scène. Cet algorithme est basé sur la somme des variances conditionnelles, une mesure qui est invariante aux variations globales de la scène observée tout en gardant une complexité computationnelle très faible. Cette dernière est utilisée dans le cadre d’une minimisation de paramètres 3D, ce qui permet de respecter une cohérence spatiale entre les plans suivis. Cette approche a été testée sur des scènes comportant des variations brutales de luminosité avec de bons résultats et s’est également montrée satisfaisante sur des séquences de suivi de bâtiments sur lesquelles un suivi simple n’était pas possible.

Références

- [1] S. Baker and I. Matthews. Lucas-kanade 20 years on : A unifying framework. *Int. Journal of Computer Vision*, 56(3) :221–255, 2004.
- [2] Benhimane and E. S. Malis. Integration of Euclidean constraints in template-based visual tracking of piecewise-planar scenes. In *IEEE/RSJ International Conference on Intelligent Robots Systems*, 2006.
- [3] S. Benhimane and E. Malis. Homography-based 2d visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA’06*, Orlando, FL, May 2006.
- [4] G. Caron, A. Dame, and E. Marchand. L’information mutuelle pour l’estimation visuelle directe de pose. In *18e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA 2012*, Lyon, France, January 2012.
- [5] F. Chaumette and S. Hutchinson. Visual servo control, Part I : Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4) :82–90, December 2006.
- [6] A.I. Comport, E. Marchand, M. Pressigout, and F. Chaumette. Real-time markerless tracking for augmented reality : the virtual visual servoing framework. *IEEE Trans. on Visualization and Computer Graphics*, 12(4) :615–628, July 2006.
- [7] A. Dame and E. Marchand. Second order optimization of mutual information for real-time image registration. *IEEE Trans. on Image Processing*, 2012.
- [8] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10) :1025–1039, October 1998.
- [9] P.-J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [10] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2) :91–110, 2004.
- [11] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence, IJCAI’81*, pages 674–679, 1981.
- [12] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *IEEE Int. Conf. on Robotics and Automation, ICRA’04*, volume 2, pages 1843–1848, New Orleans, April 2004.
- [13] E. Malis and E. Marchand. Experiments with robust estimation techniques in real-time robot vision. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS’06*, pages 223–228, Beijing, China, October 2006.
- [14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. Journal of Computer Vision*, 65 :43–72, 2005.
- [15] R. Richa, R. Sznitman, R. Taylor, and G. Hager. Visual tracking using the sum of conditional variance. In *IEEE Conference on Intelligent Robots and Systems, IROS’11*, pages 2953–2958, San Francisco, September 2011.
- [16] Glauco Garcia Scandaroli, Maxime Meilland, and Rogério Richa. Improving ncc-based direct visual tracking. In *Proceedings of the 12th European conference on Computer Vision - Volume Part VI, ECCV’12*, pages 442–455, Berlin, Heidelberg, 2012. Springer-Verlag.
- [17] CE. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27 :379–423, 623–656, 1948.
- [18] P. Viola and W. Wells. Alignment by maximization of mutual information. In *Int. Conf. on Computer Vision, ICCV’95*, Washington, DC, 1995.