

# Une approche unifiée reposant sur l'information mutuelle pour l'estimation du mouvement 2D et l'asservissement visuel

Amaury Dame<sup>1</sup>Eric Marchand<sup>2</sup><sup>1</sup> CNRS, <sup>2</sup> Université de Rennes 1<sup>1,2</sup> INRIA Rennes-Bretagne Atlantique, Lagadic  
Prenom.Nom@irisa.fr

## Résumé

*Dans cet article nous proposons une approche unifiée reposant sur l'information mutuelle et permettant la définition tant d'algorithmes de suivi différentiel (dans la lignée du KLT) que d'une alternative originale et robuste à l'utilisation d'informations visuelles géométriques en asservissement visuel. L'information mutuelle est une mesure de recalage introduite dans le cadre de la théorie de l'information et utilisée, principalement, dans le domaine de l'imagerie médicale pour le recalage multimodal. Utilisée récemment pour le suivi, cette mesure de similarité est reconnue pour sa robustesse par rapport aux changements d'illumination ainsi qu'aux occultations. Nous proposons un cadre général permettant de considérer cette mesure dans le cadre d'un processus d'estimation du mouvement 2D dans une séquence d'images. La fonction de similarité reposant sur l'entropie des images étant non-linéaire, une attention particulière sera portée à l'estimation efficace des paramètres du mouvement. Enfin, la problématique de l'estimation du mouvement étant très similaire à celui de l'asservissement visuel, nous montrerons que la même approche peut être utilisée pour réaliser des tâches de positionnement en robotique. Différents résultats valideront les développements proposés.*

## Mots Clef

Information mutuelle, suivi différentiel, asservissement visuel.

## 1 Introduction

L'estimation du mouvement dans des séquences d'images est un problème clef dans le domaine de la vision par ordinateur. Dans le cadre de cet article nous nous intéresserons uniquement au contexte de l'estimation de modèle paramétrés de mouvement.

Si de nombreuses solutions ont été proposées dans la littérature [9][12][6][1][17], la fiabilité de ces approches est souvent remise en cause principalement du fait de la faible robustesse aux changements d'illumination ou aux occultations de la fonction de similarité sous-jacente : la différence de l'intensité lumineuse (SSD). Des solutions efficaces reposant sur des processus d'estimation robustes [12] ou sur

l'estimation locale des variations d'illumination [6][17] ont été proposées pour tenter de résoudre ce problème.

Outre la robustesse aux variations d'illumination, nous avons pour objectif de pouvoir mettre en œuvre un processus de suivi dans un contexte de multimodalités (par exemple, localiser et suivre dans des images infrarouges des objets dont l'image de référence a été apprise dans une image acquise par une caméra classique). La solution proposée ne peut donc reposer sur une modification du processus d'optimisation ou sur l'ajout de paramètres permettant d'introduire des variations locales de la photométrie dans la SSD. C'est dans cette optique que nous nous intéressons à l'information mutuelle [19][14]. Cette fonction d'alignement reposant sur la quantité d'information partagée par les deux signaux images [16], est reconnue pour sa robustesse aux occultations, aux variations d'illumination et à sa robustesse face aux différentes modalités d'acquisition. Viola [19] a montré l'efficacité de cette technique, utilisée principalement pour le recalage en imagerie médicale. Mais ce n'est que récemment que des techniques permettant l'optimisation de ce critère par des approches de type Newton ont vu le jour [18][5] et que des approches différentielles (ie, similaire au KLT *modulo* la modification fondamentale de la fonction de similarité) ont été proposées [4][13][7][15].

Le suivi n'est pas le seul contexte que nous considérerons dans cet article. Dans [10], nous avons montré l'équivalence entre un grand nombre de problèmes de vision par ordinateur (calcul de pose, calibration, estimation de modèles paramétrés de mouvement,...) et l'asservissement visuel [2]. L'asservissement visuel consiste à contrôler les mouvements d'un système dynamique, tel un robot, en utilisant les informations visuelles d'images acquises par une ou plusieurs caméras. Classiquement, ce type d'approche nécessite d'extraire des primitives visuelles de l'image afin de définir la loi de contrôle. L'extraction et le suivi robuste en temps réel de ces informations visuelles est un problème classique pour l'asservissement visuel [11]. Cependant, dans [3] il a été montré que considérer l'image dans son ensemble en tant que primitive visuelle, permet de définir une loi de commande permettant de s'affranchir de l'extraction et du suivi d'indices visuels. Positionner le

robot à la position désirée revient donc à minimiser l'erreur entre l'image courante et l'image désirée. Fondamentalement, et suivant l'équivalence établie dans [10] et bien que l'espace des paramètres considérés soit différent, l'asservissement visuel photométrique et le suivi différentiel sont deux problèmes très proches. Souffrant des mêmes défauts (principalement l'absence de robustesse aux variations d'illumination) nous proposerons, comme dans le cas du suivi, de modifier la nature fondamentale de la fonction de coût en remplaçant la SSD par l'information mutuelle.

Les contributions de cet article sont donc les suivantes : nous proposons un cadre unifié permettant la définition de techniques de suivi différentiel et d'asservissement visuel reposant sur l'information mutuelle. Nous porterons un soin à la définition du processus d'optimisation et en particulier à la définition des Jacobien et Hessien de la fonction de similarité (à notre connaissance, ces méthodes utilisent toutes une approximation sur le calcul du Hessien de l'information mutuelle qui se révèle être problématique) permettant une convergence à la fois plus efficace et plus rapide de l'estimation. Finalement, nous montrerons expérimentalement que les techniques proposées sont efficaces dans les deux contextes proposés et sont robustes tant à des variations non-linéaires de l'illumination que dans le cas où le *template* de référence et l'image courante sont acquises avec des modalités différentes.

Cet article s'articule de la manière suivante. Dans la prochaine section nous montrerons comment les problèmes de suivi différentiel et d'asservissement visuel sont définis sous forme de maximisation non linéaire d'une fonction d'alignement reposant sur l'information mutuelle. La section 3 présente l'information mutuelle et détaille comment calculer ses dérivées. Enfin la section 4 présente les différents résultats obtenus pour l'asservissement visuel ainsi que pour le suivi d'images de référence dans des séquences.

## 2 Deux problèmes de recalage

Dans un premier temps nous allons décrire les deux problèmes qui vont nous intéresser dans cet article, le suivi différentiel de type KLT et l'asservissement visuel. Nous tenterons de montrer les liens entre ces deux problèmes et les solutions que nous proposons.

### 2.1 Suivi différentiel

Rappelons dans un premier temps la problématique de l'estimation du mouvement 2D par des techniques de suivi différentiel.

Le problème est formulé de la manière suivante : nous cherchons à estimer les paramètres de déplacement  $\mathbf{p}$  qui permettent de passer de l'image de référence  $T$  à l'image courante  $I$ . Ce problème est le plus communément résolu en minimisant la différence entre l'image courante et une

image de référence déplacée, on l'écrit alors [9] :

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} (SSD(T, I(\mathbf{p}))) \quad (1)$$

$$= \arg \min_{\mathbf{p}} \sum_{\mathbf{x}} [T(\mathbf{x}) - I(\mathbf{p})(\mathbf{x})]^2 \quad (2)$$

où  $I(\mathbf{p})(\mathbf{x}) = I(w(\mathbf{x}, \mathbf{p}))$ .  $w : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  est la fonction de transfert qui modifie un point  $\mathbf{x}$  de l'image de référence à son correspondant dans l'image courante. Il est possible d'ajouter à cela des paramètres supplémentaires pour prendre en compte les variations d'illumination [17]. Notre objectif étant d'être robuste aux changements d'illumination et à la multimodalité dans le processus d'acquisition des images, nous avons choisis de réécrire le problème de la manière suivante :

$$\hat{\mathbf{p}} = \arg \max_{\mathbf{p}} (MI(T, I(\mathbf{p}))) \quad (3)$$

où nous chercherons à maximiser l'information mutuelle entre les deux images permettant ainsi d'estimer le déplacement  $\mathbf{p}$  entre chaque image de la séquence.

### 2.2 Asservissement visuel

Le problème de positionnement d'un robot peut être exprimé comme suit : connaissant l'image désirée  $T(\mathbf{r}^*)$  (correspondant à une position  $\mathbf{r}^*$  de la caméra) il convient de déplacer le robot de manière à ce que l'image courante  $I(\mathbf{r})$  soit équivalente à  $T(\mathbf{r}^*)$ .

Plutôt que d'extraire des images  $T(\mathbf{r}^*)$  et  $I(\mathbf{r})$  des informations visuelles qui devront être mises en correspondance une solution, l'asservissement visuel photométrique, proposée dans [3] repose sur l'utilisation directe de la luminance comme information visuelle.

$$\hat{\mathbf{r}} = \arg \min_{\mathbf{r}} \sum_{\mathbf{x}} [T(\mathbf{r}^*)(\mathbf{x}) - I(\mathbf{r})(\mathbf{x})]^2. \quad (4)$$

Comme dans le cas précédent, il est aussi possible de redéfinir le problème sous la forme d'une maximisation de fonction d'alignement définie grâce à l'information mutuelle. On cherchera alors  $\hat{\mathbf{r}}$  tel que :

$$\hat{\mathbf{r}} = \arg \max_{\mathbf{r}} (MI(T(\mathbf{r}^*), I(\mathbf{r}))) \quad (5)$$

où  $\hat{\mathbf{r}}$  est la position désirée.

### 2.3 Généralisation

On remarque que les équations (3) et (5) sont similaires. Si les fonctions de coût sont très proches l'une de l'autre, l'espace des paramètres sur lequel doit être fait l'optimisation est très différent. Les deux problèmes peuvent alors être généralisés en utilisant dans un premier temps le terme  $\mu$  à la fois pour définir la position du robot  $\mu = \mathbf{r} \in \mathbb{SE}(3)$ , et pour le paramètre de déplacement du problème de suivi, où nous prendrons une homographie pour modéliser le mouvement,  $\mu = \mathbf{p} \in \mathbb{SL}(3)$ .

Dans ce cas, le problème revient à la maximisation du critère non-linéaire suivant :

$$\hat{\mu} = \arg \max_{\mu} (MI(T, I(\mu))). \quad (6)$$

Cette équation est typiquement résolue en utilisant une approche de type Newton. À chaque itération de la maximisation, un incrément  $\Delta\mu$  du paramètre courant  $\mu$  est calculé tel que :

$$\Delta\mu = -\lambda \mathbf{H}^{-1} \mathbf{G}^{\top} \quad (7)$$

où le scalaire  $\lambda$  est un gain appliqué au pas de l'optimisation,  $\mathbf{G}$  est le gradient de l'information mutuelle par rapport aux paramètres  $\mu$  et à  $\mathbf{H}$  sa matrice Hessienne.

Pour l'estimation du mouvement, la mise à jour du paramètre de déplacement courant est alors effectuée jusqu'à convergence :

$$\mu \leftarrow \mu \oplus \Delta\mu \quad (8)$$

où  $\oplus$  définit l'opération qui décrit la combinaison de deux matrices d'homographie.

Pour l'asservissement visuel le paramètre  $\Delta\mu$  correspond à la commande directement envoyée au robot (c'est à dire sa vitesse de translation et de rotation). C'est le système dynamique qui va donc implémenter "naturellement" l'équation (8)

### 3 Information mutuelle et optimisation non-linéaire

#### 3.1 Information mutuelle et théorie de l'information

Dans cette section nous allons définir de façon générale la fonction d'alignement utilisée dans les tâches d'alignement et de positionnement (équation (6)). L'information mutuelle a été proposée par [16] dans le contexte de la théorie de l'information. Sa définition est liée aux définitions de l'entropie et de l'entropie jointe.

**Entropie.** L'entropie  $h(X)$  d'une variable  $X$  est une mesure de sa variabilité. On la retrouve fréquemment en théorie de l'information car elle permet d'exprimer le nombre de bits nécessaire à l'encodage de chaque valeur d'une variable. Si  $x$  représente les valeurs possibles de  $X$  alors l'entropie est donnée par :

$$h(X) = - \sum_x p_X(x) \log(p_X(x)). \quad (9)$$

où  $p_X(x) = P(X = x)$  est la probabilité pour que la variable  $X$  ait la valeur  $x$ . On utilise alors par définition  $0 \log(0) = 0$ .

**Entropie jointe.** L'entropie jointe est en fait l'utilisation de l'entropie sur un système de deux variables. Ainsi l'entropie jointe  $h(X, Y)$  de  $X$  et  $Y$  définit la variabilité du couple de variables  $(X, Y)$  et est donnée par :

$$h(X, Y) = - \sum_{x,y} p_{XY}(x, y) \log(p_{XY}(x, y)) \quad (10)$$

où  $x$  et  $y$  sont respectivement les valeurs possibles de  $X$  et  $Y$  et où  $p_{XY}(x, y) = P(X = x \cap Y = y)$  est la probabilité que le couple de variables  $(X, Y)$  ait la valeur  $(x, y)$ . En théorie de l'information elle représente le nombre théorique de bits nécessaire à encoder les valeurs du couple  $(X, Y)$ .

La recherche du minimum de l'entropie jointe pourrait sembler adéquate pour une tâche d'alignement. Mais un simple exemple permet de montrer le problème que sa définition pose. Considérons un alignement parfait entre  $X$  et  $Y$  quel qu'ils soient. Nous avons alors  $X = Y$  ce qui, en utilisant l'équation (10), nous donne  $H(X, Y) = H(X)$  qui, si  $X$  est fixe, est le minimum de  $H(X, Y)$ . Les deux variables étant identiques, la variabilité du système est égale à la variabilité de l'une de ses composante. Si maintenant nous considérons  $Y = Cste$  alors en utilisant l'équation (10) on trouve de la même façon  $H(X, Y) = H(X)$ . En effet, la variabilité du système correspond à la variabilité de  $X$  puisque  $Y$  est constant. Si l'on minimise l'entropie jointe en faisant varier  $Y$  par exemple, alors  $Y = X$  et  $Y = 0$  sont solutions. Ce qui est dû à la dépendance de l'entropie jointe par rapport aux entropies des deux variables. C'est pour cette raison qu'il est nécessaire de définir l'information mutuelle.

**L'information mutuelle.** La définition de l'information mutuelle permet de résoudre le problème précédent et de créer une fonction d'alignement à partir des notions de variabilités. On définit l'information mutuelle  $MI(X, Y)$  entre  $X$  et  $Y$  de la manière suivante :

$$MI(X, Y) = h(X) + h(Y) - h(X, Y). \quad (11)$$

En utilisant les équations (9) et (10) on peut regrouper les sommes de la manière suivante :

$$MI(X, Y) = \sum_{x,y} p_{XY}(x, y) \log \left( \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \right) \quad (12)$$

Ainsi en soustrayant les entropies des variables avec leur entropie jointe, on enlève la dépendance observée précédemment. L'information mutuelle est alors la quantité d'information partagée par les deux variables. Si l'on maximise cette information alors les deux signaux doivent *a priori* être alignés.

#### 3.2 Information mutuelle sur des images

Dans les paragraphes précédents, l'expression générale de l'information mutuelle a été donnée. Considérons maintenant le cas de l'information mutuelle définie pour un couple d'images. Si dans l'équation (12)  $X$  et  $Y$  sont remplacés par  $T$  et  $I(\mu)$  (définis dans la section 2), alors l'information mutuelle utilisée dans l'équation (6) peut être exprimée par :

$$MI(T, I(\mu)) = \sum_{i,j} p_{ij}(i, j, \mu) \log \left( \frac{p_{ij}(i, j, \mu)}{p_i(i, \mu)p_j(j)} \right). \quad (13)$$

On retrouve les probabilités cette fois-ci définies aux sens de l'image, avec  $p_i(i, \boldsymbol{\mu})$  et  $p_j(j)$  respectivement les probabilités pour qu'un pixel de  $I(\boldsymbol{\mu})$  et  $T$  ait la valeur  $i$  et  $j$ . Et de même  $p_{ij}(i, j, \boldsymbol{\mu})$  la probabilité pour que le couple de valeurs d'un pixel de  $(I(\boldsymbol{\mu}), T)$  ait la valeur  $(i, j)$ . Ces probabilités sont calculées en normalisant les histogrammes et l'histogramme joint des deux images.

Les images ayant 256 niveaux de gris, on a  $((i, j) \in [0; 255]^2 \subset \mathbb{Z}^2)$ . Considérer ces 256 niveaux de gris s'avère coûteux en temps de calcul et provoque une plus forte sensibilité aux bruits. Il est donc classique de normaliser les images pour obtenir un nombre d'entrées plus petit pour chaque histogramme de la manière suivante :

$$\bar{I}(\boldsymbol{\mu})(\mathbf{x}) = I(\boldsymbol{\mu})(\mathbf{x}) \frac{N_c}{N_{c_I}} \quad \bar{T}(\mathbf{x}) = T(\mathbf{x}) \frac{N_c}{N_{c_T}}. \quad (14)$$

où  $N_{c_I}$  et  $N_{c_T}$  sont les niveaux de gris correspondant à  $I$  et  $T$ , et  $N_c$  est le nombre de niveaux de gris maximal des nouvelles images  $\bar{I}$  et  $\bar{T}$ . De cette façon le nombre d'entrées de chaque histogramme est égal à  $N_c$ . On peut finalement exprimer les probabilités comme suit :

$$\begin{aligned} p_{ij}(i, j, \boldsymbol{\mu}) &= \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi [i - \bar{I}(\boldsymbol{\mu})(\mathbf{x})] \phi [j - \bar{T}(\mathbf{x})] \\ p_i(i, \boldsymbol{\mu}) &= \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi [i - \bar{I}(\boldsymbol{\mu})(\mathbf{x})] \\ p_j(j) &= \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi [j - \bar{T}(\mathbf{x})] \end{aligned} \quad (15)$$

où  $N_{\mathbf{x}}$  est le nombre de pixels de la région d'intérêt considérée qui permet de normaliser la somme sur chaque pixel pour obtenir une valeur de probabilité. Classiquement la fonction  $\phi$  utilisée pour le calcul d'histogrammes est une gaussienne. Dans ces travaux les gaussiennes sont approximées par des B-splines qui permettent un calcul aisé de la fonction ainsi que de ses dérivées.

### 3.3 Les dérivées

Afin de mettre en œuvre l'optimisation non linéaire présentée dans la section 2, il est nécessaire de calculer les dérivées du premier et deuxième ordre de l'information mutuelle.

**Gradient.** Le calcul du gradient de l'information mutuelle est obtenu en dérivant l'équation (13). Les détails des calculs peuvent être trouvés dans [5]. Le résultat obtenu est le suivant :

$$\mathbf{G} = \sum_{i,j} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \left( 1 + \log \left( \frac{p_{ij}}{p_i} \right) \right). \quad (16)$$

Il est donc nécessaire de calculer la dérivée de la probabilité jointe en fonction du paramètre  $\boldsymbol{\mu}$ . Cette dérivée est simplement obtenu à partir de l'équation (15) :

$$\frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \frac{\partial \phi}{\partial \boldsymbol{\mu}} (i - \bar{I}(\boldsymbol{\mu})(\mathbf{x})) \phi (j - \bar{T}(\mathbf{x})) \quad (17)$$

avec :

$$\frac{\partial \phi(i - \bar{I}(\boldsymbol{\mu})(\mathbf{x}))}{\partial \boldsymbol{\mu}} = - \frac{\partial \phi(i - \bar{I}(\boldsymbol{\mu})(\mathbf{x}))}{\partial i} \frac{\partial \bar{I}(\boldsymbol{\mu})(\mathbf{x})}{\partial \boldsymbol{\mu}}. \quad (18)$$

L'expression développée de  $\partial \bar{I}(\boldsymbol{\mu})(\mathbf{x}) / \partial \boldsymbol{\mu}$  dépend de l'application. Les détails sont donnés dans la section suivante en fonction du contexte de chacune des deux applications considérées.

**Hessien.** La matrice Hessienne est la dérivée du gradient par rapport au paramètre  $\boldsymbol{\mu}$ . Si l'on applique les lois de dérivation, on obtient [5] :

$$\mathbf{H} = \sum_{i,j} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \left( \frac{1}{p_{ij}} - \frac{1}{p_i} \right) + \frac{\partial^2 p_{ij}}{\partial \boldsymbol{\mu}^2} \left( 1 + \log \frac{p_{ij}}{p_i} \right)$$

Le troisième terme d'ordre 2 est généralement considéré comme nul à convergence, ce qui revient à écrire :

$$\mathbf{H} \simeq \sum_{i,j} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \left( \frac{1}{p_{ij}} - \frac{1}{p_i} \right). \quad (19)$$

Cette approximation est commune dans les articles qui traitent de l'information mutuelle [18, 5, 4]. Elle s'inspire de la similarité avec le calcul du Hessien dans la minimisation de la SSD. L'expérience montre cependant que cette approximation classique n'est pas appropriée. En effet considérons simplement l'équation de la probabilité marginale  $p_i(i)$ , on sait que  $p_i(i) = \sum_j p_{ij}(i, j)$ . Sachant que chaque valeur de probabilité est positive, on a finalement  $p_i(i) > p_{ij}(i, j)$ . Puisque  $\frac{\partial p_{ij}}{\partial \boldsymbol{\mu}} \frac{\partial p_{ij}}{\partial \boldsymbol{\mu}}$  est positive alors la matrice Hessienne donnée par l'approximation (19) est positive. Par définition, on cherche à maximiser l'information mutuelle qui est donc concave à convergence, l'approximation donnée précédemment n'est donc pas adaptée. De plus si l'on considère le cas à convergence avec un alignement parfait, alors  $I = T$ . Ainsi  $p_{ij} = p_i$  pour tout  $i = j$  et  $p_{ij} = 0$  sinon. De plus le gradient de la probabilité jointe est quasiment nulle pour  $i \neq j$ . L'approximation de la matrice Hessienne est donc quasiment nulle à convergence ce qui rend problématique son utilisation dans un processus d'optimisation non linéaire qui utilise son inverse.

Dans nos travaux le calcul du Hessien a donc été entièrement réalisé. Pour cela, comme le montre l'expression (19), la dérivée seconde de la probabilité jointe est nécessaire. En se basant sur la dérivée au premier ordre de  $p_{ij}$  et en appliquant les règles de dérivation, on obtient finalement :

$$\frac{\partial^2 p_{ij}}{\partial (\boldsymbol{\mu})^2} (i, j, (\boldsymbol{\mu})) = \sum_{\mathbf{x}} \frac{\partial^2 \phi}{\partial \boldsymbol{\mu}^2} (i - \bar{I}(\boldsymbol{\mu})(\mathbf{x})) \phi (j - \bar{T}(\mathbf{x})) \quad (20)$$

avec :

$$\frac{\partial^2 \phi}{\partial \boldsymbol{\mu}^2} (i - \bar{I}(\boldsymbol{\mu})(\mathbf{x})) = \frac{\partial^2 \phi}{\partial i^2} \frac{\partial \bar{I}}{\partial \boldsymbol{\mu}} \frac{\partial \bar{I}}{\partial \boldsymbol{\mu}} - \frac{\partial \phi}{\partial i} \frac{\partial^2 \bar{I}}{\partial \boldsymbol{\mu}^2} \quad (21)$$

De la même manière que pour le gradient, le développement de  $\frac{\partial^2 \bar{I}}{\partial \boldsymbol{\mu}^2}$  dépend du contexte et sera donc détaillé dans la partie correspondante.

### 3.4 Application au suivi

On se limitera ici à l'application au suivi d'un plan. Le modèle de déplacement choisi est donc une homographie. Les expressions  $\frac{\partial \bar{I}}{\partial \boldsymbol{\mu}}$  et  $\frac{\partial^2 \bar{I}}{\partial \boldsymbol{\mu}^2}$  peuvent alors être détaillées en fonction du paramètre  $\mathbf{p} \in \mathbb{R}^8$  d'une homographie :

$$\frac{\partial \bar{I}(\mathbf{p})(\mathbf{x})}{\partial \mathbf{p}} = \nabla \bar{I} \frac{\partial w(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}} \quad (22)$$

$$\frac{\partial^2 \bar{I}(\mathbf{p})(\mathbf{x})}{\partial \mathbf{p}^2} = \frac{\partial w^\top}{\partial \mathbf{p}} \nabla^2 \bar{I} \frac{\partial w}{\partial \mathbf{p}} + \nabla \bar{I}_x \frac{\partial^2 w_x}{\partial \mathbf{p}^2} + \nabla \bar{I}_y \frac{\partial^2 w_y}{\partial \mathbf{p}^2}$$

où  $\nabla \bar{I}$  et  $\nabla^2 \bar{I}$  correspondent respectivement au gradient et au gradient de deuxième ordre de l'image. Les dérivés  $\frac{\partial w}{\partial \mathbf{p}}$  et  $\frac{\partial^2 w}{\partial \mathbf{p}^2}$  sont obtenus en appliquant les lois de dérivation à la fonction de transfert  $w(\mathbf{x})$ . Leurs calculs [1] n'apportent pas d'intérêt spécifique dans cet article et ne seront donc pas détaillés.

### 3.5 Application à l'asservissement visuel

Dans le cas de l'asservissement visuel, les expressions  $\frac{\partial \bar{I}}{\partial \boldsymbol{\mu}}$  et  $\frac{\partial^2 \bar{I}}{\partial \boldsymbol{\mu}^2}$  présentes dans l'équation (18) et (21) peuvent donc être détaillées en fonction de la position de la caméra  $\mathbf{r}$  tel que :

$$\frac{\partial \bar{I}(\mathbf{r})(\mathbf{x})}{\partial \mathbf{r}} = \nabla \bar{I} \mathbf{L}_x \quad (23)$$

$$\frac{\partial^2 \bar{I}(\mathbf{r})(\mathbf{x})}{\partial \mathbf{r}^2} = \mathbf{L}_x^\top \nabla^2 \bar{I} \mathbf{L}_x + \nabla \bar{I}_x \mathbf{H}_x + \nabla \bar{I}_y \mathbf{H}_y \quad (24)$$

où  $\nabla \bar{I}$  est le gradient de l'image ( $\bar{I}_x, \bar{I}_y$ ) où encore  $\left( \frac{\partial \bar{I}(\mathbf{x}, \mathbf{r})}{\partial x}, \frac{\partial \bar{I}(\mathbf{x}, \mathbf{r})}{\partial y} \right)$  avec  $x$  et  $y$  exprimés dans le repère métrique,  $\nabla^2 \bar{I}$  est le gradient du gradient.  $\mathbf{L}_x$  est la matrice d'interaction qui lie le déplacement d'un point  $(x, y)$  distant de  $Z$  avec la caméra au déplacement de la caméra  $\mathbf{v}$  (voir [2] pour l'expression détaillée). Aucun modèle de la scène n'est fourni, seule l'image désirée est connue. Une approximation doit donc être faite sur  $Z$ .  $\mathbf{H}_x$  et  $\mathbf{H}_y$  sont respectivement la dérivée de la première et seconde ligne de la matrice d'interaction par rapport au déplacement de la caméra. Le détail de leur obtention est donnée dans [8].

## 4 Résultats expérimentaux

### 4.1 Suivi différentiel

**Recalage.** Un jeu d'expérience de recalage a été réalisé dans l'objectif de justifier la méthode proposée pour calculer le Hessien de l'information mutuelle. Dans cette expérience, le template de l'image 1(d) a été sélectionné. Le paramètre initial de déplacement (vérité terrain) a été modifié en ajoutant une erreur aléatoire aux sommets définissant le template. La fréquence de convergence de l'algorithme classique avec méthode de Brent et celui proposé sans la méthode de Brent ont été étudiés sur 100 expériences d'alignement pour chaque valeur de variance d'erreur entre 1 et

12. On considère que l'algorithme converge si l'erreur finale de positionnement après 50 itérations est inférieure à 1 pixel.

Comme on peut le voir l'algorithme proposé a une fréquence de convergence nettement supérieure à la méthode classique. Une expérience d'alignement est détaillée dans la figure 1 qui en illustre la raison. Le résidu normalisé du recalage a été calculé après chaque itération ainsi que l'erreur de positionnement. Comme on le voit la solution utilisant le Hessien par la méthode proposée arrive à convergence en très peu d'itérations, ce qui est cohérent puisque plus aucune approximation n'est faite. La méthode classique quant à elle converge, mais les oscillations présentes sur l'estimation du déplacement (visibles sur l'erreur de positionnement), ne permettent que plus rarement d'atteindre une erreur inférieure à 1 pixel.

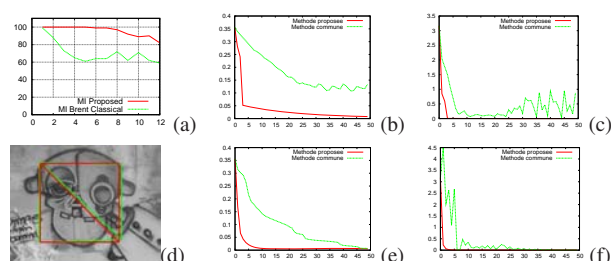


FIG. 1 – Experiences de recalage. Méthode classique en vert et proposée en rouge. (a) Fréquence de convergence en fonction de l'erreur initiale. Détail d'une expérience : (d) template d'origine (rouge) et déformé (vert), (b,e) erreur sur l'image en fonction du nombre d'itérations et (c,f) erreur sur la position des sommets, première ligne : sans méthode de Brent et deuxième ligne : avec méthode de Brent.

Les méthodes d'optimisation de type Brent peuvent être utilisées pour remédier à ces oscillations. Cette méthode consiste à évaluer l'incrément tel qu'il est défini dans l'équation (7) et à l'utiliser avec un gain  $\lambda$  optimal. Ce gain est obtenu par une recherche (coûteuse) à une dimension sur la fonction de coût par interpolation parabolique. Cette méthode a été étudiée avec et sans approximation sur le Hessien. On peut voir l'évolution des erreurs de recalage dans la figure 1. On remarque que l'évolution du résidu est toujours décroissante. Malgré tout, l'évolution avec un calcul exact du Hessien reste meilleure. De plus on observe que si la méthode de Brent permet de limiter les effets d'oscillation sur le résidu, l'erreur sur l'estimation du paramètre de déplacement reste variable.

**Suivi dans une séquence.** Cette section présente les résultats obtenus lorsque l'algorithme de suivi a été testé dans une séquence avec de forts changements d'illumination (avec de fortes spécularités) et dans une séquence où l'image de référence n'a pas la même modalité d'acquisition que la séquence elle-même.

Pour la première expérience représentée dans la figure 2, l'image de référence est définie manuellement dans la pre-

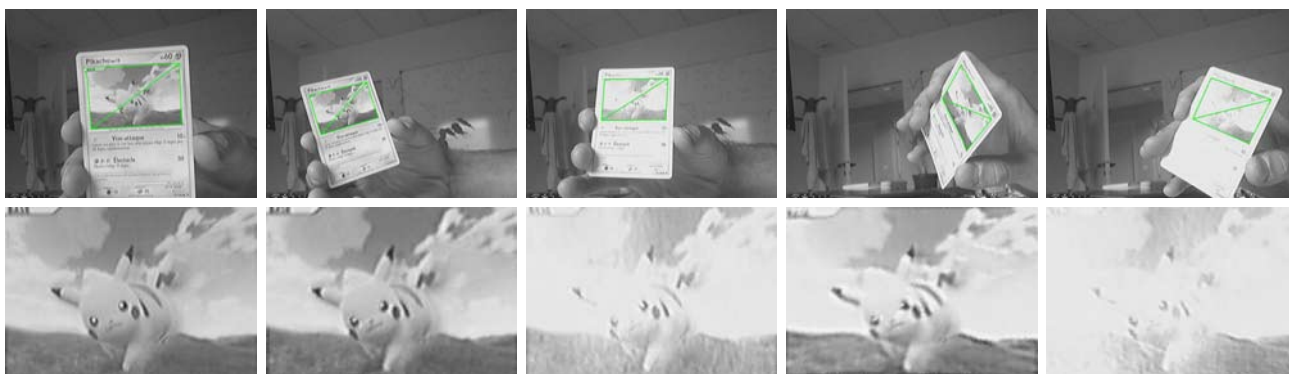


FIG. 2 – Suivi d'un plan dans une séquence d'image avec de forte variation de luminosité. Première ligne : image 0, 78, 110, 212 et 241. Le rectangle vert représente le template de référence transféré sur l'image courante à partir de l'homographie estimée. Deuxième ligne : projection de l'image courante sur l'image référence.

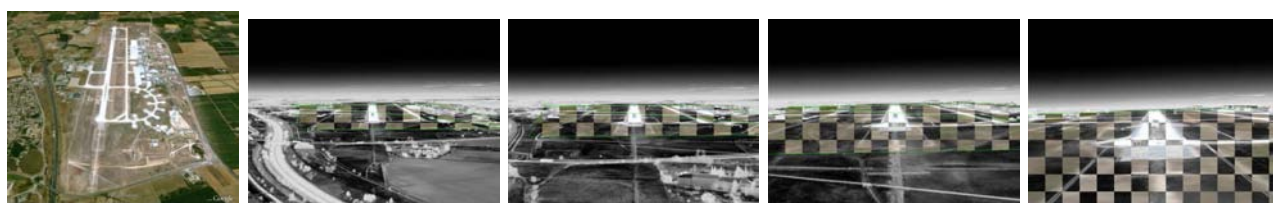


FIG. 3 – Suivi d'une image de référence de type satellite (à gauche) sur une séquence acquise en infrarouge. À gauche les images 1, 250, 500 et 750 avec l'image de référence recalée (images infrarouge de Thales Optronic et satellite de google earth).

mière image de la séquence. La zone d'intérêt sélectionnée comprend 33000 pixels. La première ligne montre les images de la séquence avec la zone d'intérêt projetée en utilisant le paramètre de déplacement estimé. La deuxième ligne représente l'image courante projetée sur la zone d'intérêt. On peut voir que le plan est suivi avec précision. Malgré les grandes variations d'illumination présente sur la zone suivie, l'image reste recalée correctement.

Dans la deuxième séquence représentée dans la figure 3 l'image de référence est une image satellite (issue de google earth). La transformation de cette image vers la première image d'une séquence acquise dans la partie infrarouge du spectre a été sélectionnée manuellement. Le recalage entre la référence et la séquence est ensuite faite automatiquement par la technique de suivi présentée dans cet article et reposant sur l'information mutuelle. Comme on peut le voir sur ces quelques images le recalage se fait correctement. Conséquence du caractère multimodal des images concernées, un suivi basé sur la différence des images (SSD) n'a pas de sens et diverge dès la première itération.

## 4.2 Asservissement visuel

**Expérience de positionnement.** Dans cette partie nous présentons les résultats obtenus pour des tâches de positionnement d'un robot à 6 degrés de liberté. L'image  $T$  est acquise à la position désirée du robot. Cette image correspond à la position  $r^*$  du robot. Cette position  $r^*$  est mémo-

risée dans le seul but de connaître la vérité terrain et ainsi de pouvoir calculer la transformation  $\Delta r$  décrivant l'erreur entre la position courante et la position désirée et qui mesure donc l'erreur de positionnement. L'algorithme a été implanté sur un PC pourvu d'un processeur 2.4 GHz. Sa vitesse d'exécution permet de l'utiliser de manière fluide à cadence vidéo (22ms par itération pour une image de  $320 \times 240$ ).

Dans une première expérience la caméra a été positionnée avec  $\Delta r_0 = (-15\text{cm}, -23\text{cm}, 30\text{cm}, -19^\circ, -11^\circ, 13^\circ)$  ce qui correspond à un mouvement relativement important de la caméra sur les 6 axes. La figure 4 montre les images acquises par la caméra ainsi que l'évolution de  $\Delta r$  au cours du temps. Malgré le large écart initial, le robot converge vers la position désirée et l'erreur finale tend vers 0. L'erreur finale enregistrée par le robot est de l'ordre de 0.1mm en translation et  $0.01^\circ$  en rotation ce qui est très précis sachant que la distance entre la scène et la caméra est d'environ 1 mètre. De plus comme on peut le voir sur la figure 5, la trajectoire de la caméra est très satisfaisante : au début le centre optique de la caméra décrit une ligne dans l'espace 3D jusqu'à ce qu'il ne reste qu'un couplage d'erreurs entre les transformations  $(t_x, r_y)$  ou  $(t_y, r_x)$  alors la caméra décrit une rotation autour de la scène jusqu'à la position désirée.

Une deuxième expérience similaire montre les résultats obtenus en présence de forts changements (non globaux) d'illumination. L'image désirée est acquise puis le robot est

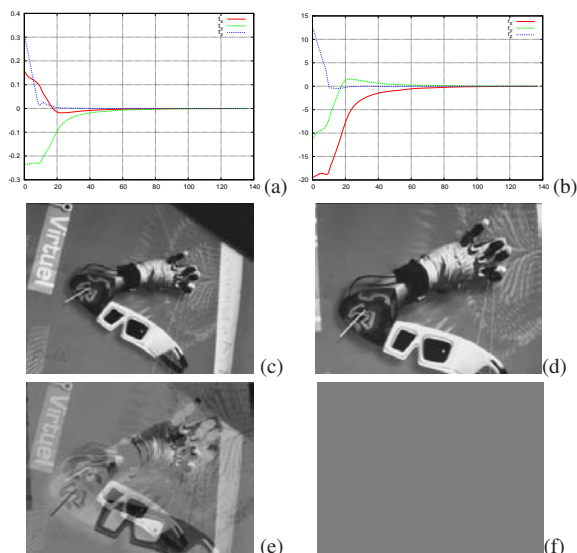


FIG. 4 – Expérience d’asservissement visuel utilisant l’information mutuelle. (a) erreur de translation en mètre, (b) erreur de rotation en (°), (c) image initiale, (d) image désirée, (e) différence d’image initiale entre l’image courante et l’image désirée et (f) différence d’image finale.

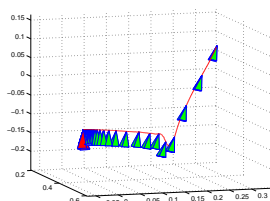


FIG. 5 – Trajectoire 3D de la caméra (axes sont en mètres). La position désirée de la caméra est représentée en rouge.

déplacé à sa position initiale et l’illumination de la scène est changée. On peut voir les images acquises ainsi que le changement d’illumination appliqué sur la figure 6. L’erreur de positionnement final d’ordre de 0.4mm en translation et de 0.05° en rotation est tout à fait satisfaisante. Une loi de commande d’asservissement visuel basée sur la différence des images [3] aurait dans ce cas divergé.

**Expérience de navigation.** Dans cette expérience la robustesse de l’information mutuelle est testée concernant la multimodalité d’acquisition. Pour cela nous considérons une tâche de navigation correspondant à une succession de tâches de positionnement. Une séquence d’images désirées  $T(t)$  est acquise sur une scène où se trouve une carte IGN. L’image aérienne correspondante à la carte IGN est positionnée sur la scène du robot de façon à ce que la position initiale du robot soit proche de la position correspondante à  $T(t_0)$ . Alors la tâche de navigation est lancée en utilisant pour image désirée  $T(t_0)$ . Lorsque le robot peut être considéré comme arrivé à convergence, c’est à dire lorsque le gradient de l’information mutuelle devient trop faible,

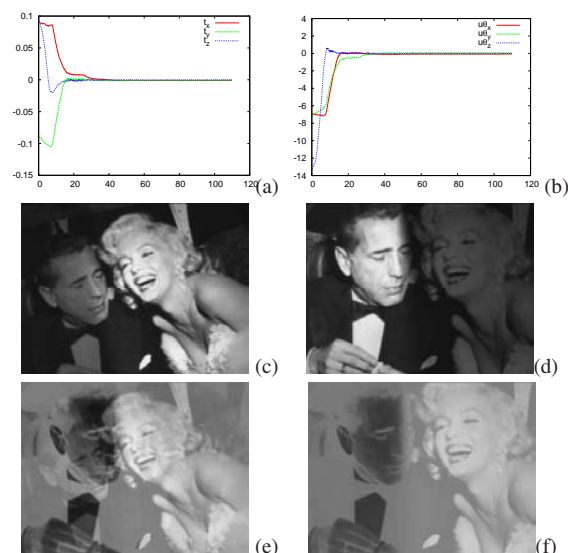


FIG. 6 – Expérience d’asservissement avec changement d’illumination. (a) erreur de translation en mètre, (b) erreur de rotation en (°), (c) image initiale, (d) image désirée, (e) différence d’image initiale entre l’image courante et l’image désirée et (f) différence d’image finale.

alors l’image désirée suivante est prise en compte. Pour évaluer le résultat obtenu nous avons représenté la trajectoire de référence et la trajectoire obtenue sur la figure 7. Certaines images désirées et courantes sont également représentées dans cette figure avec leur superposition qui permet d’en faire la comparaison. Comme on peut le voir dans ces deux figures, l’asservissement visuel reposant sur l’information mutuelle permet de faire naviguer (en utilisant les 6 ddl) le robot sur une trajectoire 3D de manière précise, ce qui montre sa robustesse face aux images multimodales.

## 5 Conclusion

Nous avons présenté dans cet article une utilisation de l’information mutuelle pour l’asservissement visuel et pour le suivi. L’information mutuelle qui est une fonction reconnue pour sa robustesse aux changements d’illumination et à la multimodalité est utilisée ici dans un cadre unifié permettant de proposer une solution aux deux problèmes. L’intérêt de ces travaux pour le suivi et pour la tâche de positionnement d’un robot est démontré et validé par des expériences qui confirment la robustesse de l’information mutuelle aux variations d’illumination et aux différentes modalités considérées. Recaler une image infrarouge avec une image aérienne ainsi que positionner un robot sur une image satellite par rapport à une carte est alors possible et cela avec précision.

**Remerciements** Amaury Dame est financé par la DGA dans le cadre d’une bourse de thèse DGA-CNRS. Les images infrarouge de la figure 3 nous ont été fournies par Thalès-Optronique.

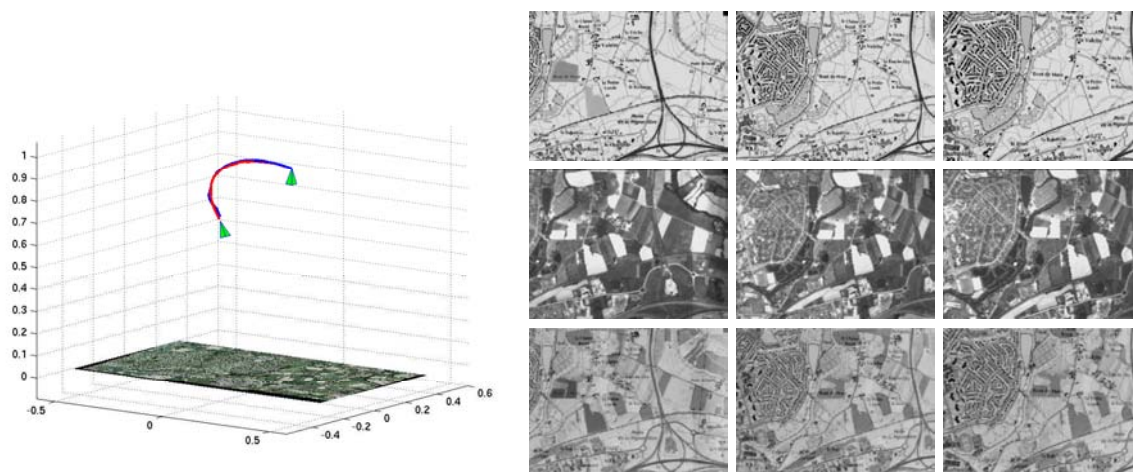


FIG. 7 – Navigation par asservissement visuel multimodal. Première ligne : images désirées ; deuxième ligne : images courantes ; troisième ligne : images désirées superposées aux l'images courantes (images satellite et cartes de l'IGN).

## Références

- [1] S. Benhimane and E. Malis. Homography-based 2d visual tracking and servoing. *Int. Journal of Computer Vision*, 26(7) :661–676, July 2007.
- [2] F. Chaumette and S. Hutchinson. Visual servoing and visual tracking. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*, chapter 24, pages 563–583. Springer, 2008.
- [3] C. Collewet, E. Marchand, and F. Chaumette. Visual servoing set free from image processing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'08*, Pasadena, CA, May 2008.
- [4] N. Dowson and R. Bowden. Mutual information for lucas-kanade tracking (milk) : An inverse compositional formulation. *IEEE Trans. on PAMI*, 30(1) :180–185, January 2008.
- [5] N.D.H. Dowson and R. Bowden. A unifying framework for mutual information methods for use in non-linear optimisation. In *ECCV'06*, volume 1, pages 365–378, June 2006.
- [6] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE PAMI*, 20(10) :1025–1039, October 1998.
- [7] K.S. Kim, J.H. Lee, and J.B. Ra. Robust multi-sensor image registration by enhancing statistical correlation. In *Int. Conf. on Information Fusion, FUSION'05*, pages 380–386, July 2005.
- [8] J.T. Lapresté and Y. Mezouar. A hessian approach to visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'04*, pages 998–1003, Sendai, Japan, September 2004.
- [9] 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.
- [10] E. Marchand and F. Chaumette. Virtual visual servoing : a framework for real-time augmented reality. In G. Drettakis and H.-P. Seidel, editors, *EUROGRAPHICS'02 Conf. Proceeding*, pages 289–298, Saarbrücken, Germany, September 2002.
- [11] E. Marchand and F. Chaumette. Feature tracking for visual servoing purposes. *Robotics and Autonomous Systems*, 52(1) :53–70, June 2005.
- [12] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4) :348–365, December 1995.
- [13] G. Panin and A. Knoll. Mutual information-based 3d object tracking. *Int. Journal of Computer Vision*, 78(1) :107–118, 2008.
- [14] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual-information-based registration of medical images : a survey. *IEEE Trans. on Medical Imaging*, 22(8) :986–1004, August 2003.
- [15] J.-P. Pons, R. Keriven, O. Faugeras and G. Hermosillo. Variational stereovision and 3D scene flow estimation with statistical similarity measures. *ICCV'03*, 2, p597-602, Nice, France, Oct 2003.
- [16] CE. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.
- [17] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. In *IEEE Computer Vision and Pattern Recognition, CVPR'07*, Minneapolis, USA, June 2007.
- [18] P. Thévenaz and M. Unser. Optimization of Mutual Information for Multiresolution Image Registration. *IEEE trans. on Image Processing*, 9(12) :2083–2099, 2000.
- [19] P. Viola and W. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2) :137–154, 1997.