

Cartographie haute résolution à partir d'images stéréoscopiques acquises à basses altitudes

High resolution terrain mapping using low altitude aerial stereo imagery

Il-Kyun Jung

Simon Lacroix

LAAS/CNRS

7, Avenue du Colonel Roche

F-31077 Toulouse Cedex 4

Pré-Nom.Nom@laas.fr

Résumé

Cet article présente une approche qui permet de construire un modèle numérique de terrain de très haute résolution, à partir d'une séquence de vues stéréoscopiques aériennes acquises à basse altitude. L'approche exploite une méthode d'estimation précise du déplacement entre deux acquisitions successives, basée sur la mise en correspondance de points d'intérêt détectés dans les images. Un filtre de Kalman étendu estime alors les 6 paramètres de position du banc stéréoscopique et les positions des points d'intérêt détectés et mémorisés au fur et à mesure des acquisitions. Des résultats exploitant des images acquises avec un ballon dirigeable sont présentés, et montrent la possibilité de construire des modèles numériques de terrain spatialement cohérents.

Mots Clef

Stéreo vision, points d'intérêt, filtrage de Kalman

Abstract

This paper presents an approach to build high resolution digital elevation maps (DEM) from a sequence of unregistered low altitude stereovision image pairs. The approach first uses a visual motion estimation algorithm that determines the 3D motions of the cameras between consecutive acquisitions, on the basis of visually detected and matched environment features. An extended Kalman filter then estimates both the 6 position parameters and the 3D positions of the memorized features as images are acquired. The paper gives details on the filter implementation, and depicts the estimation of the uncertainties on the feature observations and motion estimations. Experimental results with stereo images acquired by a blimp are presented, and show that the precision of the method enables to build spatially consistent very high resolution digital elevation maps.

Keywords

Stereovision, interest points, Kalman filter

1 Introduction

La principale difficulté pour construire des cartes de terrain à très haute résolution à l'aide d'un capteur embarqué sur un engin aérien est de déterminer *précisément* la position de ce capteur au moment de l'acquisition des données. Les techniques de navigation à l'estime, qui intègrent des données fournies par des capteurs de mouvements (centrales inertielles) ne suffisent pas pour cela. En effet, d'une part les estimées des déplacements entre deux acquisitions de données successives ne sont pas suffisamment précises, et d'autre part elles conduisent intrinsèquement à une augmentation dans le temps de l'erreur de positionnement. Des techniques précises d'estimation de déplacement visuelles ont récemment été proposées dans la littérature [1, 2], mais leurs erreurs se cumulent aussi dans le temps.

La seule solution pour réduire les erreurs de positionnement est de se baser sur la perception d'éléments stables de l'environnement, qui sont détectés, positionnés et *mémorisés* lors des déplacements du véhicule. Dans la communauté des roboticiens, il a rapidement été compris que les problèmes de la cartographie de tels éléments et de l'estimation de la position du robot sont intimement liés, et qu'il doivent par conséquent être résolus de manière unifiée [3]. Ce problème, connu sous le nom de problème SLAM ("Simultaneous Localization And Mapping"), a été largement étudié. Une présentation historique des contributions figure dans [4], et le principe du problème est résumé dans la figure 1 : par opposition aux techniques de "structure from motion" étudiées en vision, le problème SLAM permet de d'estimer *en ligne* la position optimale du robot, au fur et à mesure que des acquisitions de données sont réalisées.

Parmi les différentes approches possibles pour estimer conjointement les positions du robot et des amers, l'utilisation d'un filtre de Kalman est indiscutablement la plus populaire. Il a récemment été prouvé que l'application de cette technique converge dans le contexte du problème SLAM [4]. Quelques récentes contributions traitent de son principal défaut, à savoir sa complexité algorithmique, qui croît

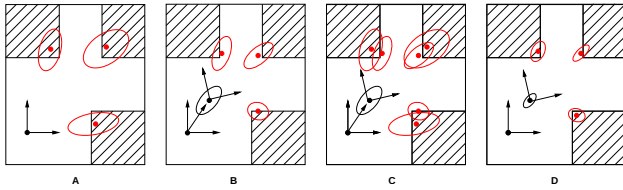


FIG. 1 — Illustration du problème “SLAM” dans le cas d’un robot évoluant dans un plan. En (A), des éléments de l’environnement (appelés “amers” par analogie avec la navigation cotière) sont détectés et positionnés avec une imprécision due au capteur. En (B), le robot se déplace, et mesure son déplacement par une technique de navigation à l’estime, avec une certaine erreur. Les amers sont alors à nouveau observés (C) : grâce à un algorithme de mise en correspondance, ils sont associés aux amers préalablement perçus. Une technique de fusion de données permet alors de réduire les erreurs sur la position courante du robot et les positions des amers (D). Le processus est itéré à chaque nouvelle perception de l’environnement effectuée par le robot.

quadratiquement avec la dimension du vecteur d’état - lequel comprend la position du robot et des amers [5]. D’autres approches ont été proposées, principalement pour s’affranchir du fait que les distributions des erreurs sont gaussiennes, hypothèse indispensable à la mise en oeuvre d’un filtre de Kalman. Ainsi les approches ensemblistes ne nécessitent que la connaissance de bornes dans lesquelles les erreurs sont comprises [6], mais sont délicates à mettre en oeuvre lorsque le nombre de paramètres de position du véhicule excède 3. D’autres techniques ont récemment été appliquées avec succès au problème SLAM [7]. En termes de types de capteurs utilisés, la plupart des validations ont été réalisées à partir de télémètres laser scrutant des environnements intérieurs selon un plan horizontal, pour des robots évoluant sur un sol plan [8, 7, 4]. A notre connaissance, il existe encore peu de contributions au problème SLAM qui exploitent la vision ([9, 10, 11]).

Cet article présente une approche au problème SLAM en trois dimensions (6 paramètres de placement du robot), sur la seule base d’un ensemble de paires stéréoscopiques non localisées, acquises à de basses altitudes. La section suivante présente notre approche du problème, et la section 3 résume les algorithmes de bases utilisés : stéréovision, détection et appariement de points d’intérêt, et estimation visuelle des déplacements. La section 4 détaille notre implémentation du filtre de Kalman, en insistant sur l’identification des différentes erreurs impliquées dans le processus. Enfin, des résultats de localisation et de construction d’un modèle numérique de terrain à partir d’images stéréoscopiques acquises depuis un ballon dirigeable volant à quelques dizaines de mètres d’altitude sont présentés dans la section 5.

2 Principe de notre approche

Les amers sont des *points d’intérêts* qui peuvent être mis en correspondance dans des images acquises de différents points de vue, et dont les coordonnées 3D sont fournies

par un algorithme de stéréovision. La position du robot et des amers est estimée par un filtre de Kalman étendu (EKF), dont le vecteur d’état est constitué de la position du banc stéréoscopique (6 paramètres) et des positions des amers mémorisés (3 paramètres par amer). L’algorithme de base qui permet à la fois l’estimation du mouvement entre deux acquisitions successives et la re-perception d’amers préalablement mémorisés est un algorithme robuste de mise en correspondance de points d’intérêt.

Les différentes étapes réalisées après chaque acquisition d’une paire stéréoscopique sont les suivantes (figure 2) :

1. Stéréovision : une image de points 3D dense est fournie par un algorithme de corrélation de pixels (section 3.1).
2. Détection des points d’intérêts et mise en correspondance avec les points détectés lors de l’acquisition précédente (section 3.2).
3. Sélection des amers : un ensemble de critères de sélection est appliqué sur les points mis en correspondance, afin de les partitionner en 3 ensembles : un ensemble de points “non-amers”, un ensemble “d’amers candidats”, et un ensemble “d’amers observés” (section 4.3).
4. Estimation visuelle du mouvement (EVM) : les points d’intérêts de l’ensemble “non-amers” sont exploités pour estimer les 6 paramètres du déplacement entre l’acquisition précédente et l’acquisition courante. Il s’agit de la phase de prédiction du filtre de Kalman.
5. Affinement de la position globale : il s’agit de la phase de mise à jour du filtre de Kalman, qui ré-estime la position courante du banc stéréoscopique et des amers mémorisés (section 4).

L’étape 3 est nécessaire pour les raisons suivantes : d’une part, seuls les points “non-amers” doivent être exploités pour estimer le déplacement entre deux acquisitions successives, afin d’éviter que les phases de prédiction et de mise à jour du filtre de Kalman soient corrélées. D’autre part, les nouveaux amers doivent être ajoutés avec soin dans le vecteur d’état, afin d’éviter une croissance trop rapide de la dimension de l’état et afin qu’ils couvrent les zones perçues de la manière la plus régulière possible. Enfin, à l’issue de ces étapes, un modèle numérique du terrain est mis à jour à partir des images de points 3D fournies par la stéréovision (section 5.2).

3 Algorithmes de base

Cette section présente brièvement les trois algorithmes à partir desquels le filtrage de Kalman a pu être mis en oeuvre.

3.1 Stéréovision

Nous utilisons un algorithme classique de stéréovision dense basé sur la corrélation de pixels, qui exploite des images acquises à partir d’un banc stéréoscopique préalablement calibré [12]. Les images sont d’abord rectifiées afin que les

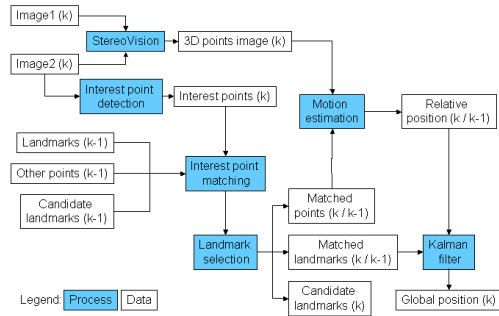


FIG. 2 — Schéma fonctionnel de notre approche du problème de la cartographie et localisation simultanée sur la seule base d’images stéréoscopiques.

lignes épipolaires soient horizontales, ce qui permet d’optimiser considérablement la phase de mise en correspondance des pixels. Une image de disparité est alors produite par la phase de corrélation de pixels : le score utilisé est une corrélation croisée normalisée. Les éventuels faux appariements sont éliminés par une phase dite de corrélation inverse [12], et une estimation sous-pixelique des valeurs de disparités est obtenue par une interpolation parabolique de la courbe de corrélation. Enfin, la phase de triangulation permet de fournir les coordonnées 3D des pixels appariés, avec des incertitudes associées dont la détermination est présentée en section 4.2. L’algorithme s’avère performant pour les images aériennes que nous considérons : la figure 3 montre que la quasi totalité des pixels perçus dans les deux images sont correctement mis en correspondance.

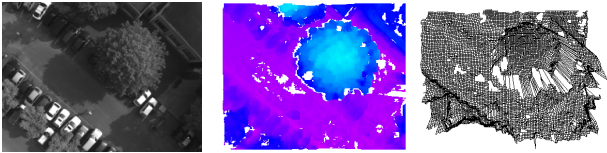


FIG. 3 — Un résultat de l’algorithme de stéréovision avec une paire d’images prises à une altitude d’environ 30 m. De gauche à droite : une des images originales, image de disparités obtenue (les disparités sont inversement proportionnelles à la profondeur des pixels, et sont montrées ici selon une échelle bleu/proche rouge/lointain), et enfin image de points 3D, représentée ici par un maillage pour des questions de lisibilité.

3.2 Détection et appariement de points d’intérêt

Les amers visuels doivent être invariants par translation, rotation, changement d’échelle, et variation de luminosité et de point de vue. Les *points d’intérêt* tels que fournis par le détecteur de Harris, ont de bonnes propriétés de stabilité [13]. Si une estimée du changement d’échelle entre deux images est connue, même approximativement, il est possible d’adapter la détection des points d’intérêt à ce

changement, de manière à ce que la répétabilité soit suffisante pour permettre la mise en correspondance des points [14]. Dans notre cas, le changement d’échelle entre deux paires consécutives est toujours négligeable, et lorsque des amers sont re-perçus après de long déplacements, quand le véhicule “ferme une boucle” par exemple, le changement d’échelle entre l’image courante et l’image où ils ont été initialement perçus peut être connu grâce à l’estimée de la position courante.

Pour mettre en correspondance les points d’intérêts détectés dans deux images, nous utilisons un algorithme d’appariement qui se base sur des groupements locaux de points d’intérêts, combinant ainsi des informations relatives au signal et à la géométrie des images. Les étapes de l’algorithme sont les suivantes (plus de détails figurent dans [15]) :

1. À partir d’un point sélectionné au hasard dans la première image, des hypothèses d’appariement sont générées sur la base d’une mesure de similarité entre les vecteurs propres de la matrice d’autocorrélation du détecteur de Harris
2. Des groupements locaux de points sont déterminés autour du point considéré et des appariements potentiels, en considérant les n voisins les plus proches¹.
3. Des hypothèses d’appariement entre les voisins sont générées avec la même mesure de similarité qu’en 1, et une estimation de la transformation affine entre les groupes considérés est déterminée. La transformation qui produit la plus grande répétabilité est confirmée par une mesure de corrélation entre les points.
4. Une fois qu’une hypothèse d’appariement entre deux groupes est confirmée, les étapes 1 à 3 sont répétées, en commençant par les points les plus proches du groupement considéré. La transformation affine estimée pour le premier appariement de groupes est exploitée pour focaliser la recherche de nouveaux appariements.

Le processus stoppe lorsque plus aucun appariement ne peut être déterminé. La figure 4 montre que ce algorithme permet la détermination de nombreux appariements, même lorsque le changement de point de vue entre les deux images considérées est conséquent.

3.3 Estimation visuelle du mouvement

Les points mis en correspondance dans deux images successives et les coordonnées 3D déterminées par l’algorithme de stéréovision fournissent donc des appariements de points 3D, qui sont utilisés pour estimer les 6 paramètres du déplacement, en appliquant la technique de minimisation par moindres carrés présentée dans [16]. Un aspect important est d’éliminer les mauvais appariements. Ceux qui sont dus à l’algorithme de mise en correspondance de points pourraient être rejetés en utilisant une estimation robuste de la matrice fondamentale, mais une telle méthode ne permettrait pas d’éliminer les mauvais appariements dus aux

1. $n = 6$ dans notre algorithme.

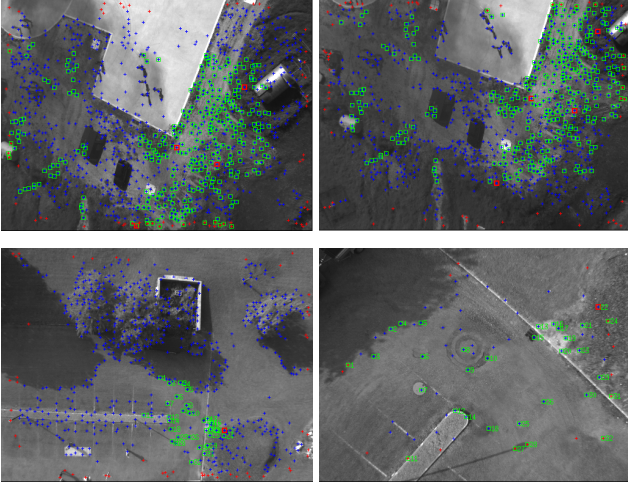


FIG. 4 — Deux résultats de notre algorithme de mise en correspondance de points d'intérêts. Les appariements déterminés sont représentés en vert

erreurs de la stéréovision, telles que celles qui apparaissent le long de discontinuités de profondeur par exemple. En effet, de telles erreurs font qu'un appariement correct de points d'intérêt est un mauvais appariement de points 3D. Nous utilisons donc la simple procédure suivante pour estimer les 6 paramètres de déplacement à partir des appariements 3D :

1. Tout d'abord, les appariements qui comprennent un point estimé avec une incertitude (voir section 4.2);
2. La transformation 3D est déterminée par moindres carrés, la moyenne et l'écart type sur le résidu sont déterminés.
3. Les appariements de points 3D dont la distance est supérieure à k fois l'écart type sur les résidus après application de la transformation trouvée sont éliminés.
4. Une nouvelle transformation 3D est déterminée, et la procédure d'élimination est réitérée jusqu'à ce que k vaille 3.

Cette procédure de rejet de mauvais appariements permet d'obtenir une estimation précise des déplacements entre deux acquisitions consécutives de paires d'images stéréoscopiques (voir résultats dans les sections 4.2 et 5.1), qui est utilisée pour l'étape de prédiction du filtre de Kalman (section 4.1).

4 Mise en œuvre du filtre de Kalman

4.1 Équations du filtre

Le filtre de Kalman étendu est une extension du filtre de Kalman linéaire standard, qui linéarise les équations de prédiction et d'observation autour de l'état prédit. L'évolution du système et son observation sont décrits par :

$$x(k+1) = f(x(k), u(k+1)) + v(k+1)$$

$$z(k) = h(x(k)) + w(k)$$

où $u(k)$ est une entrée de contrôle, et v, w sont les erreurs non corrélées dans le temps sur le modèle du système et sur son observation, de moyenne nulle et de covariances $P_v(k), P_w(k)$.

Le filtre de Kalman consiste en trois étapes : prédiction de l'état du système, observation de l'état système, et mise à jour des estimées de l'état du système et des covariances associées.

Application dans notre contexte. Dans notre approche, l'état du filtre est composé des 6 paramètres de position du banc stéréoscopique $x_p = [\phi, \theta, \psi, t_x, t_y, t_z]$ et des positions de l'ensemble des N amers mémorisés $m_i = [x_i, y_i, z_i]$, $0 < i \leq N$:

$$x(k) = [x_p, m_1 \cdots m_N]$$

La matrice de covariances du système a donc la forme suivante :

$$P(k) = \begin{bmatrix} P_{pp}(k) & P_{pm}(k) \\ P_{pm}^T(k) & P_{mm}(k) \end{bmatrix}$$

où P_{pp} représente les covariances sur la position du banc stéréoscopique, P_{mm} les covariances sur les amers et P_{pm} les covariances entre la position du banc et des amers.

Prédiction. Sous l'hypothèse que les positions des amers ne changent pas dans le temps, la prédiction de l'état est donnée par :

$$\hat{x}(k+1 | k) = f(k)(\hat{x}(k), u(k+1))$$

où $u(k+1) = (\Delta\phi, \Delta\theta, \Delta\psi, \Delta t_x, \Delta t_y, \Delta t_z)$ est le résultat de l'estimation visuelle du mouvement entre les deux positions k et $k+1$. Les covariances sur cet état prédit sont :

$$P_{pp}(k+1 | k) = \nabla_p f(k) P_{pp}(k) \nabla_p f^T(k) + \nabla_u f(k) R_u(k) \nabla_u f^T(k) + P_v(k+1) \quad (1)$$

$$P_{pm}(k+1 | k) = \nabla_p f(k) P_{pm}(k)$$

$$P_{mm}(k+1 | k) = P_{mm}(k)$$

où R_u représente les *covariances des erreurs sur l'estimation visuelle du mouvement*. Notons que la covariance sur les amers reste inchangée après l'étape de prédiction.

Observation. Lorsque l'amer i déjà mémorisé est à nouveau perçu, l'équation d'observation et la jacobienne correspondante sont :

$$\hat{z}_i(k+1 | k) = h_i(k)(\hat{x}(k+1 | k))$$

$$\nabla h_i(k) = [\nabla_p h_i(k), 0 \cdots 0, \nabla_{m_i} h_i(k), 0 \cdots 0]$$

où $h_i(k)(\hat{x}(k+1 | k))$ est la fonction de l'état prédit du robot et du i^{eme} amer dans le vecteur d'état qui passe de

l'espace d'état à l'espace d'observation. L'innovation et la covariance associées à cette observation sont :

$$\nu_i(k+1) = \mathbf{z}_i(k+1) - \hat{\mathbf{z}}_i(k+1 | k) \quad (2)$$

$$\mathbf{S}_i(k+1) = \nabla \mathbf{h}_i(k) \mathbf{P}(k+1 | k) \nabla \mathbf{h}_i^T(k) + \mathbf{R}_i(k+1) \quad (3)$$

où $\mathbf{R}_i$ représente la covariance de l'erreur d'observation du i^{eme} amer.

Mise à jour. L'étape de mise à jour du filtre fusionne la prédiction et l'observation en une nouvelle estimée de l'état et estime les covariances associées selon les formules suivantes :

$$\hat{\mathbf{x}}(k+1 | k+1) = \hat{\mathbf{x}}(k+1 | k) + \mathbf{K}_i(k+1) \nu_i(k+1) \quad (4)$$

$$\mathbf{P}(k+1 | k+1) = \mathbf{P}(k+1 | k) - \mathbf{K}_i(k+1) \mathbf{S}_i(k+1) \mathbf{K}_i^T(k+1) \quad (5)$$

où $\mathbf{K}_i(k+1) = \mathbf{P}(k+1 | k) \nabla \mathbf{h}_i^T(k) \mathbf{S}_i^{-1}(k+1)$ est la matrice de gain du filtre de Kalman. Si aucune observation n'a été faite (i.e. si aucun des amers préalablement perçus n'a été re-perçu), les étapes d'observation et de mise à jour n'ont pas lieu : l'état du système et les covariances associées sont simplement mis à jour par la phase de prédiction.

Lorsqu'un nouvel amer est détecté, il est ajouté à l'état du système qui devient (sa taille augmente donc de 3 unités) :

$$\hat{\mathbf{x}}(k) = [\hat{\mathbf{x}}_p(k), \hat{\mathbf{m}}_1(k) \cdots \hat{\mathbf{m}}_N(k), \hat{\mathbf{m}}_{N+1}(k)]$$

Le modèle d'initialisation d'un amer est :

$$\hat{\mathbf{m}}_{N+1}(k) = \mathbf{g}(k)(\hat{\mathbf{x}}_p(k), \mathbf{z}_{N+1}(k)) \quad (6)$$

$$\mathbf{P}(k) = \begin{bmatrix} \mathbf{P}_{pp}(k) & \mathbf{P}_{pm}(k) & \mathbf{P}_{pz}(k)^T \\ \mathbf{P}_{pm}^T(k) & \mathbf{P}_{mm}(k) & \mathbf{P}_{mz}(k)^T \\ \mathbf{P}_{pz}(k) & \mathbf{P}_{mz}(k) & \mathbf{P}_{zz}(k) \end{bmatrix} \quad (7)$$

où

$$\mathbf{P}_{pz}(k) = \nabla_p \mathbf{g}(k) \mathbf{P}_{pp}(k), \mathbf{P}_{mz}(k) = \nabla_p \mathbf{g}(k) \mathbf{P}_{pm}(k)$$

$$\mathbf{P}_{zz}(k) = \nabla_p \mathbf{g}(k) \mathbf{P}_{pp}(k) \nabla_p \mathbf{g}^T(k) + \nabla_z \mathbf{g}(k) \mathbf{R}_m(k) \nabla_z \mathbf{g}^T(k)$$

où $\mathbf{z}_{N+1}(k)$ représente le nouvel amer, $\mathbf{g}(k)$ représente la fonction d'initialisation basée sur l'estimation courante de la position du robot, et $\mathbf{R}_m$ est la covariance de l'erreur sur les coordonnées du nouvel amer.

4.2 Identification des erreurs

Il est fondamental d'identifier le plus correctement possible les différentes erreurs intervenant dans les équations du filtre. Dans notre contexte, les erreurs suivantes doivent être connues :

- l'erreur d'initialisation des amers ($\mathbf{R}_m$),
- l'erreur d'observation des amers ($\mathbf{R}_i$ pour l'amer i),
- et enfin l'erreur sur u , i.e. sur l'estimation visuelle du mouvement ($\mathbf{R}_u$).

Notons que l'erreur sur le modèle d'évolution du système est mise à 0, les amers étant fixes et la prédiction de l'évolution de la position du robot étant directement obtenue à partir de la position précédente et de l'estimée fournie par l'estimation visuelle du mouvement.

Erreurs d'initialisation des amers. Les amers sont détectés et appariés dans les images vidéo, leurs coordonnées 3D étant déterminées par stéréovision : la matrice de covariances $\mathbf{R}_m$ d'un amer nouvellement détecté est donc totalement définie par le modèle d'erreur de la stéréovision. Nous avons déterminé ce modèle à partir d'une analyse empirique qui a permis de déterminer les erreurs sur les disparités calculées, en analysant la répartition des disparités sur un ensemble de 100 paires stéréoscopiques acquises de la même position. Pour un pixel donné, la répartition des disparités estimées suit bien une distribution gaussienne ([17] - figure 5). Mais un aspect important est qu'il existe une corrélation assez nette entre la forme de la courbe de corrélation autour du maximum retenu et l'écart type sur la distribution : plus le pic est marqué, plus l'écart type est faible (figure 5). Cette relation est la base de notre modèle d'erreur : lors de la phase de corrélation des pixels, un écart type σ_d est associé à chaque disparité d calculée en fonction de la courbure de la courbe de corrélation à son maximum.

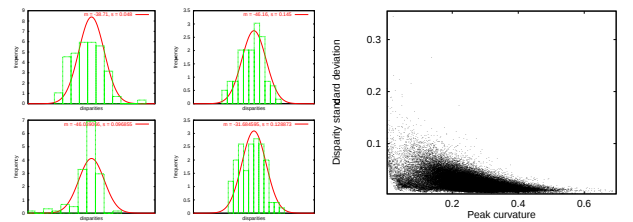


FIG. 5 – À gauche : exemples de densités de probabilité de disparités pour un pixel donnée déterminées à partir de 100 images, avec la gaussienne correspondante. À droite, écarts types sur les disparités en fonction de la courbure du score de corrélation à son pic.

Les coordonnées 3D des pixels corrélés sont déterminées à partir des relations de triangulation usuelles :

$$z = \frac{b\alpha}{d} \quad x = \beta_u z \quad y = \gamma_v z \quad (8)$$

où z est la profondeur, b l'écartement entre les deux caméras du banc stéréoscopique, et α , β_u et γ_v sont des paramètres

de calibration du banc (les deux derniers dépendant de (u, v) , position du pixel considéré dans l'image). Une approximation au premier ordre donne :

$$\sigma_z^2 \simeq \left(\frac{\partial z}{\partial d}\right)^2 \sigma_d^2 = \frac{(b\alpha)^2}{d^4} \sigma_d^2$$

en substituant la définition de z , nous avons :

$$\sigma_z = \frac{\sigma_d}{b\alpha} z^2$$

ce qui est une propriété bien connue de la stéréovision, à savoir que les erreurs croissent quadratiquement avec la profondeur, et sont inversement proportionnelles à l'écartement des caméras. La matrice de covariance des coordonnées du point est alors dérivée des équations de triangulation :

$$\mathbf{R}_m = \begin{bmatrix} 1 & \beta_u & \gamma_v \\ \beta_u & \beta_u^2 & \beta_u \gamma_v \\ \gamma_v & \beta_u \gamma_v & \gamma_v^2 \end{bmatrix} \left(\frac{\sigma_d}{b\alpha} z^2\right)^2$$

Quand un nouvel amer est détecté, ses coordonnées sont ajoutées à l'état du filtre, et les covariances sont déterminées en fonction des équations (6) et (7).

Erreur d'observation. Dans notre cas, l'observation des amers est basée sur la mise en correspondance des points d'intérêts. Deux types d'erreurs peuvent survenir à ce niveau : mauvais appariements et erreurs sur la localisation des points d'intérêt dans le plan image. Les mauvais appariement étant supposés éliminés par la procédure présentée en 3.3, seules les erreurs sur la position des points dans le plan image et sur la détermination de leurs coordonnées 3D interviennent pour déterminer les erreurs d'observation. La version précise du détecteur de Harris permet d'obtenir les coordonnées sous-pixelliques d'un point d'intérêt $\mathbf{p}$. Lorsque ce point est à nouveau perçu dans une image proche (dans deux images consécutives notamment), nous fixons empiriquement l'erreur d'appariement à 0.5 pixels. Par contre, lorsque les points de vue sont très différents, lors de la re-perception d'un amer après un long déplacement par exemple, la déformation projective est plus importante : nous fixons dans ce cas l'erreur d'appariement à 1.0 pixels (ce qui est consistant avec l'erreur maximale de 1.5 pixel généralement tolérée pour valider un appariement de points d'intérêt [13]).

Cette erreur d'appariement dans le plan image est combinée avec les erreurs sur les coordonnées des points 3D correspondants. Le principe de cette combinaison est illustré figure 6 : l'erreur d'observation est définie par la projection de l'erreur d'appariement dans le plan image sur les points 3D. Lorsque l'erreur d'appariement est de 1 pixels, les coordonnées 3D d'un point observé $\mathbf{p}_0$ sont pondérées par celles de ses 8 plus proches voisins :

$$\bar{\mathbf{X}} = \frac{1}{9} \sum_{i=0}^8 \mathbf{X}_i, \quad \sigma_{\bar{\mathbf{X}}}^2 = \frac{1}{9} \sum_{i=0}^8 (\bar{\mathbf{X}} - \mathbf{X}_i)^2 + \sigma_i^2$$

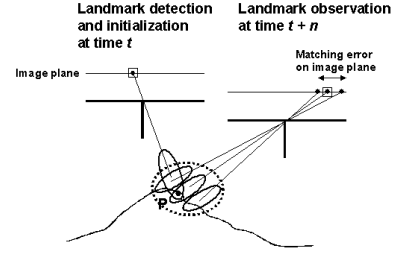


FIG. 6 — Principe de la combinaison des erreurs d'appariement et de stéréovision. Les petites ellipses représentent les erreurs de la stéréovision, la grande ellipse en pointillé est l'erreur d'observation résultante.

où $\mathbf{X}_0$ et $\mathbf{X}_k$ sont les coordonnées 3D de $\mathbf{p}_0$ et de ses voisins, et σ_0 et σ_k sont les variances associées.

Quand l'erreur d'appariement dans le plan image est de 0.5 pixels, les coordonnées des points 3D fournies par la stéréovision étant déterminées pour des valeurs entières des pixels, nous supposons que la variation de la surface locale est linéaire. Les coordonnées de l'amer $\mathbf{p}_0$ observé et les variances associées sont alors :

$$\bar{\mathbf{X}} = \frac{1}{9} (\mathbf{X}_0 + \sum_{i=1}^8 (\frac{\mathbf{X}_0 + \mathbf{X}_i}{2}))$$

$$\sigma_{\bar{\mathbf{X}}}^2 = \frac{1}{9} \sum_{i=0}^8 \left(\left(\frac{\bar{\mathbf{X}} - \mathbf{X}_i}{2} \right)^2 + \left(\frac{\sigma_0 + \sigma_i}{2} \right)^2 \right)$$

Ces coordonnées et les variances associées sont exploitées dans les équations (2) et (3).

Erreurs sur l'estimation de mouvement. Étant donné un ensemble de points 3D appariés

$$\hat{\mathbf{Q}} = [\mathbf{X}_1, \dots, \mathbf{X}_N, \mathbf{X}'_1, \dots, \mathbf{X}'_N]$$

la fonction de coût qui est minimisée pour déterminer le déplacement associé est la suivante [16]:

$$J(\hat{\mathbf{u}}, \hat{\mathbf{Q}}) = \sum_{n=1}^N (\mathbf{X}'_n - R(\hat{\phi}, \hat{\theta}, \hat{\psi}) \mathbf{X}_n - [\hat{t}_x, \hat{t}_y, \hat{t}_z]^T)^2$$

où $\hat{\mathbf{u}} = (\hat{\phi}, \hat{\theta}, \hat{\psi}, \hat{t}_x, \hat{t}_y, \hat{t}_z)$. $\hat{\mathbf{u}}$ et $\hat{\mathbf{Q}}$ peuvent être écrits comme :

$$\hat{\mathbf{u}} = \mathbf{u} + \Delta \mathbf{u}, \quad \hat{\mathbf{Q}} = \mathbf{Q} + \Delta \mathbf{Q}$$

où les vraies grandeurs $\mathbf{u}$ et $\mathbf{Q}$ sont inconnues. En supposant que l'estimée du déplacement optimale maximise la fonction de coût et sa jacobienne, les incertitudes sur les positions des points $\mathbf{X}_n$ et leur observation $\mathbf{X}'_n$ peuvent être propagées à l'aide d'un développement au premier ordre de la jacobienne autour de $\mathbf{u}$ et $\mathbf{Q}$, comme expliqué dans [18] : en considérant que $\mathbf{X}_n$ et $\mathbf{X}'_n$ ne sont pas corrélées, l'estimée de la covariance $\mathbf{P}_{\hat{\mathbf{u}}}$ peut-être écrite comme :

$$\mathbf{P}_{\hat{\mathbf{u}}} = \left(\frac{\partial g}{\partial \mathbf{u}}(\hat{\mathbf{u}}, \hat{\mathbf{Q}}) \right)^{-1} (\Lambda_{\mathbf{X}} + \Lambda_{\mathbf{X}'}) \left(\frac{\partial g}{\partial \mathbf{u}}(\hat{\mathbf{u}}, \hat{\mathbf{Q}}) \right)^{-1}$$

où $g = \frac{\partial J}{\partial \mathbf{u}}$ est la jacobienne de la fonction de coût, et

$$\Lambda_{\mathbf{X}} = \sum_{n=1}^N \frac{\partial g}{\partial \mathbf{X}_n}(\hat{\mathbf{u}}, \mathbf{X}_n) \mathbf{P}_{\mathbf{X}_n} \left(\frac{\partial g}{\partial \mathbf{X}_n}(\hat{\mathbf{u}}, \mathbf{X}_n) \right)^T$$

$$\Lambda_{\mathbf{X}'} = \sum_{n=1}^N \frac{\partial g}{\partial \mathbf{X}'_n}(\hat{\mathbf{u}}, \mathbf{X}'_n) \mathbf{P}_{\mathbf{X}'_n} \left(\frac{\partial g}{\partial \mathbf{X}'_n}(\hat{\mathbf{u}}, \mathbf{X}'_n) \right)^T$$

$\mathbf{P}_{\hat{\mathbf{u}}} = \mathbf{R}_u$ est la matrice de covariance utilisée dans l'équation (1) pour estimer la variance sur l'état à l'issue de l'étape de prédiction.

Les moyennes des variances calculées sur l'estimation des 6 paramètres du déplacement et leur dispersion pour une cinquantaine de déplacements de quelques mètres estimés sont présentées dans le tableau 1. L'estimation visuelle du mouvement mesure des déplacements de quelques mètres avec une précision de l'ordre de 1 % en translation, et de l'ordre de 0.1° en rotation, et ce avec une assez bonne régularité.

	Φ	Θ	Ψ	t_x	t_y	t_z
Erreurs moy.	0.098	0.089	0.037	3.6	3.8	1.1
σ des erreurs	0.024	0.030	0.008	1.3	1.1	0.3

TAB. 1 — Statistiques sur les erreurs estimées de l'EMV calculées sur environ 50 déplacements de quelques mètres (en degrés et centimètres).

4.3 Sélection des amers

Comme expliqué en section 2, les appariements de points 3D fournis entre deux images consécutives sont séparés en trois ensembles. L'ensemble des amers observés correspond simplement aux points qui figurent déjà dans le vecteur d'état du filtre. Les appariements restants sont étudiés afin de déterminer les amers candidats selon les trois critères suivants :

- **Observabilité.** De bon amers doivent être observés dans plusieurs images consécutives,
- **Stabilité.** Les coordonnées d'un amer doivent être les plus précises possible,
- **Représentativité.** Les amers doivent couvrir le plus régulièrement possible la scène observée.

Parmi l'ensemble des points appariés qui ne sont pas des amers ré-observés, seul un pourcentage assez faible² va être étudié : une procédure heuristique de satisfaction de ces critères permet de sélectionner ceux qui vont être conservés comme amers. Les autres sont utilisés pour estimer le mouvement.

5 Résultats

Nos développements ont été testés à partir de centaines d'images acquises à bord d'un ballon dirigeable évoluant

2. nous utilisons 10 %

à des altitudes de 20 à 40 m. Les caméras du banc stéréoscopique de 2.2 m de largeur sont des caméras numériques avec un capteur CCD d'un demi-pouce, de résolution 1024 × 768 pixels, et sont dotées d'objectifs de 4.8 mm de longueur focale (champ de vue de 67° × 53°). Les images sont sous-échantillonnées d'un facteur 2 avant d'être traitées.

5.1 Erreurs de localisation

Ne disposant pas d'un système de localisation absolu précis (tel qu'un GPS centimétrique à différence de phase par exemple), l'évaluation des erreurs de positionnement est faite ainsi : lorsque le ballon "ferme une boucle", l'estimation visuelle du déplacement appliquée entre la première image de la séquence et l'image courante est considérée comme étant la référence. Elle est en effet suffisamment précise par rapport au cumul des erreurs des estimations de mouvements effectuées tout au long de la séquence.

La figure 7 présente une comparaison entre les trajectoires reconstruites par simple cumul des estimations locales de déplacement et reconstruite en appliquant le filtre de Kalman à partir d'une quarantaine d'images, et la figure 8 montre l'évolution des écarts types sur les 6 paramètres de localisation du ballon. On distingue deux phases dans cette dernière figure : jusqu'à la 25^{ème} étape, les écarts types croissent (mais cependant plus lentement que si seule l'EVM était appliquée pour déterminer la position du ballon). Dans les étapes suivantes, quelques amers préalable-ment détectés sont re-perçus : les écarts types des erreurs décroissent.

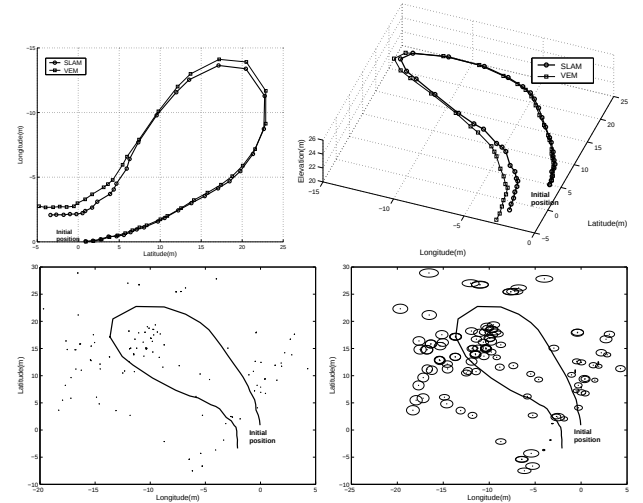


FIG. 7 — Un premier résultat de l'application du filtre sur une séquence de 40 paires stéréoscopiques, prises à des altitudes autour de 25 m. Les images en haut représentent la trajectoire reconstituée par la technique d'EVM et le filtre (projection orthogonale et vue 3D). Les images du bas montrent les 120 amers détectés et mémorisés, avec les ellipses d'incertitude correspondantes (taille réelle à gauche, grossies d'un facteur 40 à droite).

Le tableau 5.1 compare les résultats de l'estimation de la position finale entre l'EVM seule et l'application du filtre.

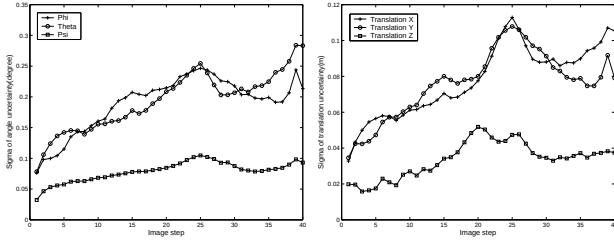


FIG. 8 — Évolution des écarts types de la position du banc stéréoscopique durant le vol montré figure 7.

	Référence : Écart type	EVM : Erreur absolue	SLAM : Ecart type	SLAM : Erreur absolue
Φ	0.10°	3.30°	0.21°	0.20°
Θ	0.09°	2.40°	0.28°	0.61°
Ψ	0.04°	0.56°	0.09°	0.11°
t_x	0.04 m	0.91 m	0.11 m	0.24 m
t_y	0.04 m	1.47 m	0.08 m	0.07 m
t_z	0.01 m	0.91 m	0.04 m	0.10 m

TAB. 2 — Comparaison de la localisation finale entre le cumul des EVM et l'application du filtre, la référence étant l'EVM calculée directement entre les images 1 et 40

L'amélioration apportée par le filtre est claire : les erreurs de translation sont inférieures à 0.1 m au bout d'une trajectoire d'environ 60 m , et les erreurs angulaires sont inférieures à un demi degré.

Les figures 9 à 10 montrent les résultats lors d'une trajectoire durant laquelle 400 images ont été acquises, les comparaisons numériques étant présentées dans le tableau 3. Dans ce cas, une diminution notable des erreurs de positionnement est observée à la fin de la trajectoire, lorsque le ballon perçoit à nouveau les amers détectés au début de la trajectoire.

	Reference : Écart type	EVM : Erreur absolue	SLAM : Ecart type	SLAM : Erreur absolue
Φ	0.79°	0.15°	0.94°	0.14°
Θ	0.65°	9.02°	1.12°	0.44°
Ψ	0.10°	3.57°	0.67°	0.40°
t_x	0.33 m	11.22 m	-2.81 m	0.23 m
t_y	0.40 m	2.57 m	1.53 m	0.24 m
t_z	0.08 m	16.27 m	3.48 m	0.08 m

TAB. 3 — Comparaison de la localisation finale entre le cumul des EVM et l'application du filtre pour la trajectoire de la figure 9, la référence étant l'EVM calculée directement entre les images 1 et 400.

5.2 Modèle numérique de terrain

Grâce à l'estimation de la localisation des caméras fournie par le filtre, il est possible de construire un modèle

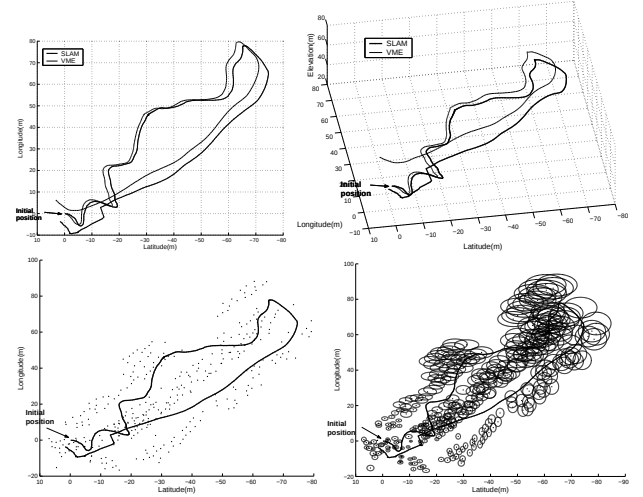


FIG. 9 — Un autre résultat de trajectoire estimée à partir d'une séquence de 400 images stéréoscopiques (même conventions que dans la figure 7).

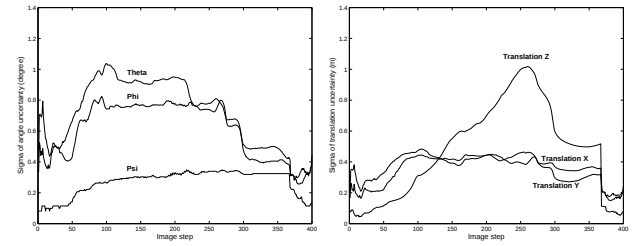


FIG. 10 — Évolution des écarts types sur la position du banc stéréoscopique lors de la trajectoire montrée figure 9.

numérique du terrain qui décrit l'environnement survolé par une fonction $z = f(x, y)$, déterminée sur chaque cellule (x_i, y_i) d'une grille cartésienne régulière.

Nous utilisons pour cela un algorithme extrêmement simple, qui calcule l'élévation z des points d'une cellule par moyennage de tous les points qui se projettent verticalement sur la surface qu'elle couvre. L'écart type de l'élévation est aussi calculé, et comme une valeur de luminance est associée à chaque point 3D par la stéréovision, il est aussi possible de déterminer une valeur moyenne de luminance pour chaque cellule. La figure 11 montre un modèle numérique construit à partir d'une centaine d'images : la résolution du modèle est de 10 cm , et l'orthoimage correspondante ne fait pas apparaître de décalage notable.

6 Conclusions

Nous avons présenté un moyen d'estimer précisément *en ligne* la position d'un banc stéréoscopique évoluant à quelques dizaines de mètres d'altitude, en appliquant une technique de cartographie et localisation simultanée basée sur une filtre de Kalman. Il s'agit à notre connaissance d'une des premières applications de ce type de technique en utilisant *exclusivement* des informations fournies par la vision. L'utilisation de points d'intérêts pour détecter et mettre en correspondance les amers permet à l'algorithme de fonc-

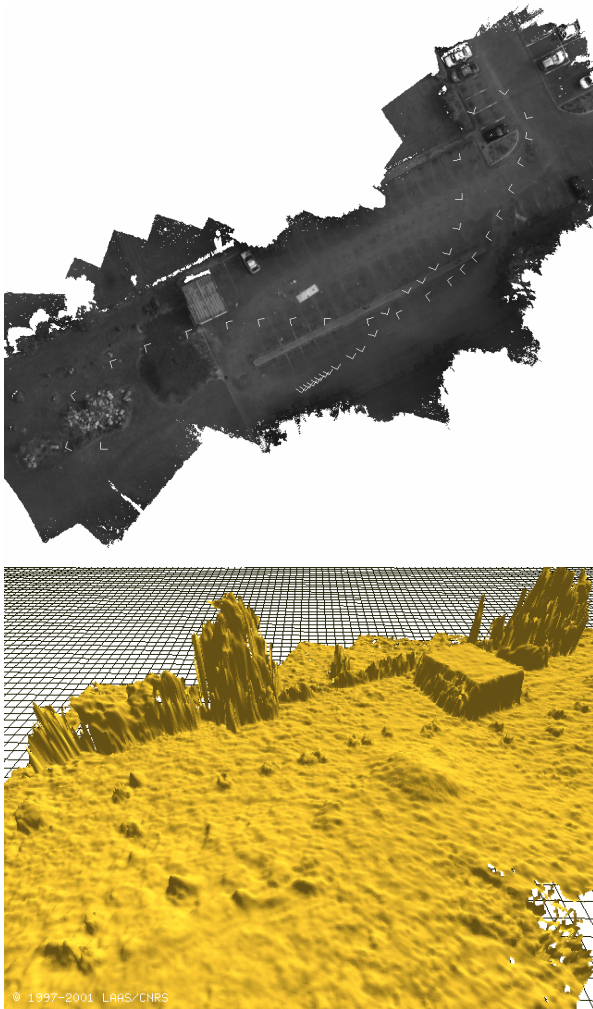


FIG. 11 — Modèle numérique de terrain construit à partir d'une centaine d'images : orthoimage et vue 3D de la partie gauche de l'orthoimage. Le modèle couvre une surface d'environ 3500 m^2 .

tionner sans que des amers artificiels soient présents dans l'environnement. En particulier, l'efficacité et la robustesse de l'algorithme de mise en correspondance des points d'intérêts permet la ré-observation des amers détectés, mêmes avec des points de vue sensiblement différents. la détermination des différentes erreurs intervenant dans le processus a permis de mettre en oeuvre une implémentation du filtre sans avoir recours à une phase empirique de "tuning" du filtre.

Références

- [1] A. Mallet, S. Lacroix, and L. Gallo. Position estimation in outdoor environments using pixel tracking and stereovision. In *IEEE International Conference on Robotics and Automation, San Francisco, Ca (USA)*, pages 3519–3524, April 2000.
- [2] C. Olson, L. Matthies, M. Schoppers, and M. Maimone. Stereo ego-motion improvements for robust rover navigation. In *IEEE International Conference on Robotics and Automation*, pages 1099–1104, May 2001.
- [3] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *Robotics Research: The Fourth International Symposium, Santa Cruz (USA)*, pages 468–474, 1987.
- [4] G. Dissanayake, P. M. Newman, H-F. Durrant-Whyte, S. Clark, and M. Csorba. A solution to the simultaneous localization and map building (slam) problem. *IEEE Transaction on Robotic and Automation*, 17(3):229–241, May 2001.
- [5] J. Guivant and E. Nebot. Optimization of the simultaneous localization and map building algorithm for real time implementation. *IEEE Transactions on Robotics and Automation*, 17(3):242–257, June 2001.
- [6] M. Kieffer, L. Jaulin, E. Walter, and D. Meizel. Robust autonomous robot localization using interval analysis. *Reliable Computing*, 6(3):337–362, Aug. 2000.
- [7] S. Thrun, W. Burgard, and D. Fox. A real-time algorithm for mobile robot with applications to multi-robot and 3d mapping. In *IEEE International Conference on Robotics and Automation, San Francisco, CA (USA)*, 2000.
- [8] O. Wijk and H.I. Christensen. Triangulation based fusion of sonar data for robust robot pose tracking. *IEEE Transactions on Robotics and Automation*, 16(6):740–752, 2000.
- [9] S. Se, D. Lowe, and J. Little. Local and global localization for mobile robots using visual landmarks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, Maui, Hawaii (USA)*, pages 414–420, October 2001.
- [10] A. Davison and. Real-time simultaneous localisation and mapping with a single camera. In *IEEE International Conference on Computer Vision, Nice (France)*, pages 1403–1410, Oct. 2003.
- [11] D. Nister. An efficient solution to the five-point relative pose problem. In *IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wi. (USA)*, June 2003.
- [12] O. Faugeras, T. Vieville, E. Theron, J. Vuillemin, B. Hotz, Z. Zhang, L. Moll, P. Bertin, H. Mathieu, P. Fua, G. Berry, and C. Proy. Real-time correlation-based stereo : algorithm, implementations and application. Technical Report RR 2013, INRIA, August 1993.
- [13] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Sixth International Conference on Computer Vision, Bombay (India)*, pages 230–235, Jan. 1998.
- [14] Y. Dufournaud, C. Schmid, and R. Horaud. Matching images with different resolutions. In *International Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC (USA)*, pages 612–618, Juin 2000.
- [15] I-K. Jung and S. Lacroix. A robust interest point matching algorithm. In *8th International Conference on Computer Vision, Vancouver (Canada)*, July 2001.
- [16] R. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim. Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1426–1446, Nov/Dec 1989.
- [17] L. Matthies. Toward stochastic modeling of obstacle detectability in passive stereo range imagery. In *IEEE International Conference on Computer Vision and Pattern Recognition, Champaign, Illinois (USA)*, pages 765–768, 1992.
- [18] R.M. Haralick. Propagating covariances in computer vision. In *International Conference on Pattern Recognition*, pages 493–498, 1994.