

Mise en correspondance par corrélation avec prise en compte des occultations

Occlusion handling in correlation-based matching

Sylvie Chambon, Alain Crouzil

Institut de Recherche en Informatique de Toulouse, IRIT, Université Paul Sabatier, UPS,
118, route de Narbone, 31062 Toulouse Cedex 9
chambon@irit.fr, crouzil@irit.fr

Manuscrit reçu le 22 septembre 2006

Résumé et mots clés

En stéréovision binoculaire, la mise en correspondance est une étape cruciale pour réaliser la reconstruction 3D de la scène. De très nombreuses publications traitent ce problème. Ainsi, le premier objectif est de proposer un état de l'art des méthodes de mise en correspondance. Nous synthétisons cette étude en présentant un algorithme générique complet faisant intervenir des éléments constitutifs permettant de décrire les différentes étapes de la recherche de correspondances. Une des plus grandes difficultés, au cours de l'appariement, provient des occultations. C'est pourquoi le second objectif est de présenter un état de l'art des méthodes qui prennent en compte cette difficulté. Enfin, le dernier objectif est de présenter de nouvelles méthodes hybrides, dans le cadre des méthodes locales à base de corrélation. Nous nous appuyons sur l'utilisation de deux mesures de corrélation permettant de mieux prendre en compte le problème des occultations. Les résultats mettent en évidence la meilleure méthode qui consiste à fusionner deux cartes de disparités obtenues avec des mesures différentes.

Stéréovision, mise en correspondance, occultations, corrélation.

Abstract and key words

In binocular stereovision, the accuracy of the 3D reconstruction depends on the accuracy of matching results. Consequently, matching is an important task. Our first goal is to present a state of the art of matching methods. We define a generic and complete algorithm based on essential components to describe most of the matching methods. Occlusions are one of the most important difficulties and we also present a state of the art of methods dealing with occlusions. Finally, we propose matching methods using two correlation measures to take into account occlusions. The results highlight the best method that merges two disparity maps obtained with two different measures.

Stereovision, matching, occlusions, correlation.

1. Introduction

En vision par ordinateur, le but est de trouver, à partir d'images acquises par un ou plusieurs capteurs, des informations relatives aux formes, au mouvement et au relief de la scène. La reconnaissance du relief concerne l'extraction automatique d'informations sur la structure 3D de la scène à partir d'une image, de deux images (stéréovision binoculaire) ou d'un ensemble d'images. Nous nous plaçons dans le cas de la stéréovision binoculaire où trois étapes fondamentales permettent de retrouver le relief de la scène : le calibrage, la mise en correspondance et la reconstruction 3D. La mise en correspondance, appelée aussi l'appariement, s'avère être une tâche délicate dont la qualité du résultat détermine directement celle de la reconstruction 3D. Elle est plus difficile en présence de changements de luminosité, d'objets dont la texture est uniforme, de bruit et d'occultations.

La mise en correspondance stéréoscopique consiste à retrouver dans les images gauche et droite, les éléments homologues, c'est-à-dire les éléments qui sont la projection de la même entité de la scène. Les éléments appariés peuvent être des pixels ou des primitives structurées de l'image. L'utilisation de primitives structurées limite la combinatoire lors de la recherche des correspondances et offre la possibilité d'utiliser des attributs discriminants permettant de lever des ambiguïtés. En revanche, il est nécessaire de détecter au préalable ces primitives structurées. De plus, leur mise en correspondance ne permet pas d'obtenir une reconstruction dense de la scène. Dans le cadre de cet article, nous nous intéressons à la mise en correspondance dense de pixels.

Un des problèmes les plus délicats en vision par ordinateur est le problème des occultations. De très nombreux chercheurs ont tenté de prendre en compte cette difficulté. Dans de précédents travaux [19], nous avons d'ailleurs proposé de nouvelles mesures de corrélation robustes aux occultations. Les résultats obtenus avec ces mesures sont les meilleurs dans les zones des occultations alors que dans les zones sans occultation, ces mesures n'obtiennent pas toujours les meilleurs résultats. Ainsi, nous proposons de nouvelles méthodes s'appuyant sur l'utilisation d'une mesure classique dans les zones sans occultation et d'une mesure robuste dans les zones des occultations.

Dans un premier temps, notre objectif est de répertorier les approches qui peuvent être appliquées au cas de la mise en correspondance stéréoscopique de pixels et de les caractériser par leurs éléments constitutifs, puis de proposer une nouvelle classification. Par la suite, après avoir distingué les différentes manières de prendre en compte les occultations, nous introduisons les méthodes proposées. Les résultats exposés en dernière partie mettent en évidence la meilleure méthode qui consiste à fusionner deux cartes des disparités.

2. Appariement de pixels

Il existe de nombreuses publications sur la mise en correspondance et il est difficile d'en donner un état de l'art exhaustif [39, 17]. Nous adoptons une décomposition des méthodes en éléments constitutifs qui interviennent dans un algorithme générique complet.

2.1. Éléments constitutifs

Généralement, les méthodes de mise en correspondance sont décrites sous la forme d'un problème de minimisation d'une fonction d'énergie, le coût global, ou de plusieurs fonctions d'énergie, les coûts locaux. L'état de l'art le plus abouti est celui de Scharstein et Szeliski [80] qui distinguent quatre éléments constitutifs pour caractériser les méthodes de mise en correspondance stéréoscopique de pixels¹. Nous avons complété ces travaux en distinguant les éléments constitutifs suivants :

- les *primitives à appairer et leurs attribut* (pixels ou primitives structurées [53]);
- le *prétraitement des images* (tente notamment de supprimer le bruit);
- la *coût global de mise en correspondance* qui détermine à quel point les correspondances sont en accord avec le modèle choisi (ce coût comprend un coût de correspondance faisant intervenir des coûts locaux et un coût des contraintes composé de coûts de voisinage);
- la *zone d'agrégation* (pixels voisins des pixels étudiés qui sont pris en compte pour calculer les différents termes du coût global de mise en correspondance);
- le *support* (ensemble des pixels pour lesquels on cherche un correspondant);
- la *zone de recherche* (pour chaque pixel du support, il s'agit des pixels candidats);
- la *méthode d'optimisation* (permet de trouver la valeur minimale du coût global et de déterminer ainsi les couples de pixels retenus comme correspondants);
- l'*affinement des résultats* (tente de corriger *a posteriori* d'éventuelles erreurs);
- les *passages multiples* dans le cas où la méthode effectuée plusieurs fois la mise en correspondance avec des techniques différentes pour combiner les avantages de chacune;
- la *multirésolution* (réduit le temps de calcul).

Cette caractérisation semble convenir pour décrire la plupart des méthodes existantes. Certains de ces éléments constitutifs sont essentiels : les primitives, le coût global, le support, la zone d'agrégation et la méthode d'optimisation. Par la suite, nous présentons l'algorithme générique complet qui met en évidence l'utilisation de ces dix éléments constitutifs, puis, nous les décrivons.

¹ <http://vision.middlebury.edu/stereo/>

2.2. Algorithme générique complet

Les méthodes existantes ne passent pas automatiquement par toutes les étapes de l'algorithme 2.1 que nous appelons l'algorithme générique complet. Les étapes 1 et 2e correspondent au schéma multirésolution. L'étape 2a n'est effectuée que par les méthodes qui nécessitent des calculs préliminaires. L'étape 2d n'est réalisée que par les méthodes qui traitent les erreurs d'appariement.

2.3. Définitions et notations

Nous notons I_l , avec $l = g, d$, les images gauche et droite, de taille $N_l = N_l^{\text{lig}} \times N_l^{\text{col}}$. Le pixel de l'image I_l de coordonnées $(i, j)^T$ est noté $\mathbf{p}_l^{i,j}$ et le niveau de gris associé est $I_l^{i,j}$. Le vecteur transposé de \mathbf{v} est noté \mathbf{v}^T et $\|\mathbf{v}\|$ est sa norme euclidienne. Nous nous plaçons dans le cas où la mise en correspondance est effectuée de l'image gauche vers l'image droite mais toutes les explications et les formules peuvent être utilisées dans le cas inverse. Une manière de représenter le résultat d'une mise en correspondance consiste à associer à chaque pixel $\mathbf{p}_g^{i,j}$ de l'image gauche un vecteur appelé disparité, défini par ses composantes $(u - i \quad v - j)^T$ où $(u \quad v)^T$ sont les coordonnées dans l'image droite du point correspondant au pixel $\mathbf{p}_g^{i,j}$. Nous pouvons assimiler le problème de la mise en correspondance à la recherche d'une fonction de disparité d qui attribue une disparité à chaque pixel $\mathbf{p}_g^{i,j}$:

$$d : \mathbb{N}^2 \longrightarrow \mathbb{R}^2 \quad (1)$$

$$\mathbf{p}_g^{i,j} \longmapsto d(\mathbf{p}_g^{i,j}) = (u - i \quad v - j)^T. \quad (2)$$

Lorsqu'une disparité réelle est calculée, il s'agit d'un appariement au sous-pixel. La plupart des méthodes calculent des disparités entières, et dans ce cas, l'expression (1) devient: $d : \mathbb{N}^2 \longrightarrow \mathbb{Z}^2$.

2.4. Contraintes

Une contrainte est une propriété liée à une correspondance qui découle d'hypothèses faites sur: la géométrie du capteur (contrainte épipolaire), la géométrie de la scène (contraintes d'unicité, d'ordre et de symétrie) ou la réflectance de la surface des objets ainsi que le type et la position de la source lumineuse. Cette dernière hypothèse concerne la contrainte de dissimilarité, cf. § 2.7. Nous présentons les contraintes les plus utilisées, en détaillant celles que nous utilisons par la suite.

La *contrainte épipolaire* peut être utilisée pour faciliter la mise en correspondance en réduisant la zone de recherche. Elle permet d'obtenir une relation entre les points de l'image gauche et de l'image droite qui ne dépend que des paramètres des caméras et des coordonnées image et qui est indépendante des coordonnées du point de la scène [4]. La *rectification épipolaire*

Algorithme 2.1. Mise en correspondance stéréoscopique de pixels. La zone d'agrégation et la zone de recherche interviennent aux étapes A et C.

Pour chaque passage faire

1. Construire pyramide d'images
2. Pour chaque niveau (pyramide) faire
 - (a) Prétraiter images
 - (b) Partitionner ensemble des pixels en supports
 - (c) Pour chaque support faire
 - i. Pour chaque itération faire
 - A. Calculer coûts locaux
 - B. Calculer coût de correspondance
 - C. Calculer coûts de voisinage
 - D. Calculer coût des contraintes
 - E. Calculer coût global
 - F. Déterminer correspondances (optimisation)
 - (d) Affiner résultats
 - (e) Propager résultats

consiste à se ramener à une configuration du capteur dans laquelle deux pixels qui se correspondent se trouvent sur la même ligne [43]. Cette configuration simplifie la méthode de recherche des correspondants car la disparité est seulement une différence de colonnes et l'expression (2) devient: $\mathbf{p}_g^{i,j} \longmapsto d(\mathbf{p}_g^{i,j}) = (0 \quad v - j)^T$. À partir de ce paragraphe, nous supposons toujours que nous travaillons avec des **images rectifiées**.

La *contrainte d'unicité*, largement employée en stéréovision, est donnée par:

$$\text{Si } (\mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v})$$

$$\text{alors } \forall j_2 \neq j_1, \mathbf{p}_g^{i,j_2} + d(\mathbf{p}_g^{i,j_2}) \neq \mathbf{p}_d^{i,v}.$$

Si deux pixels différents ont le même correspondant, alors la contrainte d'unicité n'est pas vérifiée. Cette contrainte peut être violée: quand un plan de la scène est très incliné par rapport à l'une des deux caméras, l'effet de raccourcissement peut apparaître et ainsi tous les pixels n'ont pas forcément un correspondant unique.

La *contrainte d'ordre* est très utilisée en stéréovision et elle est définie par:

$$\text{Si } (\mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v_1})$$

$$\text{et } (\mathbf{p}_g^{i,j_2} + d(\mathbf{p}_g^{i,j_2}) = \mathbf{p}_d^{i,v_2})$$

$$\text{alors } (j_1 - j_2)(v_1 - v_2) \geq 0.$$

Elle signifie que l'ordre des pixels de l'image gauche le long de la droite épipolaire doit être le même que celui de leurs correspondants. Cette contrainte peut aussi être violée lorsque, par exemple, l'angle entre le plan image et la direction induite pas les positions des deux objets est important. Les contraintes

d'ordre et d'unicité ont été très utilisées; une variante de ces contraintes, la consistance faible qui permet d'interdire moins de correspondances, a même été proposée dans [58].

La *contrainte de symétrie* ou *consistance forte* [58] est souvent utilisée et elle est définie par :

$$\text{Si } (\mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v}) \text{ alors } \mathbf{p}_d^{i,v} + d(\mathbf{p}_d^{i,v}) = \mathbf{p}_g^{i,j}.$$

Deux mises en correspondance sont effectuées, de la gauche vers la droite puis de la droite vers la gauche. Si un pixel $\mathbf{p}_g^{i,j}$ a pour correspondant le pixel $\mathbf{p}_d^{i,v}$, alors, lors de la seconde mise en correspondance, $\mathbf{p}_d^{i,v}$ doit avoir pour correspondant $\mathbf{p}_g^{i,j}$. La contrainte de symétrie assure la contrainte d'unicité. Cette contrainte est forte et, dans certains travaux [25], les auteurs ont proposé une version qui tolère une erreur de quelques pixels, la symétrie faible.

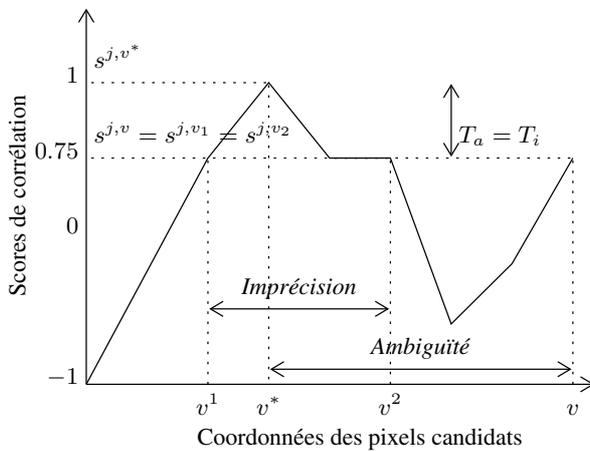


Figure 1. Ambiguïté et imprécision – Nous illustrons ce calcul en utilisant un coût local dont l'intervalle de variation est $[-1; 1]$, en supposant que l'on cherche un maximum et avec $T_a = T_i = 0.25$.

Les *contraintes d'ambiguïté* et *d'imprécision* sont spécifiques au cas des méthodes locales [27], cf. figure 1, et permettent d'évaluer la qualité des résultats obtenus après le calcul des coûts locaux. Les deux critères évalués proviennent d'une analyse de la courbe constituée par les coûts locaux sur la zone de recherche. L'ambiguïté permet de mesurer la gravité d'une grosse erreur que l'on a pu commettre en choisissant un pixel alors qu'il en existe un autre pour lequel le coût local (score de corrélation) était proche. Plus précisément, s'il existe un score de corrélation $s^{j,v}$ proche de $s^{j,v*}$, c'est-à-dire, tel que $|s^{j,v} - s^{j,v*}| \leq T_a$ (T_a est un seuil à fixer) alors l'ambiguïté correspond à $|v - v^*|$. L'imprécision permet de quantifier l'erreur de localisation, c'est-à-dire la gravité d'une petite erreur que l'on a pu commettre. Plus précisément, s'il existe v_1 et v_2 tels que $v^* \in [v_1; v_2]$ et quel que soit $v \in [v_1; v_2]$ le score $s^{j,v}$ est proche du score $s^{j,v*}$, c'est-à-dire, tels que $|s^{j,v} - s^{j,v*}| < T_i$ (T_i est un seuil à fixer) alors l'imprécision correspond à $|v_1 - v_2|$.

Il existe d'autres contraintes, comme la contrainte de continuité figurale [70], la limite du gradient de disparité [50, p. 185] et la contrainte de rang [76].

2.5. Attributs des pixels

Les correspondances peuvent être recherchées pour une partie des pixels – mise en correspondance partielle – ou pour tous les pixels – mise en correspondance dense. Pour la mise en correspondance dense, l'attribut utilisé peut être le niveau de gris (c'est le plus populaire) ou la couleur. Pour la mise en correspondance partielle, on met en correspondance des pixels de l'image qui correspondent à des points particuliers dans la scène, comme par exemple : les points d'intérêt [42, 64, 84, 86] ou les points contour [18, 83].

2.6. Prétraitement des images

Le prétraitement des images est lié aux images utilisées, aux pixels pris en compte et aux attributs associés aux pixels choisis. Les différents cas que nous pouvons distinguer sont les suivants :

- *Pour améliorer les images* – Pour tenter d'éliminer d'éventuels bruits dans les images, avant d'effectuer une détection de pixels particuliers, Kim *et al.* [56] utilisent un laplacien de gaussien.
- *Pour détecter les pixels particuliers de l'image* – Il est nécessaire d'appliquer un opérateur de détection au préalable pour localiser les pixels particuliers et les attributs associés, cf. § 2.5.
- *Pour extraire les attributs des pixels* – Certaines méthodes s'appuient soit sur le calcul des dérivées des images [81, 94], soit sur des transformations de rang [103].
- *Pour effectuer une segmentation en régions* – Ces dernières années ont vu se multiplier les méthodes coopératives qui intègrent une segmentation [48, 97].

2.7. Coûts de mise en correspondance, zones d'agrégation et support

Le *coût global de mise en correspondance* évalue à quel point les correspondances sont valides :

$$E_{\text{global}}(d) = (1 - \lambda)E_{\text{corresp}}(d) + \lambda E_{\text{cont}}(d), \tag{3}$$

où $\lambda \in [0; 1]$ permet d'ajuster l'influence des deux termes : E_{corresp} , le coût de correspondance, terme d'attache aux données, et E_{cont} , le coût des contraintes modélisant les interactions entre les pixels considérés. Résoudre le problème de la mise en correspondance, formulé de cette manière, consiste à trouver la fonction de disparité d qui minimise ce coût E_{global} .

Le *coût de correspondance* évalue à quel point un ensemble de correspondances est fiable ; il est déterminé par la somme sur le

support \mathcal{S} des coûts locaux et il est de la forme :

$$E_{\text{corresp}}(d) = \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$$

avec $\mathbf{p}_d^{i,v} = \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j})$.

Le *coût local* correspond au coût d'une correspondance et il est défini par :

$$E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) = \sum_{\substack{\mathbf{p}_g^{i',j'} \in \text{ZA}(\mathbf{p}_g^{i,j}) \\ \mathbf{p}_d^{i',v'} \in \text{ZA}(\mathbf{p}_d^{i,v})}} E_{\text{diss}}(\mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'}), \quad (4)$$

où $\text{ZA}(\mathbf{p}_g^{i,j})$ est la zone d'agrégation ou fenêtre de corrélation (ensemble de pixels connexes au pixel étudié). Le *coût de dissimilarité* E_{diss} évalue à quel point deux pixels sont différents.

Le *coût des contraintes* modélise les contraintes entre les pixels du support et ceux de la zone d'agrégation :

$$E_{\text{cont}}(d) = \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} E_{\text{vois}}(\mathbf{p}_g^{i,j}).$$

Le *coût du voisinage* permet d'évaluer l'effet de la contrainte sur le pixel étudié et ses voisins :

$$E_{\text{vois}}(\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in \text{ZA}(\mathbf{p}_g^{i,j})} E_{\text{liss}}(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'}). \quad (5)$$

Le *coût de lissage* E_{liss} permet de comparer les disparités associées aux deux pixels considérés.

Une manière très répandue de classer les méthodes de mise en correspondance consiste à distinguer les méthodes locales et globales [16]. Les méthodes locales effectuent N_g minimisations des coûts locaux et la taille de la fenêtre de corrélation est toujours strictement supérieure à un pixel. Les méthodes globales effectuent une minimisation du coût global. Nous reprenons cette classification en la complétant et en distinguant : les méthodes locales, les méthodes globales dont la zone d'agrégation est réduite à un pixel, les méthodes globales dont la zone d'agrégation n'est pas réduite à un pixel.

Les *méthodes locales* sont aussi appelées méthodes par corrélation. Dans cet article, une mesure de corrélation évalue le degré de dissimilarité entre deux ensembles de pixels. Il y a N_g supports, chacun étant réduit à un singleton noté $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$. La zone ZA peut être une fenêtre carrée centrée sur le pixel considéré ou une fenêtre adaptative, cf. § 3.4. Le coût E_{diss} peut être une mesure de corrélation [5, 9, 20] ou d'information mutuelle [45]. Il n'y a pas de coût des contraintes et la disparité d est donnée par une méthode de recherche exhaustive, *winner take all*, notée WTA :

$$d(\mathbf{p}_g^{i,j}) = (0 \quad v - j)^T$$

avec $\mathbf{p}_d^{i,v} = \underset{\mathbf{p}_d^{i,v'} \in \mathcal{Z}_d(\mathbf{p}_g^{i,j})}{\text{argmin}} E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v'})$,

où $\mathcal{Z}_d(\mathbf{p}_g^{i,j})$ correspond à la zone de recherche associée au pixel $\mathbf{p}_g^{i,j}$, cf. algorithme 2.2.

Pour les *méthodes globales*, la zone d'agrégation du coût local est réduite à un pixel et l'équation (4) devient : $E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) = E_{\text{diss}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$. Le support \mathcal{S} peut être l'image entière (une seule optimisation est à réaliser) ou une zone plus restreinte. La plupart des coûts de dissimilarité s'appuient sur une fonction de la différence des niveaux de gris qui peut être : une fonction exponentielle [89], la valeur absolue [14], le carré [57] ou la dissimilarité s'appuyant sur les niveaux de gris interpolés linéairement autour des deux pixels étudiés [10]. Le coût de lissage E_{liss} a pour but de pénaliser les disparités différentes des disparités du voisinage [7, 38]. Il peut utiliser soit des comparaisons avec les disparités voisines [7], soit un terme de pénalité. Celui-ci peut s'appuyer sur un seuil sur la différence de disparité avec le voisinage [12], sur une détection des occultations [87], sur la contrainte d'unicité [69, 101], d'ordre [40] ou de symétrie [65].

Enfin, nous appelons *méthodes mixtes* celles dont le coût global respecte l'équation (4), contrairement aux méthodes globales. Le coût local peut correspondre à une mesure de corrélation [1]. Les contraintes utilisées sont, entre autres, les mêmes que celles décrites dans le § 2.7.

Algorithme 2.2. Méthode locale de mise en correspondance de pixels – C'est une instance de l'algorithme 2.1 (cf. § 2.7). Les étapes B à E ne sont pas effectuées et la boucle 2(c)i est absente.

Pour chaque passage faire

1. Construire pyramide d'images
2. Pour chaque niveau (pyramide) faire
 - (a) Prétraiter images
 - (b) $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$
 - (c) Pour chaque $\mathcal{S}^{i,j}$ faire
 - A. Pour chaque $\mathbf{p}_d^{i,v} \in \mathcal{Z}_d(\mathbf{p}_g^{i,j})$ faire calculer $E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$ suivant $\text{ZA}(\mathbf{p}_g^{i,j})$
 - F. Déterminer correspondant de $\mathbf{p}_g^{i,j}$ selon méthode WTA
 - (d) Affiner résultats
 - (e) Propager résultats

2.8. Méthodes d'optimisation

Pour minimiser le coût global, nous distinguons les méthodes à variables continues [100] (appariement au sous-pixel) des méthodes à variables discrètes (appariement « au pixel » : la disparité est entière). Nous nous intéressons aux méthodes à variables discrètes car, en vision par ordinateur, elles sont les plus utilisées. Une optimisation discrète est réalisée et l'espace des solutions est fini car la zone de recherche est bornée. De

nombreuses méthodes d'appariement ont utilisé des techniques d'optimisation qui font appel à la même modélisation: les champs de Markov aléatoires :

- *Programmation dynamique* – Bellman et Dreyfus [8] en sont les fondateurs et de très nombreuses méthodes de mise en correspondance l'utilisent [30, 36].

- *Recuit simulé* – L'échantillonneur de Gibbs ou l'algorithme de Metropolis [29] ont été utilisés pour la mise en correspondance dans [6, 73].

- *Relaxation* – Davis et Rosenfeld [26] sont les premiers à avoir proposé cette technique en vision par ordinateur.

Pour la mise en correspondance, le principe est exposé dans [105].

- *Coupure de graphe* – Introduite par Roy et Cox [78], l'utilisation de flot de graphe pour la mise en correspondance a été formalisée par Veksler [95] et développée dans [15, 57, 90].

- *Réseaux de neurones* – Hopfield [49] a été le premier à aborder cette technique qui a été adaptée à la mise en correspondance [98].

- *Propagation de croyance* – Cette méthode, *belief propagation*, introduite par Pearl [74] en intelligence artificielle, a été reprise en vision par ordinateur dans [98].

- *Algorithmes génétiques* – Holland [47] est le premier à avoir exposé des algorithmes génétiques fondés sur la théorie de l'évolution de Darwin. Ces algorithmes ont été adaptés à la mise en correspondance dans [79].



2.9. Approche à passages multiples

Nous distinguons les méthodes itératives (optimisation globale) des méthodes à passages multiples (plusieurs calculs des disparités mais avec des techniques différentes). Afin d'obtenir une estimation initiale des disparités, certaines méthodes effectuent une première mise en correspondance locale [12] ou globale [7, 65]. Dans [66], après une première mise en correspondance de points particuliers, la zone de recherche est adaptée pour les autres pixels. Dans [32, 58], une nouvelle forme de fenêtre est déterminée à partir des résultats d'un premier appariement. Dans [67, 106], les coûts locaux sont modifiés en fonction de ceux obtenus à l'étape précédente. Enfin, les méthodes par croissance de germes effectuent une détection de points d'intérêt [62] ou une segmentation [97], mettent en correspondance une partie des pixels puis augmentent progressivement le nombre de pixels.

2.10. Affinement des résultats

Plusieurs techniques ont été proposées pour effectuer un calcul plus fin, au sous-pixel, cf. [91]: les positions des correspondants ne sont plus des valeurs entières mais réelles. La technique la plus répandue utilise une interpolation quadratique des coûts locaux, scores de corrélation obtenus pour les pixels voisins du

correspondant trouvé [85]. Dans certains cas, lorsque l'appariement n'est pas dense, les auteurs décident de calculer une carte dense en interpolant les disparités trouvées [33, 58]. Dans de nombreux articles, l'ensemble des correspondances finales subit des corrections afin d'éliminer des erreurs. Nous pouvons citer deux buts: supprimer des correspondances [56, 88, 104] et corriger les erreurs de disparités [32].

2.11. Approche multirésolution

Le but de la multirésolution est d'utiliser différents niveaux de résolution pour diminuer les temps de calcul. Le niveau le plus grossier de la pyramide d'images ne conserve que les grandes structures et le niveau le plus fin contient tous les détails. On met en correspondance au niveau le plus grossier et on répercute les résultats sur les niveaux plus fins. Il existe des pyramides régulières [17] et des pyramides irrégulières [68] (jamais utilisées pour la mise en correspondance). Les pyramides régulières les plus utilisées sont les pyramides gaussiennes [6, 65], moyennes [23, p. 75-82] et non linéaires [24]. Pendant la propagation, en effectuant la mise en correspondance au niveau n , la méthode employée est contrainte par les résultats du niveau $n + 1$. Dans le cas d'une méthode locale [23, p. 79-82], les disparités fournies au niveau $n + 1$ peuvent permettre de réduire la zone de recherche au niveau n .

2.12. Classification des méthodes

L'élément constituant le plus discriminant est le coût global. De plus en plus de méthodes effectuent plusieurs appariements en séquence avec différentes méthodes pour obtenir de meilleurs résultats qu'en utilisant une seule technique; c'est pourquoi, nous distinguons quatre catégories de méthodes (cf. tableau 1): locales, globales, mixtes et à passages multiples.

3. Étude des occultations

Le problème des occultations au cours de la mise en correspondance étant crucial, nous proposons, d'une part, une étude des méthodes existantes et, d'autre part, de nouvelles méthodes.

3.1. Méthodes avec prétraitements

Une première approche [37] s'appuie sur un masque des occultations. Les méthodes diffèrent par leur façon de le calculer mais, généralement, elles s'appuient sur cette hypothèse: les occultations sont au niveau des contours de type « marche ». En effet, souvent, les discontinuités de profondeur se trouvent aux frontières des objets. Les techniques habituelles de détection des

Tableau 1. Classification des méthodes de mise en correspondance – Nous précisons, entre parenthèses, les années et le nombre de publications pour chaque type de méthodes.

TYPE		MÉTHODE D'OPTIMISATION	RÉFÉRENCES
<i>Méthode locale</i> (1989-2003, 16)		Recherche exhaustive	[5, 9, 25, 31, 33, 34, 37, 46, 55, 63, 72, 81, 93, 94, 99, 103]
<i>Méthode globale</i> (1989-2005, 24)		Recherche exhaustive	[89]
		Programmation dynamique	[10, 12, 14, 30, 36, 38, 104]
		Recuit simulé	[6, 61, 73]
		Relaxation	[105]
		Coupure de graphe	[15, 28, 48, 52, 57, 95]
		Réseaux de neurones	[51, 69]
		Propagation de croyance	[87, 90]
<i>Méthode mixte</i> (1989-2004, 8)		Algorithmes génétiques	[41, 79]
		Programmation dynamique	[92]
		Recuit simulé	[60]
		Relaxation	[3, 11, 82]
		Coupure de graphe	[1, 78]
<i>Méthode à passages multiples</i> (1992-2005, 12)	<i>Locale</i>	Algorithmes génétiques	[40]
		Recherche exhaustive	[32, 58, 66]
		Croissance de germes	[62, 96, 97]
	<i>Globale</i>	Méthode coopérative	[67, 106]
		Programmation dynamique	[56, 70]
		Recuit simulé	[7]
		Relaxation	[65]

Tableau 2. Classification des méthodes de mise en correspondance prenant en compte explicitement le problème des occultations – Nous précisons, entre parenthèses, les années et le nombre de publications.

TYPE	RÉFÉRENCES
<i>Méthode avec prétraitements</i> (2002-2005, 6)	[28, 37, 61, 87, 97, 104]
<i>Méthode locale</i> (1997-2005, 4)	[22, 55, 59, 103]
<i>Méthode globale</i> (1989-2005, 12)	[1, 6, 7, 12, 28, 30, 36, 38, 52, 57, 65, 92]
<i>Méthode locale à passages multiples</i> (1992-2005, 11)	[34, 35, 36, 46, 54, 63, 71, 72, 99, 102, 104]
<i>Méthode avec affinement</i> (1993-2003, 10)	[2, 30, 31, 33, 58, 66, 67, 88, 104, 106]

contours sont le seuillage de la norme du vecteur gradient [75, p. 503] ou la détection des passages par zéro de la dérivée seconde dans la direction du gradient [83]. Une autre méthode consiste à utiliser une segmentation initiale en régions [29, 87]. Les pixels proches des frontières des régions sont supposés être occultés.

3.2. Méthodes locales

Une autre manière de détecter les occultations est d'utiliser les transformations locales [20]. Dans [55, 103], les auteurs utilisent non plus les niveaux de gris mais l'ordre des niveaux de gris pour tenter de prendre en compte le problème des occulta-

tions. Dans [22, 59], les auteurs utilisent des outils de statistiques robustes pour détecter les pixels occultés en considérant leur niveau de gris comme des données aberrantes lors de l'estimation de la transformation qui relie les niveaux de gris des voisinages des pixels homologues (les pixels des fenêtres de corrélation comparées).

3.3. Méthodes globales

Les méthodes globales qui prennent en compte le problème des occultations ajoutent une contrainte d'occultation dans le coût des contraintes. Nous distinguons les possibilités suivantes faisant intervenir : la limite du gradient de disparité [6], un coût de pénalité (qui peut dépendre de la contrainte d'unicité [52] ou des disparités voisines [1, 92]), les disparités dans la zone d'agrégation [12, 36] ou un modèle probabiliste de données aberrantes [44].

3.4. Méthodes à passages multiples

Ils s'agit de méthodes locales à plusieurs passages. Une technique très populaire consiste à utiliser des *fenêtres adaptatives* [71, 72]. Pour Lotti et Giraudon [63], la taille des fenêtres est déterminée à partir d'une carte des contours : la fenêtre ne doit pas contenir de contour. Dans [35, chap. 2, et 3], une carte des disparités initiale est utilisée pour estimer l'ensemble des paramètres des transformations locales des fenêtres de corrélation. Une deuxième mise en correspondance est réalisée en prenant en compte ces transformations. Le principe des autres méthodes est de faire varier la taille et la forme des fenêtres de corrélation en fonction de la variation locale des niveaux de gris ainsi que de la disparité. La mise à jour de la fenêtre de corrélation peut être réalisée : en tenant compte d'une mesure d'incertitude [54, 71], en observant les disparités des pixels voisins [46], en testant seulement un certain nombre de fenêtres [34].

Dans le cadre des méthodes à base de *mesures pondérées*, celle de Zoghliami *et al.* [108] s'appuie sur le calcul de deux cartes des disparités : les scores de corrélation calculés lors d'une première étape de mise en correspondance sont utilisés pour attribuer à chaque pixel un poids qui sera utilisé dans une seconde étape pour calculer un score de corrélation pondéré. De cette manière, les pixels dont le score est mauvais dans la première étape ont peu d'influence dans la seconde.

Il existe aussi un nombre important de *méthodes hybrides*. Weng *et al.* [99] effectuent une estimation initiale des disparités. Sur chaque droite épipolaire, les variations de ces disparités sont étudiées : si le sens de variation change, cela implique une occultation. Kanade [106] utilise une méthode locale coopérative. Après la première étape, pour chaque étape suivante, la mesure de corrélation intègre un coût de pénalité qui dépend des disparités des pixels voisins. Dans [58], les composants de disparités, régions où la disparité est la même pour tous les pixels, sont extraits de la première mise en correspondance. La

deuxième étape prend en compte la forme de ces composants pour recalculer le score de corrélation. Eklund et Farag [32] appliquent un seuil sur les scores de corrélation obtenus lors de la première étape. C'est seulement pour les pixels qui se trouvent au dessus de ce seuil que l'appariement va être effectué une deuxième fois avec des formes variables de fenêtres.

3.5. Méthodes avec affinement

Un pixel peut être détecté comme occulté si son correspondant et lui-même ne respectent pas une contrainte (symétrie [2, 31], ordre [30, 106] ou unicité [58, 104]), s'ils obtiennent une mesure d'imprécision trop élevée [33] ou si leur score de corrélation est au dessus d'un seuil [88]. Un histogramme local des disparités pour un pixel donné peut être utilisé [31]. Si cet histogramme contient deux modes, c'est-à-dire deux pics, alors le pixel est considéré comme occulté.

4. Méthodes hybrides utilisant deux mesures de corrélation

Suite aux travaux que nous avons déjà réalisés [20, 21], nous proposons une nouvelle approche faisant intervenir deux mesures de corrélation avec des images en couleur : une mesure classique utilisée dans les zones sans occultation et une mesure robuste utilisée dans la zone des occultations. Nos algorithmes de mise en correspondance sont de quatre formes différentes, selon le moment où la détection de la zone des occultations est réalisée.

La taille des fenêtres de corrélation est $(2N_v + 1) \times (2N_h + 1)$ et le nombre de pixels dans cette fenêtre est $N_f = (2N_v + 1)(2N_h + 1)$. Les niveaux de gris des pixels des fenêtres de corrélation sont stockés dans les vecteurs $\mathbf{f}_l = (\dots I_l^{i+p, j+q} \dots)^T = (\dots f_l^k \dots)^T$ où f_l^k est l'élément d'indice k de \mathbf{f}_l , $p \in [-N_v; N_v]$, $q \in [-N_h; N_h]$ et $k \in [0; N_f - 1]$. Une mesure de corrélation est notée Mes . Nous notons d , d_{cla} et d_{rob} les fonctions de disparité obtenues respectivement avec les méthodes proposées, la mesure classique, notée Mes_{cla} , et la mesure robuste, notée Mes_{rob} . Nous utilisons « cla » (resp. « rob ») en indice lorsqu'il s'agit d'un calcul effectué avec la mesure classique (resp. robuste). Les vecteurs $\mathbf{c}_l^{i,j} = (x_l^{i,j} \ y_l^{i,j} \ z_l^{i,j})^T$, sont les couleurs des pixels de coordonnées $(i \ j)^T$ dans I_l . Les matrices \mathbf{F}_l de taille $N \times 3$, contiennent les composantes des couleurs des pixels dans les fenêtres de corrélation gauche et droite : $\mathbf{F}_l = (\dots \ \mathbf{c}_l^{i+p, j+q} \ \dots)^T$.

Tout d'abord nous présentons les zones d'occultations prises en compte dans cette approche, puis, nous détaillons les quatre types d'algorithmes proposés pour prendre en compte les occultations.

4.1. Zones des occultations considérées

Nous considérons ces zones (cf. figure 2) :

- *Les pixels occultés* – Ils ne possèdent pas de correspondant dans l'autre image :

$$O(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ est un pixel occulté} \\ 0 & \text{sinon.} \end{cases}$$

- *La zone des occultations* – Notée $ZO(I_g)$, elle contient tous les pixels occultés dans I_g :

$$ZO(I_g) = \{\mathbf{p}_g^{i,j} \mid O(\mathbf{p}_g^{i,j}) = 1\}.$$

- *Les pixels proches des pixels occultés* – Ce sont les pixels dans le voisinage (la fenêtre de corrélation) d'un pixel occulté. Cela correspond à la dilatation morphologique de la zone des occultations en utilisant comme élément structurant la fenêtre de corrélation :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } O(\mathbf{p}_g^{i,j}) = 0 \text{ et } V(\mathbf{p}_g^{i,j}) > 0 \\ 0 & \text{sinon,} \end{cases} \quad (6)$$

$$\text{et } V(\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in ZA(\mathbf{p}_g^{i,j})} O(\mathbf{p}_g^{i',j'}).$$

- *La zone d'influence des occultations* – Elle est notée $ZI(I_g)$ et contient tous les pixels proches des pixels occultés dans l'image I_g :

$$ZI(I_g) = \{\mathbf{p}_g^{i,j} \mid PO(\mathbf{p}_g^{i,j}) = 1\}.$$

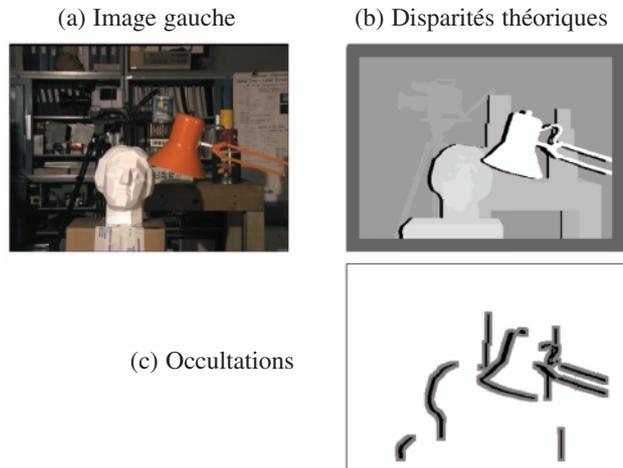


Figure 2. Zones des occultations considérées – Nous avons calculé la zone totale des occultations, ZT , en (c), grâce à la vérité terrain, fournie dans [80]. En (b), nous avons la carte des disparités (de référence) : chaque pixel représente l'amplitude de la disparité, c'est-à-dire, la distance entre la position du pixel de l'image gauche et celle de son correspondant dans l'image droite. Plus le pixel est clair et plus la disparité est importante. En (b) et (c), les pixels noirs représentent les pixels occultés. En (c), les pixels gris appartiennent à la zone d'influence des occultations.

- *La zone totale des occultations* – Elle est notée $ZT(I_g)$ et elle est l'union de la zone des occultations et de la zone d'influence des occultations dans l'image I_g :

$$ZT(I_g) = ZO(I_g) \cup ZI(I_g) \\ = \{\mathbf{p}_g^{i,j} \mid (O(\mathbf{p}_g^{i,j}) = 1) \text{ ou } (PO(\mathbf{p}_g^{i,j}) = 1)\}.$$

4.2. Méthodes utilisant les contours

Le principe de ces méthodes, cf. algorithme 4.1, est d'effectuer une détection de contours et de considérer les pixels détectés comme des pixels de la zone totale des occultations. Ensuite, nous utilisons la mesure robuste dans cette zone et la mesure classique dans le reste de l'image. Pour le prétraitement, c'est-à-dire l'étape 2a de l'algorithme 4.1, l'équation (6) devient :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ est un point contour} \\ 0 & \text{sinon.} \end{cases} \quad (7)$$

Nous n'avons considéré que des contours de type « marche ». Nous distinguons une méthode par seuillage de la norme du vecteur gradient calculé par les masques de Sobel, la méthode CONTOUR-SOB, et une méthode où les contours sont détectés par passage par zéro de la dérivée seconde en utilisant la méthode SDEF de Shen et Castan [83], la méthode CONTOUR-SDEF.

Algorithme 4.1. Méthode CONTOUR – Il s'agit d'une instance de l'algorithme 2.2 où la détection de la zone totale des occultations précède l'appariement, en utilisant les contours (prétraitement).

Pour chaque passage faire

- 2a Pour chaque $\mathbf{p}_g^{i,j}$ faire calculer $PO(\mathbf{p}_g^{i,j})$, cf. équation (7)
- 2b $S^{i,j} = \{\mathbf{p}_g^{i,j}\}$
- 2c Pour chaque $S^{i,j}$ faire
 - A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ Faire $Mes(\mathbf{F}_g, \mathbf{F}_d) =$

$$\begin{cases} Mes_{cla}(\mathbf{F}_g, \mathbf{F}_d) & \text{si } PO(\mathbf{p}_g^{i,j}) = 0 \\ Mes_{rob}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases}$$
 - F. Déterminer correspondant de $\mathbf{p}_g^{i,j}$ selon méthode WTA
- 2d Si passage 1 alors appliquer contrainte de symétrie, affiner au sous-pixel

Passage 0 – Appariement gauche -> droite

Passage 1 – Appariement droite -> gauche

4.3. Méthodes utilisant le principe des mesures de corrélation pondérée

Le principe, cf. algorithme 4.2, est de déterminer au moment du calcul du score de corrélation la mesure à utiliser. Ce choix s'ef-

Algorithme 4.2. Méthode CORRÉLATION – Cet algorithme est une instance de l'algorithme 2.2 où la détection de la zone totale des occultations est réalisée pendant la mise en correspondance.

Algorithme 4.3. Méthode POST-DÉTECTION – Cet algorithme est une instance de l'algorithme 2.2 où la détection de la zone totale des occultations est faite après la mise en correspondance.

Pour chaque passage faire

2b $S^{i,j} = \{p_g^{i,j}\}$

2c Pour chaque $S^{i,j}$ faire

A Pour $p_d^{i,v} \in Z_d(p_g^{i,j})$ faire

A.1 Évaluer $PO(p_g^{i,j})$ (méthode Kaneko ou Lan)

A.2 Calculer $Mes(\mathbf{F}_g, \mathbf{F}_d) =$

$\begin{cases} Mes_{cla}(\mathbf{F}_g, \mathbf{F}_d) & \text{si } PO(p_g^{i,j}) = 0 \\ Mes_{rob}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases}$

F. Déterminer correspondant de $p_g^{i,j}$ selon méthode WTA

2d Si passage 1 alors appliquer contrainte de symétrie, affiner au sous-pixel

Passage 0 – Appariement gauche -> droite
Passage 1 – Appariement droite -> gauche

Pour chaque passage faire

2a Si passage 2 ou passage 3 alors

Pour chaque $p_g^{i,j}$ faire calculer $PO(p_g^{i,j})$

2b $S^{i,j} = \{p_g^{i,j}\}$

Si passage 0 ou 1 alors prendre en compte tous les pixels sinon

Si $PO(p_g^{i,j}) = 1$ ou $O(p_g^{i,j}) = 1$ alors prendre en compte $p_g^{i,j}$

2c Pour chaque $S^{i,j}$ faire

A. Pour chaque $p_d^{i,v} \in Z_d(p_g^{i,j})$

Faire $Mes(\mathbf{F}_g, \mathbf{F}_d) =$

$\begin{cases} Mes_{cla}(\mathbf{F}_g, \mathbf{F}_d) & \text{si passage } < 2 \\ Mes_{rob}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases}$

F. Déterminer correspondant de $p_g^{i,j}$ selon méthode WTA

2d Si passage 1 alors $O(p_g^{i,j}) =$

$\begin{cases} 1 & \text{si } p_g^{i,j} + d_{cla}(p_g^{i,j}) = p_d^{i,v} \text{ et } p_d^{i,v} + d_{cla}(p_d^{i,v}) \neq p_g^{i,j} \\ 0 & \text{sinon} \end{cases}$

Si passage 3 alors appliquer contrainte de symétrie, affiner au sous-pixel

Passage 0 – gauche -> droite et Mes_{cla}
Passage 1 – droite -> gauche et Mes_{cla}
Passage 2 – gauche -> droite et Mes_{rob}
Passage 3 – droite -> gauche et Mes_{rob}



fectue en étudiant l'ensemble des pixels dans les fenêtres de corrélation : à chaque calcul du score de corrélation, un poids est attribué à chaque pixel du voisinage. Suivant les valeurs de ces poids, nous utilisons soit la mesure classique, soit la mesure robuste. Nous proposons d'utiliser la méthode ordinale de Kaneko *et al.* [55], la méthode CORRÉLATION-KAN, ou les moindres carrés médians, comme dans [59], la méthode CORRÉLATION-LAN.

4.4. Méthodes avec post-détection

Il s'agit d'effectuer une mise en correspondance avec une mesure classique, puis de déterminer la zone totale des occultations et enfin d'effectuer une mise en correspondance partielle (sur la zone totale des occultations) avec une mesure robuste. Pour le post-traitement, étape 2a de l'algorithme 4.3, les solutions envisagées pour détecter les pixels dans la zone totale des occultations sont :

- Méthode POST-SEU – Un seuil T_S (choisi de manière empirique et qui dépend de la mesure utilisée) est appliqué au score :

$$PO(p_g^{i,j}) = \begin{cases} 1 & \text{si } Mes_{cla}(\mathbf{F}_g, \mathbf{F}_d) > T_S \\ 0 & \text{sinon.} \end{cases}$$

- Méthode POST-AMB – Nous notons $A_{cla}(p_g^{i,j})$ l'ambiguïté obtenue pour le pixel $p_g^{i,j}$ [27] :

$$PO(p_g^{i,j}) = \begin{cases} 1 & \text{si } A_{cla}(p_g^{i,j}) > T_A \\ 0 & \text{sinon.} \end{cases}$$

Le seuil T_A a été déterminé empiriquement et correspond à 60 % de la valeur maximale de l'ambiguïté sur toute l'image.

- Méthodes POST-SYM, POST-UNI, POST-ORD – Nous utilisons respectivement la contrainte de symétrie, la contrainte d'unicité, la contrainte d'ordre, cf. § 2.4.

La méthode POST-SYM semble la plus prometteuse car il n'y a pas de seuil à fixer et les cartes des disparités que nous avons obtenues dans [19] montrent que la contrainte de symétrie détecte de nombreux pixels comme occultés parmi les pixels proches des pixels occultés. Toutefois, le but de la méthode proposée est d'améliorer les résultats dans la zone d'influence des occultations, or la contrainte de symétrie détecte des pixels dans la zone d'influence des occultations mais aussi les pixels véritablement occultés. Ainsi cette méthode peut détériorer les résultats obtenus par la mesure classique. C'est pourquoi, nous proposons d'autres variantes utilisant aussi la contrainte de symétrie :

- Méthode POST-MÉD – Nous effectuons un filtrage médian sur la carte des disparités en suivant cette règle : un pixel occulté (resp. non occulté) dont le nombre de voisins occultés est supérieur à un seuil T_{F_1} (resp. T_{F_2}) est considéré comme appartenant à la zone d'influence des occultations. Nous avons déterminé

empiriquement $T_{F_1} = \frac{N_f}{10}$ et $T_{F_2} = \frac{N_f}{7}$, avec N_f le nombre de pixels de la fenêtre de corrélation, et $PO(\mathbf{p}_g^{i,j})$ est évalué par :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } (O(\mathbf{p}_g^{i,j}) = 1 \text{ et } V(\mathbf{p}_g^{i,j}) > T_{F_1}) \\ & \text{ou} \\ & (O(\mathbf{p}_g^{i,j}) = 0 \text{ et } V(\mathbf{p}_g^{i,j}) > T_{F_2}) \\ 0 & \text{sinon.} \end{cases}$$

- Méthode POST-DIL – La contrainte de symétrie est appliquée et nous calculons le dilaté, obtenu par dilatation conditionnelle au sens de la morphologie mathématique, par la fenêtre de corrélation, des zones des occultations de la carte initiale des disparités. La dilatation est conditionnée par le nombre de pixels occultés dans la fenêtre de corrélation :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } O(\mathbf{p}_g^{i,j}) = 0 \text{ et } V(\mathbf{p}_g^{i,j}) > T_D = \frac{3N_f}{10} \\ 0 & \text{sinon.} \end{cases}$$

Le seuil T_D a été choisi en étudiant les différentes possibilités sur le nombre et la disposition des pixels occultés dans la fenêtre de corrélation. Nous supposons que si plus de 30 % de la fenêtre est occultée alors, le pixel considéré possède une forte probabilité d'être dans la zone d'influence des occultations. Si la carte des disparités initiale possède trop de faux négatifs (ce sont les pixels détectés comme occultés alors qu'ils ne le sont pas), cet algorithme risque d'augmenter leur nombre. C'est pourquoi nous proposons la variante suivante.

- Méthode POST-DIL-CONT – Cette fois, la dilatation est conditionnée par le nombre de pixels occultés et le nombre de points contour dans la fenêtre de corrélation. Nous limitons la dilatation des pixels occultés et en supposant que nous pouvons faire confiance aux pixels détectés comme proches de pixels occultés lorsque ceux-ci sont proches d'un contour :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } V(\mathbf{p}_g^{i,j}) > T_D \text{ et } PF(\mathbf{p}_g^{i,j}) > T_{C_2} \\ 0 & \text{sinon,} \end{cases}$$

$$PF(\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in ZA(\mathbf{p}_g^{i,j})} F(\mathbf{p}_g^{i',j'}), \quad (8)$$

$$F(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ pixel contour} \\ 0 & \text{sinon.} \end{cases}$$

Nous avons déterminé empiriquement le seuil $T_{C_2} = \frac{N_f}{20}$.

- Méthode POST-ORD-SYM – Nous combinons les contraintes d'ordre et de symétrie, c'est-à-dire que les pixels impliqués dans une correspondance qui ne respecte pas la contrainte d'ordre ou la contrainte de symétrie sont considérés comme occultés.

- Méthode POST-AMB-SYM – Nous combinons la contrainte d'ambiguïté et de symétrie :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } (A_{\text{cla}}(\mathbf{p}_g^{i,j}) > T_A) \\ & \text{ou } (\mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \\ & \text{et } \mathbf{p}_d^{i,v} + d_{\text{cla}}(\mathbf{p}_d^{i,v}) \neq \mathbf{p}_g^{i,j}) \\ 0 & \text{sinon.} \end{cases}$$

4.5. Méthode de fusion de résultats

Une dernière technique envisagée, la méthode FUSION-DISP, consiste à calculer les cartes des disparités avec une mesure classique et une mesure robuste et à fusionner les résultats obtenus. Quand un pixel est occulté, nous notons sa disparité symbolique **occ** et les règles de fusion utilisées sont :

- Si les deux mesures de corrélation donnent le même résultat (pixel occulté ou un même correspondant), alors ce résultat est conservé.

- Si une seule des deux mesures détecte le pixel comme occulté, alors les résultats obtenus avec cette mesure dans le voisinage de ce pixel sont examinés. Nous supposons que s'il y a de nombreux pixels détectés comme occultés dans son voisinage alors il y a une forte probabilité pour que ce pixel soit occulté, sinon il s'agit d'un faux négatif. Plus précisément, si plus de la moitié des voisins sont occultés alors le pixel est occulté, sinon la disparité obtenue est conservée :

$$d(\mathbf{p}_g^{i,j}) = \begin{cases} \text{occ} & \text{si } (O_{\text{cla}}(\mathbf{p}_g^{i,j}) = 1 \text{ et} \\ & V_{\text{cla}}(\mathbf{p}_g^{i,j}) > \frac{N_f}{2} \text{ ou} \\ & (O_{\text{rob}}(\mathbf{p}_g^{i,j}) = 1 \text{ et} \\ & V_{\text{rob}}(\mathbf{p}_g^{i,j}) > \frac{N_f}{2}) \\ d_{\text{cla}}(\mathbf{p}_g^{i,j}) & \text{si } O_{\text{rob}}(\mathbf{p}_g^{i,j}) = 1 \\ d_{\text{rob}}(\mathbf{p}_g^{i,j}) & \text{sinon.} \end{cases}$$

- Si les deux mesures donnent des disparités différentes, nous supposons qu'il y a une forte probabilité pour que ce pixel possède un correspondant mais, en revanche, que ce pixel se situe dans une zone difficile à apparier. Nous pensons que plus il y a de pixels détectés comme occultés dans son voisinage, qui sont probablement des faux négatifs, plus la mesure utilisée est ambiguë ; c'est pourquoi nous privilégions la mesure qui a donné le moins de pixels occultés dans le voisinage :

$$d(\mathbf{p}_g^{i,j}) = \begin{cases} d_{\text{cla}}(\mathbf{p}_g^{i,j}) & \text{si } V_{\text{rob}}(\mathbf{p}_g^{i,j}) > V_{\text{cla}}(\mathbf{p}_g^{i,j}) \\ d_{\text{rob}}(\mathbf{p}_g^{i,j}) & \text{sinon.} \end{cases}$$

5. Protocole d'évaluation

Bien que le protocole de Scharstein et Szeliski² soit le plus pertinent à notre connaissance, nous pensons qu'il n'est pas adapté aux études comparatives que nous souhaitons réaliser :

- Des méthodes globales sont évaluées alors que nous travaillons avec des méthodes locales.

- Il n'y a que trois critères d'évaluation et nous souhaitons réaliser une évaluation plus fine. En effet, nous proposons une évaluation sur dix critères et nous mettons particulièrement en évi-

² <http://cat.middlebury.edu/stereo/>

dence les résultats obtenus dans la zone des discontinuités et dans la zone d'influence des occultations (le protocole de Scharstein et Szeliski ne fait pas cette distinction).

Nous présentons notre protocole en décrivant les images utilisées, les zones d'évaluation considérées, les critères utilisés et la synthèse des critères permettant d'obtenir le classement des méthodes.

5.1. Images testées

Dans le but d'utiliser de nouvelles images de référence, nous avons proposé une méthode d'obtention de la vérité terrain [13]. Nous considérons des scènes constituées d'objets dont la surface est un ensemble de facettes planes dans le but d'utiliser des contraintes géométriques sur la scène. L'approche se déroule en cinq étapes : segmentation des images, mise en correspondance de points d'intérêt par corrélation avec application de la contrainte de symétrie et un affinement au sous-pixel par l'interpolation des niveaux de gris, estimation des homographies et calcul des disparités avec localisation des occultations. Benoît Bocquillon met à la disposition de la communauté les couples d'images proposés ainsi que l'outil d'obtention de la vérité terrain : <http://www.irit.fr/~Alain.Crouzil/images.html>. Les quatorze couples d'images testés dans notre protocole sont donc (cf. exemples dans la figure 3) : un stéréogramme aléatoire utilisé dans [19], les images de synthèse proposées dans [13], les images réelles proposées dans [80], le couple d'images réelles proposé dans [13].

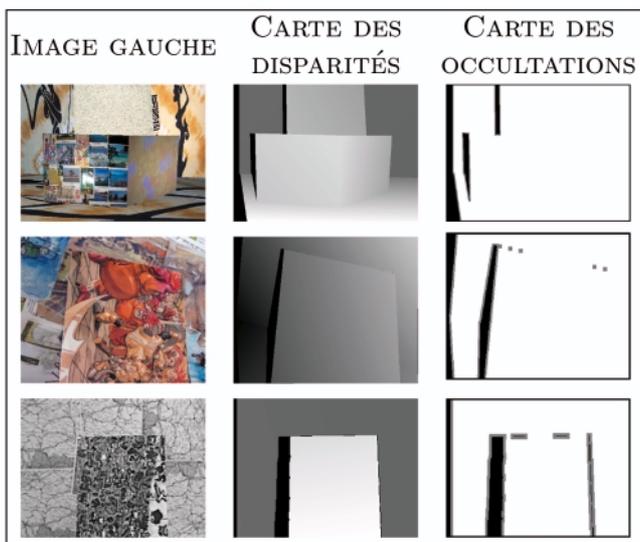


Figure 3. Couples d'images utilisées par le protocole. Sur la première ligne, il s'agit d'images de synthèse [13] alors que les autres images sont des images réelles [13, 80].

5.2. Zones d'évaluation

Nous utilisons les zones présentées au § 4.1 et la zone des discontinuités (pixels en gris sur la carte des discontinuités dans la figure 4), en distinguant :

- *Les pixels proches d'une discontinuité* – Ce sont les pixels dans le voisinage, défini par rapport à la zone d'agrégation, d'un pixel qui possède une disparité suffisamment différente :

$$PD(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \exists \mathbf{p}_g^{i,v} \in ZA(\mathbf{p}_g^{i,j}) \mid \\ & \|d(\mathbf{p}_g^{i,v}) - d(\mathbf{p}_g^{i,j})\| > T_{Dis} \\ 0 & \text{sinon.} \end{cases}$$

Le seuil T_{Dis} a été déterminé de manière empirique pour chaque image.

- *La zone des discontinuités* – Elle est notée $ZD(I_g)$ et contient tous les pixels proches d'une discontinuité dans l'image I_g : $ZD(I_g) = \{\mathbf{p}_g^{i,j} \mid PD(\mathbf{p}_g^{i,j}) = 1\}$.

5.3. Critères d'évaluation

Nous notons d_{ref} la fonction de disparité théorique et $Err_g^{i,j} = \|d(\mathbf{p}_g^{i,j}) - d_{ref}(\mathbf{p}_g^{i,j})\|$. Le calcul de $Err_g^{i,j}$ n'a de signification que si $d(\mathbf{p}_g^{i,j}) \neq \mathbf{occ}$ et $d_{ref}(\mathbf{p}_g^{i,j}) \neq \mathbf{occ}$ et les pourcentages évalués sont :

1. *Appariement corrects*, COR – $Err_g^{i,j} < 1$.
2. *Appariements acceptés*, ACC – $1 \leq Err_g^{i,j} < 2$.
3. *Appariements mauvais*, MAU – $2 \leq Err_g^{i,j} < 3$.
4. *Appariements erronés*, ERR – $Err_g^{i,j} \geq 3$.
5. *Faux positifs*, FPO – Un appariement est un faux positif lorsqu'une correspondance a été établie pour un pixel alors qu'il est occulté, c'est-à-dire, si $d(\mathbf{p}_g^{i,j}) \neq d_{ref}(\mathbf{p}_g^{i,j})$ et $d_{ref}(\mathbf{p}_g^{i,j}) = \mathbf{occ}$.
6. *Faux négatifs*, FNE – Un appariement est un faux négatif lorsqu'un pixel est considéré comme occulté alors qu'il ne l'est pas, c'est-à-dire, si $d(\mathbf{p}_g^{i,j}) \neq d_{ref}(\mathbf{p}_g^{i,j})$ et $d(\mathbf{p}_g^{i,j}) = \mathbf{occ}$.

Le critère 1 est évalué sur chaque zone. Afin d'obtenir un résultat plus visuel, nous fournissons aussi des cartes des appariements corrects : si le pixel est blanc alors l'appariement est incorrect, s'il est noir alors c'est un vrai négatif et s'il est gris alors l'appariement est correct. Nous avons ajouté deux cartes pour analyser les résultats : la carte des occultations qui représente l'ensemble des pixels (pixels noirs) détectés, par la méthode évaluée, comme appartenant à la zone totale des occultations et la carte des occultations correctes qui représente la qualité d'une carte des occultations. Elle met en évidence les vrais négatifs (pixels noirs) et les pixels correctement détectés proche d'une occultation (pixels gris).

5.4. Mesures de corrélation

Les premiers résultats obtenus dans [21] nous ont permis de déterminer la meilleure combinaison de mesure : ZNCC et

Tableau 3. Méthodes proposées pour prendre en compte le problème des occultations – Il y a quatre types de méthodes et quinze variantes différentes.

TYPE	VARIANTE	CARACTÉRISTIQUES
CONTOUR	CONTOUR-SOB	Utilisation du filtre de Sobel
	CONTOUR-SDEF	Utilisation de la méthode SDEF
CORRÉLATION	CORRÉLATION-KAN	Utilisation de la méthode de Kaneko
	CORRÉLATION-LAN	Utilisation de la méthode de Lan
POST-DÉTECTION	POST-SEU	Seuil sur le score de corrélation
	POST-AMB	Seuil sur l'ambiguïté
	POST-SYM	Contrainte de symétrie
	POST-UNI	Contrainte d'unicité
	POST-ORD	Contrainte d'ordre
	POST-MÉD	Contrainte de symétrie suivie d'un filtrage médian
	POST-DIL	Contrainte de symétrie suivie d'une dilatation des zones des occultations
	POST-DIL-CONT	Contrainte de symétrie suivie d'une dilatation des zones des occultations prenant en compte les contours
	POST-ORD-SYM	Contrainte d'ordre et contrainte de symétrie
	POST-AMB-SYM	Seuil sur l'ambiguïté et contrainte de symétrie
FUSION	FUSION-DISP	Fusion des deux cartes des disparités

SMPD₂. La mesure de corrélation croisée centrée normalisée, notée ZNCC, (*Zero mean Normalized Cross-Correlation*), est le coefficient de corrélation linéaire classique en statistiques :

$$\text{ZNCC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{(\mathbf{f}_g - \overline{\mathbf{f}_g}) \cdot (\mathbf{f}_d - \overline{\mathbf{f}_d})}{\|\mathbf{f}_g - \overline{\mathbf{f}_g}\| \|\mathbf{f}_d - \overline{\mathbf{f}_d}\|}.$$

La mesure robuste s'appuie sur l'estimateur SMAD (*Smooth Median Absolute Deviation*), qui est une estimation robuste de la variance [77]. Les valeurs ordonnées du vecteur \mathbf{f}_l sont notées : $(f_l)_{0:N_f-1} \leq \dots \leq (f_l)_{k:N_f-1} \leq \dots \leq (f_l)_{N_f-1:N_f-1}$. Dans [19], nous définissons la somme des h premières³ puissances des écarts à la médiane, notée SMPD_P (*Smooth Median Powered Deviation*) :

$$\text{SMPD}_P(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{h-1} (|\mathbf{f}_g - \mathbf{f}_d - \text{med}(\mathbf{f}_g - \mathbf{f}_d)|^P)_{k:N_f-1}.$$

La méthode élémentaire est la méthode de mise en correspondance utilisant uniquement une de ces mesures. Par ailleurs, nous utilisons des images couleur. Les résultats obtenus dans [20], nous permettent de choisir cette stratégie : calcul des

scores de corrélation sur chaque composante du système XYZ, puis fusion par minimum.

5.5. Méthodes évaluées

Nous avons évalué tous les algorithmes proposés, cf. § 4, ainsi que les différentes variantes, cf. tableau 3. De plus, nous avons comparé les résultats obtenus avec les méthodes hybrides avec ceux obtenus avec la méthode de mise en correspondance élémentaire avec ZNCC ou SMPD₂.

5.6. Synthèse des critères et classement des méthodes

Pour effectuer le classement des méthodes, nous attribuons un rang pour chaque critère, puis nous effectuons le classement des méthodes en fonction des moyennes de ces rangs. Dans les tableaux de résultats, nous utilisons la notation RT (rang total) pour indiquer le rang de chaque méthode parmi toutes les méthodes évaluées. Afin d'obtenir une présentation la plus lisible possible dans les tableaux, nous ne présentons des résultats que pour les six meilleures méthodes : les deux méthodes élémentaires utilisant respectivement ZNCC et SMPD₂, les méthodes POST-SYM, POST-DIL, POST-DIL-CONT et FUSION-DISP.

³ Pour obtenir une mesure très tolérante aux données aberrantes, nous avons pris $h = \frac{N_f}{2}$.

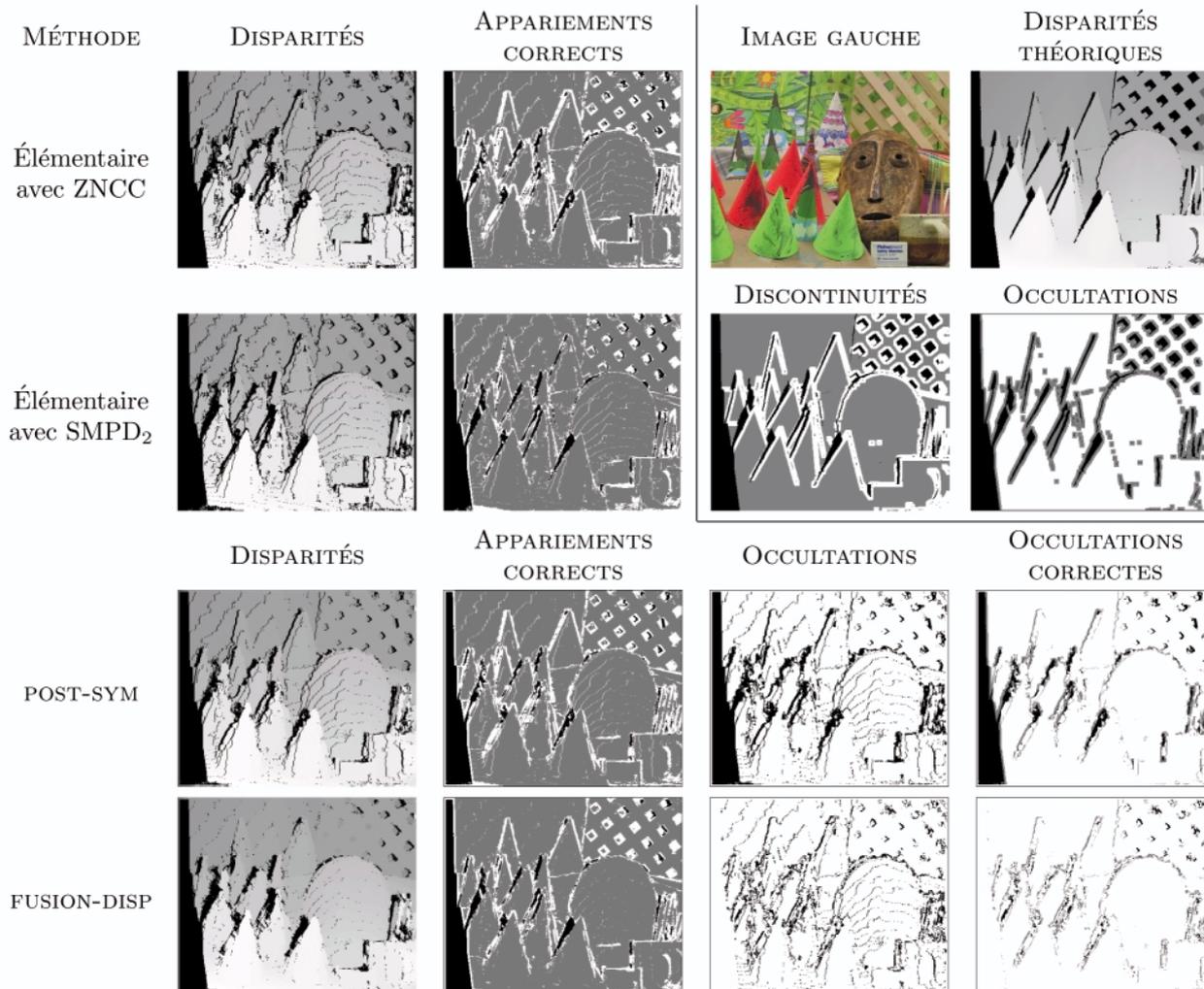


Figure 4. Cartes de résultats du couple cones – Ces images sont celles qui contiennent la plus grande zone totale des occultations (33 %). La méthode FUSION-DISP obtient la carte des disparités la plus nette et la carte des appariements corrects la plus dense.

Les résultats en gras correspondent aux meilleurs résultats par critère. Dans les cartes de résultats fournis dans les figures, nous ne donnons les résultats que pour les deux méthodes élémentaires et les méthodes POST-SYM et FUSION-DISP, ces deux dernières étant les plus performantes.

6. Résultats

En termes de temps de calcul, la méthode CONTOUR-SDEF est la plus rapide, après l’algorithme élémentaire utilisant une mesure classique. La méthode FUSION-DISP est la plus coûteuse.

Dans [21], nous avons mis en évidence que, pour la méthode CONTOUR, quel que soit le masque utilisé, les pourcentages de pixels corrects près des zones des occultations ne sont pas toujours meilleurs que ceux obtenus avec la méthode élémentaire. Cette méthode pose un problème délicat : le choix des seuils. Nous avons déterminé, de manière empirique, les seuils opti-

maux pour chaque image et il est clairement apparu qu’il était très difficile de les déterminer automatiquement. Pour la méthode CORRÉLATION, les résultats ne sont pas meilleurs. Le même problème se pose pour le choix des seuils. De plus, la méthode de Lan est beaucoup plus coûteuse en temps de calcul que toutes les autres méthodes. Enfin, les deux dernières catégories d’algorithmes s’avèrent les plus efficaces. Les résultats sont globalement améliorés par rapport à ceux obtenus avec l’algorithme élémentaire, en particulier en utilisant les techniques fondées sur la contrainte de symétrie : méthodes POST-SYM, POST-DIL et POST-DIL-CONT. La méthode FUSION-DISP donne aussi de bons pourcentages de pixels corrects et améliore les résultats obtenus avec la méthode élémentaire utilisant une mesure classique, pour tous les critères, excepté le pourcentage de pixels erronés.

Le tableau 4 permet de mettre en évidence que les meilleurs résultats sont toujours obtenus avec la méthode FUSION-DISP qui fournit toujours les meilleurs résultats dans la zone d’influence des occultations et dans la zone des discontinuités (cette

Tableau 4. Classement des méthodes hybrides de mise en correspondance d'images couleur – Nous avons synthétisé les résultats obtenus sur les quatorze couples de tests avec le protocole que nous avons décrit dans le paragraphe 5.

Chaque case contient le rang moyen obtenu sur les quatorze couples, pour chaque critère, pour chaque méthode. La dernière colonne permet d'indiquer le classement global (qui est la moyenne des rangs obtenus pour chaque critère) de toutes les méthodes.

Nous pouvons remarquer que seule la méthode POST-DIL est moins bien classée que la méthode élémentaire utilisant ZNCC.

La seule méthode qui donne de meilleurs résultats que la méthode élémentaire utilisant SMPD₂ est la méthode FUSION-DISP.

Celle-ci est classée première sur six critères différents.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
Élémentaire avec ZNCC	4	6	6	3	2	5	3	2	6	5	5
Élémentaire avec SMPD ₂	5	4	3	1	1	6	1	1	3	6	2
POST-SYM	2	2	1	6	6	2	4	6	2	2	3
POST-DIL	6	5	4	4	5	4	6	5	5	3	6
POST-DIL-CONT	3	3	5	2	4	3	5	4	4	4	4
FUSION-DISP	1	1	1	5	3	1	2	3	1	1	1

méthode obtient la première place pour six critères et elle occupe donc la première place par rapport au rang total). La méthode FUSION-DISP est aussi classée en première position pour le pourcentage d'appariements corrects. Cette méthode allie les avantages d'une mesure classique, c'est-à-dire l'obtention du meilleur pourcentage d'appariements corrects, aux avantages d'une mesure robuste, c'est-à-dire l'obtention des meilleurs pourcentages d'appariements corrects dans la zone d'influence des occultations et dans la zone des discontinuités. Nous donnons le détail des résultats obtenus sur un couple d'images : *cones*.

Ces images possèdent la plus grande zone totale des occultations parmi toutes les images testées. Bien que la méthode élémentaire utilisant SMPD₂ soit meilleure que celle utilisant ZNCC, la méthode FUSION-DISP est toujours la mieux classée, cf. tableau 5. Les améliorations sont nettement visibles si on observe les cartes de disparités, cf. figure 4. Notons aussi pour

ces images que la méthode POST-DIL est classée deuxième. Cela vient du fait qu'elle se comporte bien dans la zone d'influence des occultations et dans la zone des discontinuités. De même, elle fournit moins de faux négatifs qu'avec les autres images.

7. Conclusion

Les contributions de cet article portent sur trois aspects. D'une part, nous proposons un état de l'art des méthodes de mise en correspondance de pixels, notamment en proposant un algorithme générique complet dont les instances permettent de décrire les méthodes de mise en correspondance par l'intermédiaire d'éléments constitutifs. Les lecteurs intéressés peuvent trouver davantage de références, sur les méthodes de mise en correspondance sur la page suivante :

S

Tableau 5. Résultats du couple cones – Nous indiquons entre parenthèses le rang de la méthode pour chaque critère et pour ces images. La méthode FUSION-DISP est globalement la meilleure mais elle n'obtient pas un pourcentage d'appariements corrects aussi élevé que celui obtenu par la méthode élémentaire utilisant SMPD₂. Cependant, contrairement à la majorité des images, pour ce couple, la méthode élémentaire utilisant SMPD₂ obtient un pourcentage d'appariements corrects nettement meilleur que celui de la méthode élémentaire utilisant ZNCC.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
Élémentaire avec ZNCC	81.22 (5)	1.32 (5)	0.62 (4)	2.7 (3)	4.06 (2)	10.08 (6)	70.09 (5)	70.97 (2)	69.47 (6)	78.11 (6)	5
Élémentaire avec SMPD ₂	85.86 (1)	0.46 (6)	0.2 (6)	1.22 (1)	2.91 (1)	9.07 (4)	77.4 (1)	79.2 (1)	76.14 (3)	78.87 (5)	2
POST-SYM	82.58 (4)	2.01 (1)	0.67 (1)	3.32 (6)	5.51 (6)	5.92 (2)	71.29 (3)	60.68 (6)	78.71 (2)	89.12 (2)	3
POST-DIL	80.02 (6)	1.97 (2)	0.66 (2)	3.09 (4)	5.18 (5)	9.08 (5)	67.35 (6)	63 (5)	70.4 (5)	81.49 (4)	5
POST-DIL-CONT	83.36 (3)	1.64 (4)	0.49 (5)	2.37 (2)	4.65 (3)	7.49 (3)	70.82 (4)	66.77 (3)	73.65 (4)	82.68 (3)	4
FUSION-DISP	85.23 (2)	1.96 (3)	0.65 (3)	3.2 (5)	4.95 (4)	4.02 (1)	75.79 (2)	64.66 (4)	83.57 (1)	90.48 (1)	1

<http://www.irit.fr/~Alain.Crouzil/appariement.html>. D'autre part, nous présentons un état de l'art sur la façon de prendre en compte les occultations. Une liste de références est également disponible sur la page citée ci-dessus. Enfin, nous proposons des méthodes hybrides, de quatre types différents, qui combinent les avantages d'une mesure classique et d'une mesure robuste.

Les résultats obtenus montrent que la meilleure solution consiste à fusionner deux cartes de disparités obtenues avec deux mesures différentes. Ces résultats, très encourageants, nous permettent d'envisager trois types de perspectives. Dans un premier temps, nous souhaitons utiliser plus de deux cartes de disparités lors de la fusion, notamment, en intégrant une carte de disparités obtenue avec une mesure qui obtient des résultats plus performants que les autres mesures dans les zones peu texturées. Le problème à résoudre est la façon dont cette fusion va être réalisée. Ensuite, nous voulons intégrer la méthode de fusion dans un algorithme de mise en correspondance globale. Il faut alors décider quelle méthode d'optimisation est la plus adaptée et comment réaliser l'intégration de la méthode décrite dans cet article. Enfin, l'utilisation de plusieurs mesures de corrélation dans un processus de mise en correspondance par croissance de germes est aussi une piste intéressante que nous avons commencé à suivre et qui semble prometteuse. Un point crucial de cette approche réside dans la stratégie d'intégration des différentes mesures dans un algorithme itératif de croissance de germes.



Références

- [1] M. AGRAWAL et L. DAVIS, « Windowbased, discontinuity preserving stereo ». Dans *IEEE CVPR*, volume 1, pages 66-73, 2004.
- [2] L. ALVAREZ, R. DERICHE, T. PAPADOPOULOS et J. SÁNCHEZ, « Symmetrical Dense Optical Flow Estimation with Occlusions Detection ». Dans *ECCV*, volume 1, pages 721-735, 2002.
- [3] P. ANANDAN, « A Computational Framework and an Algorithm for the Measurement of Visual Motion ». *IJCV*, 2(3):283-310, 1989.
- [4] X. ARMANGUÉ et J. SALVI, « Overall view regarding fundamental matrix estimation ». *Image and Vision Computing*, 21(2):205-220, 2003.
- [5] P. ASCHWANDEN et W. GUGGENBÜL, Experimental results from a comparative study on correlation type registration algorithms. Dans W. Förstner et S. Rüwiedel, éditeurs, *Robust computer vision: Quality of Vision Algorithms*, pages 268-282. Wichmann, 1992.
- [6] S. T. BARNARD, « Stochastic Stereo Matching over Scale ». *IJCV*, 3(1):17-32, 1989.
- [7] P. N. BELHUMEUR et D. MUMFORD, « A bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions ». Dans *IEEE CVPR*, pages 506-512, 1992.
- [8] R. E. BELLMAN et S. E. DREYFUS, *La programmation dynamique et ses applications*. Dunod, 1962.
- [9] D. N. BHAT et S. K. NAYAR, « Ordinal Measures for Image Correspondence ». *IEEE TPAMI*, 20(4):415-423, 1998.
- [13] B. BOCQUILLON, S. CHAMBON et A. CROUZIL, « Segmentation semi-automatique en plans pour la génération de cartes denses de disparités ». Dans *ORASIS*, 2005. CDROM.
- [14] A. BOVYRIN, V. ERUHIMOV, S. MOLINOV, V. MOSYAGIN et V. PISAREVSKY, « Fast and robust dense stereo correspondence by column segmentation ». Dans *IEEE ICIP*, volume 2, pages 1033-1036, 2003.
- [15] Y. BOYKOV, O. VEKSLER et R. ZABIH, « Fast approximate energy minimization via graph cuts ». *IEEE T-PAMI*, 23(11):1222-1239, 2001.
- [16] M. Z. BROWN, D. BURSCHKA et G. D. HAGER, « Advances in Computational Stereo ». *IEEE T-PAMI*, 25(8):993-1008, 2003.
- [17] P. J. BURT et E. H. ADELSON, « The Laplacian Pyramid as a Compact Image Code ». *IEEE T-Communications*, 31(4):532-540, 1983.
- [18] J. F. CANNY, « A computational approach to edge detection ». *IEEE T-PAMI*, 8(6):679-698, 1986.
- [19] S. CHAMBON et A. CROUZIL, « Dense matching using correlation: new measures that are robust near occlusions ». Dans *BMVC*, volume 1, pages 143-152, 2003.
- [20] S. CHAMBON et A. CROUZIL, « Mesures de corrélation pour des images couleur ». *Trait. Signal*, 21(6):635-659, 2004.
- [21] S. CHAMBON et A. CROUZIL, « Towards correlation-based matching algorithms that are robust near occlusions ». Dans *IEEE ICPR*, volume 3, pages 20-23, 2004.
- [22] H. CHEN et P. MEER, « Robust Regression with Projection Based M-estimators ». Dans *IEEE ICCV*, volume 2, pages 878-885, 2003.
- [23] A. CROUZIL, « Perception du relief et du mouvement par analyse d'une séquence stéréoscopique d'images ». Thèse de doctorat, université Paul Sabatier, Toulouse III, 1997.
- [24] A. CROUZIL, X. DESCOMBES et J.-D. DUROU, « A Multiresolution Approach for Shape from Shading Coupling Deterministic and Stochastic Optimization ». *IEEE TPAMI*, 25(11):1416-1421, 2003.
- [25] A. CROUZIL, L. MASSIP-PAILHES et S. CASTAN, « A New Correlation Criterion Based on Gradient Fields Similarity ». Dans *IEEE ICPR*, volume 1, pages 632-636, 1996.
- [26] L. S. DAVIS et A. ROSENFELD, « Cooperating processes for low-level vision: A survey ». *Artificial Intelligence*, 17(1-3):245-263, 1981.
- [27] O. De JOINVILLE, G. MAILLET, H. MAÎTRE et M. ROUX, « Évaluation a priori de la qualité d'un MNS ». Dans *ORASIS*, pages 67-76, 2001.
- [28] Y. DENG, Q. YANG, X. LIN et X. TANG, « A Symmetric Patch-Based Correspondence Model for Occlusion Handling ». Dans *IEEE ICCV*, volume 2, pages 1316-1322, 2005.
- [29] X. DESCOMBES, « Méthodes stochastiques en analyse d'image : des champs de Markov aux processus ponctuels marqués ». Habilitation à diriger des recherches, université de Nice-Sophia Antipolis, 2004.
- [30] U. R. DHOND et J. K. AGGARWAL, « Stereo Matching in the Presence of Narrow Occluding Objects Using Dynamic Disparity Search ». *IEEE T-PAMI*, 17(7):719-724, 1995.
- [31] G. EGNAL et R. P. WILDES, « Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches ». *IEEE T-PAMI*, 24(8):1127-1133, 2002.
- [32] M. P. EKLUND et A. A. FARAG, « Robust correspondence methods for stereo vision ». *Pattern Recognition and Artificial Intelligence*, 17(7):1059-1079, 2003.
- [33] P. FUA, « A parallel stereo algorithm that produces dense depth maps and preserves image features ». *Machine Vision and Applications*, 6(1):35-49, 1993.
- [34] A. FUSIELLO, V. ROBERTO et E. TRUCCO, « Ecient Stereo with Multiple Windowing ». Dans *IEEE CVPR*, pages 858-863, 1997.
- [35] D. GARCIA, « Mesures de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images ». Thèse de doctorat, école des mines d'Albi, 2001.
- [36] D. GEIGER, B. LADENDORF et A. YUILLE, « Occlusions and Binocular Stereo ». Dans *IEEE ICIP*, volume 14, pages 211-226, 1995.

- [37] A. GHAFOR, R. NAVEED IQBAL et S. SHOAD KHAN, « Image Matching Using Distance Transform ». Dans *SCIA*, pages 654-660, 2003.
- [38] M. GONG, « Motion Estimation using Dynamic Programming with Selective Path Search ». Dans *IEEE ICPR*, volume 4, pages 203-206, 2004.
- [39] L. GOTTSFELD BROWN, « A Survey of Image Registration Techniques ». *ACM Computing Surveys*, 24(4):325-376, 1992.
- [40] J. Y. GOULERMAS et P. LIATSIS, « A Collective-Based Adaptive Symbiotic Model for Surface Reconstruction in Area-Based Stereo ». *IEEE T-EC*, 7(5):482-502, 2003.
- [41] K.-P. HAN, K.-W. SONG, E.-Y. CHUNG, S.-J. CHO et Y.-H. HA, « Stereo matching using genetic algorithm with adaptive chromosomes ». *Pattern Recognition*, 34(9):1729-1740, 2001.
- [42] C. HARRIS et M. STEPHENS, « A Combined Corner and Edge Detector ». Dans *Alvey Vision Conference*, pages 147-151, 1988.
- [43] R. I. HARTLEY, « Theory and Practice of Projective Rectification ». *IJCV*, 35(2):115-127, 1999.
- [44] D. HASLER, L. SBAIZ, S. SÜSTRUNK et M. VETTERLI, « Outlier Modelling in Image Matching ». *IEEE T-PAMI*, 25(3):301-315, 2003.
- [45] H. HIRSCHMÜLLER, « Accurate and efficient stereo processing by semi-global matching and mutual information ». Dans *IEEE CVPR*, volume 2, pages 807-814, 2005.
- [46] H. HIRSCHMÜLLER, P. R. INNOCENT et J. GARIBALDI, « Real-Time Correlation-Based Vision with Reduced Border Errors ». *IJCV*, 47(1-3):229-246, 2002.
- [47] J. HOLLAND, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, 1975.
- [48] L. HONG et G. CHEN, « Segment-Based Stereo Matching Using Graph Cuts ». Dans *IEEE CVPR*, volume 1, pages 74-81, 2004.
- [49] J. J. HOPFIELD, « Neural networks and physical systems with emergent collective computational abilities ». *Proceedings of the National Academy of Sciences*, 79:2254-2258, 1982.
- [50] R. HORAUD et O. MONGA, *Vision par ordinateur, outils fondamentaux*. Traité des nouvelles technologies, série Informatique. Hermès, 1993.
- [51] J. HU et P. SIY, « Stereo Correspondence through Multiple Constraint Neural Networks ». Dans *IEEE ICNN*, volume 2, pages 332-342, 1993.
- [52] H. ISHIKAWA et D. GEIGER, « Occlusions, Discontinuities, and Epipolar Lines in Stereo ». Dans *ECCV*, volume 1, pages 232-248, 1998.
- [53] C. V. JAWAHAR et P. J. NARAYANAN, « Generalised correlation for multi-feature correspondence ». *Pattern Recognition*, 35(6):1303-1313, 2002.
- [54] T. KANADE et M. OKUTOMI, « A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment ». *IEEE TPAMI*, 16(9):920-932, 1994.
- [55] S. KANEKO, I. MURASE et S. IGARASHI, « Robust Image registration by Increment Sign Correlation ». *Pattern Recognition*, 35(10):2223-2234, 2002.
- [56] C. KIM, K. M. LEE, B. T. CHOI et S. U. LEE, « A Dense Stereo Matching Using Two-Pass Dynamic Programming with Generalized Ground Control Points ». Dans *IEEE CVPR*, volume 2, pages 1075-1082, 2005.
- [57] V. KOLMOGOROV et R. ZABIH, « Multicamera Scene Reconstruction via Graph Cuts ». Dans *ECCV*, volume 2, pages 82-96, 2002.
- [58] J. KOSTKOVÁ et R. SÁRA, « Stratified Dense Matching for Stereoopsis in Complex Scenes ». Dans *BMVC*, volume 1, pages 339-348, 2003.
- [59] Z. D. LAN, « Méthodes robustes en vision : application aux appareillages visuels ». Thèse de doctorat, Institut National Polytechnique de Grenoble, 1997.
- [60] G. Le BESNERAIS et H. ORIOT, « Disparity estimation for high resolution stereoscopic reconstruction using GNC approach ». Dans *IEEE ICIP*, volume 2, pages 594-597, 1998.
- [61] S. H. LEE, Y. KANATSUGU et J.-I. PARK, « MAP-Based Stochastic Diffusion for Stereo Matching and Line Fields Estimation ». *IJCV*, 47(1-3):195-218, 2002.
- [62] M. LHUILLIER et L. QUAN, « Robust Dense Matching Using Local and Global Geometric Constraints ». Dans *IEEE ICPR*, volume 1, pages 968-972, 2000.
- [63] J. L. LOTTI et G. GIRAUDON, « Correlation algorithm with adaptive window for aerial image in stereo vision ». Dans *ESSRS Workshop*, pages 2315-2325, 1994.
- [64] D. G. LOWE, « Distinctive Image Features from Scale-Invariant Keypoints ». *IJCV*, 60(2):91-110, 2004.
- [65] A. LUO et H. BURKHARDT, « An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions ». *IJCV*, 15(3):171-188, 1995.
- [66] R. MANDUCHI et C. TOMASI, « Distinctiveness Maps for Image Matching ». Dans *ICIAI*, pages 26-31, 1999.
- [67] H. MAYER, « Analysis of Means to Improve Cooperative Disparity Estimation ». Dans *ISPRS*, pages 25-31, 2003.
- [68] A. MONTANVERT, P. MEER et A. ROSENFELD, « Hierarchical Image Analysis Using Irregular Tessellations ». *IEEE T-PAMI*, 13(4):307-316, 1991.
- [69] N. M. NASRABADI et C. Y. CHOO, « Hopfield Network for Stereo Vision Correspondence ». *IEEE T-PAMI*, 3(1):5-13, 1992.
- [70] Y.-I. OHTA et T. KANADE, « Stereo by Intra- and Inter-Scanline Search Using 18 Dynamic Programming ». *IEEE T-PAMI*, 7(2):139-154, 1985.
- [71] M. OKUTOMI et T. KANADE, « A Locally Adaptive Window For Signal Matching ». *IJCV*, 7(2):143-162, 1992.
- [72] N. PAPANODITIS, M. CORD, M. JORDAN et J.-P. COCQUEZ, « Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery ». *CVIU*, 72(2):122-142, 1998.
- [73] J. P. PASCUAL STARINK et E. BACKER, « Finding point correspondences using simulated annealing ». *Pattern Recognition*, 28(2):231-240, 1995.
- [74] J. PEARL, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.
- [75] W. K. PRATT, *Digital image processing*. Wiley-Interscience Publication, 1978.
- [76] C. RAO, A. GRITAI et M. SHAH, « Viewinvariant Alignment and Matching of Video Sequences ». Dans *IEEE ICCV*, volume 2, pages 939-945, 2003.
- [77] P. J. ROUSSEEUW et C. CROUX, *L1-Statistical Analysis and Related Methods*. Dans Y. Dodge, éditeur, *Explicit Scale Estimators with High Breakdown Point*, pages 77-92. Elsevier, 1992.
- [78] S. ROY et I. J. COX, « A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem ». Dans *IEEE ICCV*, pages 492-499, 1998.
- [79] H. SAITO et M. MORI, « Application of genetic algorithms to stereo matching of images ». *Pattern Recognition Letters*, 16(8):815-821, 1995.
- [80] D. SCHARSTEIN et R. SZELISKI, « A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms ». *IJCV*, 47(1):7-42, 2002.
- [81] P. SEITZ, « Using local orientational information as image primitive for robust object recognition ». Dans *Visual Communication and Image Processing IV*, volume SPIE-1199, pages 1630-1639, 1989.
- [82] J. SHAO, « Generation of Temporally Consistent Multiple Virtual Camera Views from Stereoscopic Image Sequences ». *IJCV*, 47(2):171-180, 2002.

- [83] J. SHEN et S. CASTAN, « An optimal linear operator for step edge detection ». *CVGIP*, 24(2):112-133, 1992.
- [84] J. SHI et C. TOMASI, « Good features to track ». Dans *IEEE CVPR*, pages 593-600, 1994.
- [85] M. SHIMIZU et M. OKUTOMI, « Sub-Pixel Estimation Error Cancellation on Area-Based Matching ». *IJCV*, 63(3):207-224, 2005.
- [86] S. SMITH et J. BRADY, « SUSAN – a new approach to low level image processing ». *IJCV*, 23(1):45-78, 1997.
- [87] C. SUN et S. PELEG, « Fast Panoramic Stereo Matching Using Cylindrical Maximum Surfaces ». *IEEE T-SMC*, 34(1):760-765, 2004.
- [88] R. SZELISKI et D. SCHARSTEIN, « Symmetric Sub-Pixel Stereo Matching ». Dans *ECCV*, volume 2, pages 525-540, 2002.
- [89] H. TAO, H. S. SAWHNEY et R. KUMAR, « A Global Matching Framework for Stereo Computation ». Dans *IEEE ICCV*, volume 2, pages 532-539, 2001.
- [90] M. F. TAPPEN et W. T. FREEMAN, « Comparison of Graph Cuts with Belief Propagation for Stereo, using Identical MRF Parameters ». Dans *IEEE ICCV*, volume 2, pages 900-907, 2003.
- [91] Q. TIAN et M. N. HUHN, « Algorithms for Subpixel Registration ». *CVGIP*, 35:220-233, 1986.
- [92] P. H. S. TORR et A. CRIMINISI, « Dense stereo using pivoted dynamic programming ». *Image and Vision Computing*, 22(10):795-806, 2004.
- [93] T. TUYTELAARS et L. Van GOOL, « Wide Baseline Stereo Matching based on Local, Anely Invariant Regions ». Dans *BMVC*, volume 1, pages 412-422, 2000.
- [94] F. ULLAH, S. KANEKO et S. IGARASHI, « Orientation Code Matching For Robust Object Search ». *IEICE Transactions on Information and Systems*, E-84-D(8):999-1006, 2001.
- [95] O. VEKSLER, « *Efficient graph-based energy minimization methods in computer vision* ». PhD thesis, Cornell University, États-Unis, 1999.
- [96] L. WANG, S. B. KANG, H.-Y. SHUM et G. XU, « Cooperative segmentation and stereo using perspective space search ». Dans *ACCV*, pages 366-371, 2004. 1996.
- [97] Y. WEI et L. QUAN, « Region-based Progressive Stereo Matching ». Dans *IEEE CVPR*, volume 1, pages 106-113, 2004.
- [98] Y. WEISS et W. T. FREEMAN, « On the optimality of solutions of the maxproduct belief-propagation algorithm in arbitrary graphs ». *IEEE T-IT*, 47(2):736-744, 2001.
- [99] J. WENG, N. AHUJA et T. S. HUANG, « Matching Two Perspective Views ». *IEEE T-PAMI*, 14(8):806-825, 1992.
- [100] J. XIAO et M. SHAH, « Two-Frame Wide Baseline Matching ». Dans *IEEE ICCV*, volume 1, pages 603-609, 2003.
- [101] Y. YANG, A. YUILLE et J. LU, « Local, Global, and Multilevel Stereo Matching ». Dans *IEEE CVPR*, pages 274-279, 1993.
- [102] K.-J. YOON et I.-S. KWEON, « Locally Adaptive Support-Weight Approach for Visual Correspondence Search ». Dans *IEEE CVPR*, volume 2, pages 924-931, 2005.
- [103] R. ZABIH et J. WOODFILL, « Nonparametric Local Transforms for Computing Visual Correspondence ». Dans *ECCV*, pages 151-158, 1994.
- [104] Y. ZHANG et C. KAMBHAMETTU, « Stereo Matching with Segmentation-based Cooperation ». Dans *ECCV*, volume 2, pages 556-571, 2002.
- [105] Z. ZHANG, « Le problème de la mise en correspondance : l'état de l'art ». Rapport de recherche INRIA RR-2146, 1993.
- [106] C. L. ZITNICK et T. KANADE, « A cooperative Algorithm for Stereo Matching and Occlusion Detection ». *IEEE T-PAMI*, 22(7):675-684, 2000.
- [107] B. ZITOV et J. FLUSSER, « Image registration methods: A survey ». *Image and Vision Computing*, 21(11):977-1000, 2003.
- [108] I. ZOGLAMI, O. FAUGERAS et R. DERICHE, « Traitement des occultations pour la modification d'objet plan dans une séquence d'image ». Dans *ORASIS*, pages 93-103, 1996.



Sylvie Chambon

Sylvie Chambon a obtenu son doctorat en Informatique de l'Université Paul Sabatier de Toulouse en 2005. Son travail portait sur la mise en correspondance d'images couleur en présence d'occultations. Elle effectue actuellement un séjour post-doctoral en imagerie thoracique sur le recalage 3D d'images TDM (tomodensitométriques) et TEP (tomographie par émission de positons) à l'ENST (École Nationale Supérieure des Télécommunications).



Alain Crouzil

Alain Crouzil a obtenu son doctorat en Informatique à l'Université Paul Sabatier de Toulouse en 1997. Il est, depuis 1998, Maître de Conférences à l'Université Paul Sabatier et membre de l'équipe TCI (Traitement et Compréhension d'Images) de l'IRIT (Institut de Recherche en Informatique de Toulouse). Ses recherches portent sur la stéréovision, le calibrage et l'autocalibrage de caméra et le *Shape From Shading*.