

Appariements de régions pour l'estimation de cartes de profondeur

Region matching for depth map estimation

Jean-Charles Bricola

Michel Bilodeau

Serge Beucher

MINES ParisTech, PSL Research University
CMM – Centre de Morphologie Mathématique
35 rue St Honoré, 77300 Fontainebleau, France

jean-charles.bricola@mines-paristech.fr

Résumé

En vision stéréoscopique binoculaire, les coordonnées spatiales d'un point de la scène peuvent être déduites de ses projections dans les deux images formées. Trouver les projections qui se correspondent constitue cependant la principale difficulté du problème et ce pour deux raisons : l'ambiguïté des appariements au travers des régions homogènes et l'impossibilité de trouver une correspondance valide en cas d'occultation.

Cet article propose une méthode d'estimation de cartes de profondeur utilisant l'appariement de régions plutôt que de points. Ces appariements sont effectués à deux échelles de segmentation obtenues au moyen de critères morphologiques assurant l'existence d'une relation hiérarchique entre les partitions fine et grossière. La hiérarchie peut alors être exploitée afin d'obtenir des cartes de disparité régionales fines non bruitées desquelles les cartes de profondeur finales sont déduites.

Enfin, nous montrons comment cette méthode s'adapte à différents types d'images stéréoscopiques : des images fortement texturées acquises dans les mêmes conditions d'illumination à celles ayant une faible dynamique de disparités et dont les contours constituent la seule source d'information pertinente.

Mots Clef

stéréo, appariements de régions, hiérarchies de segmentation, ligne de partage des eaux

Abstract

Given a pair of stereo images, the spatial coordinates of a scene point can be derived from its projections onto the two considered image planes. Finding the correspondences between such projections however remains the main difficulty of the depth estimation problem : the matching of points across homogeneous regions is ambiguous and occluded points cannot be matched as their projections do not exist in one of the image planes.

Instead of searching for dense point correspondences, this article proposes an approach to the estimation of depth

map which is based on the matching of regions. The matchings are performed at two different levels of segmentation obtained by morphological criteria which ensure the existence of an hierarchy between the coarse and fine partitions. The hierarchy is then exploited to compute fine regional disparity maps which are accurate and free from noisy measurements.

We finally show how this method can be employed on a variety of stereo images : from highly textured and constant illumination databases to low-baseline stereo images which relevant information resides in the contours only.

Keywords

stereo, region matching, hierarchical segmentation, watershed

1 Introduction

L'estimation de cartes de profondeur à partir de données stéréoscopiques est un problème étudié depuis les années 70. Il revient le plus souvent à un problème de mise en correspondance, les profondeurs pouvant être déduites des appariements de pixels entre les images stéréo. Dans la suite de cet exposé, on se place dans une configuration où les images sont rectifiées : tous les points de la scène se projettent alors dans les images avec la même ordonnée, mais des abscisses différentes. La différence d'abscisse entre les points appariés, appelée aussi *disparité*, est alors inversement proportionnelle à la profondeur recherchée [12].

Les premiers travaux sur ce sujet partent des appariements de points remarquables et tentent de trouver une fonction de profondeur passant par ces points [10]. Otha [16] propose également une méthode appariant des points le long des contours détectés tout en rendant les résultats obtenus cohérents entre les différentes lignes épipolaires. Ces méthodes sont intéressantes dans la mesure où elles tiennent compte de données fiables dans le processus d'estimation. Mais leur limitation réside dans le fait que le nombre de singularités disponibles est généralement insuffisant.

Les méthodes actuelles sont basées sur des correspon-

dances de pixels dites *denses* et s'inscrivent dans un *framework* maintes fois éprouvé [18]. La première étape consiste à déterminer les coûts d'appariements de points pour chaque ligne de l'image et chaque disparité. Le coût est basé sur une mesure de dissimilarité entre les patches ayant pour centres les pixels mis en correspondance. Le lecteur trouvera un inventaire de ces mesures dans [9]. Les composantes d'un patch peuvent être pondérées en fonction de leur appartenance à un même objet. DeMaeztu [6] utilise pour cela une distance basée sur la couleur. La deuxième étape consiste à veiller à la cohérence des disparités entre les différentes lignes de recherche. Dans les méthodes locales réputées pour leur rapidité d'exécution, les coûts sont simplement diffusés entre les lignes de recherche. La disparité retenue pour un pixel donné sera celle qui minimise le coût d'agrégation. Dans les méthodes globales, on associe à la carte de disparité recherchée une énergie tenant compte du coût de disparité local et de la cohérence de voisinage, tolérant ainsi les discontinuités en fonction du gradient de l'image [8].

L'exploitation des régions dans les algorithmes d'estimation de profondeur gagne en popularité. Zitnick [22] propose une méthode effectuant des appariements sur une sur-segmentation de l'image en imposant un critère de lissage sur les segments ayant des distributions couleur similaires. A l'opposé, on trouve des méthodes partant de cartes de disparités bruitées afin d'estimer des plans de profondeur, lesquels sont ensuite segmentés et associés à une équation de plan [5, 20]. A l'inverse de la méthode basée sur les superpixels, ces approches utilisant des régions orientées objet ont deux qualités principales : définir les frontières des objets [2] et définir l'appartenance d'un pixel occulté à une région semi-occultée [15, 13].

Par rapport aux méthodes existantes, l'approche régionale proposée dans cet article se distingue en trois points, représentant les contributions principales de ce travail :

- Aucune disparité n'est déduite d'un appariement de pixels. Ces derniers peuvent en effet compromettre les résultats sur des images difficiles (bruit, présence des motifs périodiques, régions fortement homogènes, etc.), les méthodes existantes partant implicitement de l'hypothèse qu'il existe une majorité d'appariements corrects.
- Les disparités sont estimées en cascade, à deux niveaux d'une hiérarchie de segmentations. Cela permet de choisir le niveau de précision ou la rapidité avec laquelle la carte de disparités doit être produite en fonction de l'application.
- L'approche intègre les contours dans le processus d'estimation. Souvent, les disparités de contour sont mal exploitées, car elles nécessitent la connaissance d'appartenance du contour à un objet. Ce problème a été peu soulevé à notre connaissance, en dehors des travaux de [19] exploitant la structure des jonctions de frontières.

Aperçu de l'approche. L'appariement d'une région est associé à une *disparité régionale* estimant la disparité moyenne de ses pixels entre les deux images stéréosco-

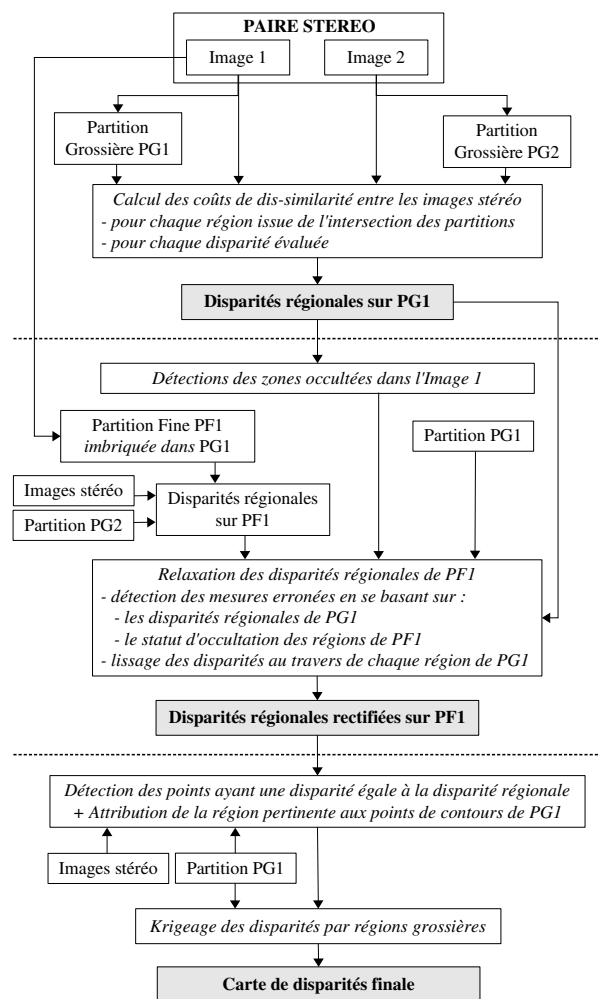


FIGURE 1 – Présentation des étapes d'estimation et d'affinage de cartes de profondeurs réalisées dans cette approche.

piques rectifiées. Les cartes de disparités régionales fournissent donc une approximation des cartes de disparités recherchées. La perte de précision provient notamment des larges régions de l'image qui ne sont pas fronto-parallèles à la caméra.

Les cartes de disparités régionales sont calculées à deux niveaux de segmentation comme le montre le diagramme présenté en figure 1. Au niveau grossier les appariements régionaux sont plus fiables du fait de la singularité des régions segmentées. Les disparités régionales résultantes mettent en relief l'ordre de profondeur entre les objets segmentés. Au niveau fin les erreurs d'appariements sont plus fréquentes mais les variations de disparités sont mieux capturées du fait que les régions soient plus petites. Un processus de relaxation est ensuite appliqué sur les disparités régionales fines. Le processus s'effectue de façon indépendante pour chaque ensemble de régions fines incluses dans la même région grossière et tient compte des disparités régionales grossières dans la détection des mesures de dispa-

Opérateurs morphologiques

B and H sont des éléments structurants

H est un élément structurant isotrope de taille élémentaire

Symbole	Description
$\delta_B(f)$	Dilatation de f : $\delta_B(f)[x] = \sup_{x \in B} f[x]$
$\varepsilon_B(f)$	Érosion de f : $\varepsilon_B(f)[x] = \inf_{x \in B} f[x]$
$D_g^1(f)$	Dilatation géodésique de f dans le masque g : $D_g^1(f) = \inf(\delta_H(f), g)$
$R_g(f)$	Reconstruction géodésique de g à partir du marqueur f t.q. $f \leq g$: $R_g(f) = D_g^{+\infty}(f) = D_g^1(\dots(D_g^1(f))\dots)$
$R_g^*(f)$	Reconstruction géodésique duale de g à partir du marqueur f t.q. $f \geq g$: $R_g^*(f) = -R_{-g}(-f)$

TABLE 1 – Notation

rité erronées.

La viabilité de cette approche repose sur l'utilisation adéquate de méthodes de segmentations basées sur la Ligne de Partage des Eaux contrôlée par des marqueurs. Les marqueurs sont en effet obtenus par un mécanisme d'inondation permettant d'obtenir une relation hiérarchique entre les régions des partitions fine et grossière. En outre, les marqueurs peuvent être transférés entre deux images stéréo produisant ainsi des segmentations stéréo équivalentes, utiles par exemple, pour les appariements de points de contour utilisés dans le lissage de la carte de profondeur.

2 Segmentations hiérarchiques

La segmentation basée sur la *Ligne de Partage des Eaux* [3] ou LPE est particulièrement bien adaptée à l'utilisation de marqueurs désignant les zones de l'image visées par l'extraction de régions. Les marqueurs sont des composantes connexes labellisées habituellement placées dans les creux de la surface topographique à inonder, typiquement le gradient de l'image. Des marqueurs naissent des lacs labellisés grandissant au fil de l'inondation de la surface topographique. Les points de rencontre de lacs ayant des labels différents conduisent à la formation d'une ligne de partage de eaux constituant alors la segmentation de l'image.

Notre objectif revient donc à choisir les marqueurs de façon judicieuse afin d'obtenir une segmentation grossière mettant en valeur les structures de l'image les plus saillantes et une segmentation fine parfaitement imbriquée dans la segmentation grossière.

2.1 Marqueurs issus des h -minima

Soit g une fonction transformant un point de l'image $p = (x, y)$ en une altitude $g[p]$. Dans notre application, g correspond à la magnitude du gradient couleur de l'image d'entrée. Les h -minima de g sont définis par l'équation 1 :

$$\mathcal{M}(g, h)[p] = \begin{cases} 1 & \text{si } (R_g^*(g + h) - g)[p] > 0 \\ 0 & \text{sinon} \end{cases} \quad (1)$$

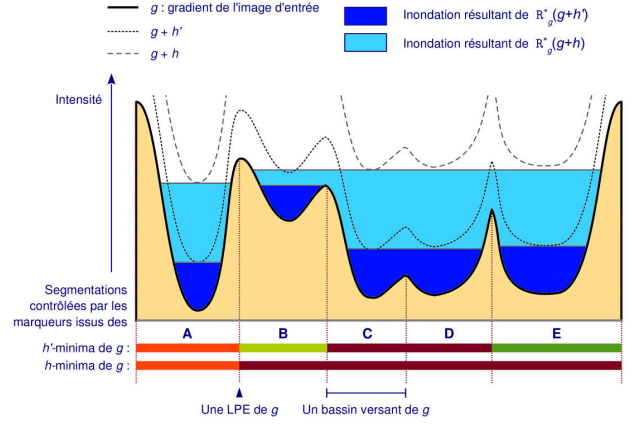


FIGURE 2 – h -minima d'une surface topographique g et hiérarchies de segmentations induites par les marqueurs issus des h - et des h' -minima pour $h' < h$. L'inondation induite par $R_g^*(g + h')$ produit les lacs représentés en bleu foncé. Ces lacs ont une profondeur égale à h' à l'exception du lac contenu dans le bassin versant B, ce dernier étant absorbé par le lac chevauchant les bassins versants C et D.

Marqueurs de contraste. $R_g^*(g + h)$ représente la surface topographique obtenue après inondation des bassins versants de g jusqu'à ce que les lacs ainsi formés aient une profondeur égale à h , ou à défaut soient absorbés par d'autres bassins versants. La figure 2 illustre ce phénomène.

Tous les bassins versants voisins ne pouvant contenir chacun un lac de profondeur au moins égale à h sont donc recouverts par une seule composante connexe issue des h -minima de g . Cette propriété permet donc de fusionner les bassins versants entre lesquels le contraste est faible, h contrôlant l'importance du contraste.

Marqueurs de volume. Par rapport à l'inondation directement induite par les h -minima, les inondations volumiques retardent la fusion des lacs de grande envergure et conduisent à des hiérarchies de segmentation plus cohérentes vis-à-vis de la perception humaine [21]. On peut simuler cette inondation en faisant varier la valeur de h au niveau de chaque bassin versant jusqu'à ce que le volume des lacs recherché soit atteint, ou, à défaut, que les lacs soient absorbés. Les marqueurs de volume finalement retenus correspondent aux lacs non absorbés.

Hiérarchie induite. Les propriétés des h -minima sont présentées en détail dans [4]. On y trouve notamment la preuve que $\mathcal{M}(g, h') \subseteq \mathcal{M}(g, h)$ pour tout $h' < h$. Il en résulte deux choses. Premièrement les minima de la surface inondée sont toujours contenus dans ses h -minima. Il suffit donc de transférer à chaque minima le label du h -minimum dans lequel il est contenu afin d'obtenir la segmentation contrôlée par les marqueurs résultants des h -minima. Deuxièmement, si deux minima ne sont pas contenus dans le même h -minimum, il leur est impossible d'ap-

partenir au même h' -minimum. De fait, deux minima héritant des labels différents des h -minima héritent des labels différents des h' -minima. On en déduit que la LPE issue des h -minima est incluse dans la LPE issue des h' -minima conduisant à une relation hiérarchique entre les deux partitions ainsi obtenues.

3 Disparités régionales

La disparité régionale est une mesure attribuée à une région de l'image de référence. Elle vise à estimer la disparité moyenne des pixels formant la région considérée. Nous cherchons pour cela le déplacement que doit effectuer une région de l'image de référence pour être en superposition optimale avec l'image ciblée.

3.1 Mesures de similarité

On représente une région \mathbf{R} par une fonction indicatrice R telle que $R[x, y] = 1 \Leftrightarrow (x, y) \in \mathbf{R}$. On note $\mathbf{R}^{(d)}$ la région obtenue en décalant la région \mathbf{R} de d pixels dans le sens de l'axe horizontal, et dont la fonction indicatrice est définie par $R^{(d)}[x, y] = R[x - d, y]$.

Les partitions grossières pour l'image de référence et l'image sœur de la paire stéréo sont notées P_{G1} et P_{G2} respectivement. On note également P_{F1} la partition fine obtenue pour l'image de référence. Chaque partition est simplement décrite par l'ensemble des régions qui la compose. La disparité d^* est assignée à une région de P_{G1} si la superposition avec une région de P_{G2} décalée de d^* pixels minimise le coût de dissimilarité calculé à partir des critères suivants que l'on choisira en fonction des conditions d'acquisition des images stéréoscopiques :

Critères de recouvrement. La distance de Jaccard [14] mesure le recouvrement de deux régions en calculant le ratio entre l'aire de leur intersection et l'aire de leur union, soit :

$$c_{\text{Jaccard}}(d, \mathbf{R}_i, \mathbf{R}_j) = 1 - \frac{|\mathbf{R}_i \cap \mathbf{R}_j^{(d)}|}{|\mathbf{R}_i \cup \mathbf{R}_j^{(d)}|} \quad (2)$$

Si la segmentation grossière P_{G2} comporte significativement moins de régions que P_{G1} , par exemple en raison d'une inondation plus importante lors de la génération de P_{G2} , il peut être préférable d'utiliser un ratio non symétrique en remplaçant $|\mathbf{R}_i \cup \mathbf{R}_j^{(d)}|$ par $|\mathbf{R}_i|$ dans l'équation 2.

Critère d'illumination. La Somme des Différences Absolues (SAD) est fréquemment utilisée sur des images acquises dans les mêmes conditions d'illumination. En raison des semi-occultations que subissent les régions entre les images stéréo, il est préférable de calculer la SAD sur des intersections de régions plutôt que sur une région tout entière de l'image de référence. On définit la moyenne des différences absolues entre régions par l'équation 3 :

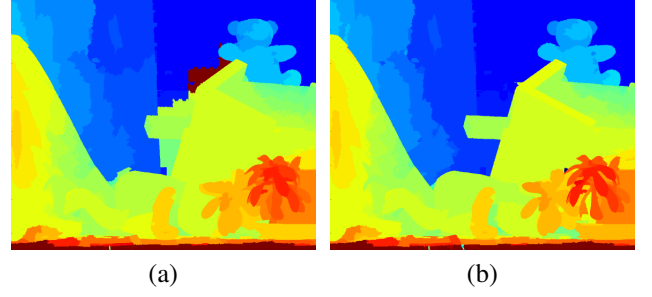


FIGURE 3 – Cartes de disparités régionales grossières obtenues pour *Teddy* (a) Sans partition contrainte, (b) Avec partition contrainte P_{G2} . On remarque que les régions semi-occultées sont moins affectées par les erreurs d'appariements avec l'utilisation d'une partition contrainte.

$$\begin{aligned} \mathbf{R}_{i,j}^{(d)} &= \mathbf{R}_i \cap \mathbf{R}_j^{(d)} \\ G_{1,2}^{(d)} &= |I_1[x, y] - I_2[x - d, y]| \\ c_{\text{Illum}}(d, \mathbf{R}_i, \mathbf{R}_j) &= \frac{1}{|\mathbf{R}_{i,j}^{(d)}|} \sum_{(x,y) \in \mathbf{R}_{i,j}^{(d)}} G_{1,2}^{(d)}[x, y] \quad (3) \end{aligned}$$

où I_1 et I_2 représentent les deux images de la paire stéréo en niveaux de gris prenant ses valeurs dans une échelle redimensionnée entre 0 et 1. On peut bien entendu étendre l'équation 3 aux images couleur, en moyennant les coûts obtenus sur les différents canaux couleur.

Combinaison recouvrement + illumination. Les coûts obtenus en résolvant l'équation 3 pour des régions se recouvrant à peine sont peu significatifs et peuvent néanmoins être suffisamment bas pour provoquer un appariement. La combinaison des critères de recouvrement et d'illumination consiste à ne pas tenir compte de telles intersections.

La disparité régionale $d^*(\mathbf{R}_i)$ attribuée à une région \mathbf{R}_i appartenant à P_{G1} ou à P_{F1} est définie par la relation 4, soit :

$$\begin{aligned} c_A &= \max(\Theta(d, \mathbf{R}_i, \mathbf{R}_j), c_{\text{Illum}}(d, \mathbf{R}_i, \mathbf{R}_j)) \\ d^*(\mathbf{R}_i) &= \arg \min_d c_A(d, \mathbf{R}_i, \mathbf{R}_j), \forall \mathbf{R}_j \in P_{G2} \quad (4) \end{aligned}$$

La variante asymétrique du critère de recouvrement est ici utilisée et on définit Θ par une fonction indicatrice égale à 1 lorsque $c_{\text{Jaccard}}(d, \mathbf{R}_i, \mathbf{R}_j) > 0.75$. Les coûts obtenus au niveau des intersections recouvrant moins de 25% de la région d'origine de l'image de référence sont pénalisés.

Le fait de calculer les coûts pour chaque intersection et donc d'utiliser une partition grossière P_{G2} permet de tenir compte du phénomène d'occlusion comme en témoigne la figure 3.

Critère de contours. Dans les images peu texturées ou encore acquises dans des conditions d'illumination différentes, la mise en correspondance des gradients peut

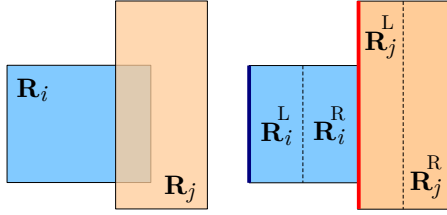


FIGURE 4 – Attribution de disparités calculées sur le gradient. Chaque région est coupée le long de son squelette vertical en deux sous-régions. Lorsqu’une région est semi-occultée, c’est-à-dire qu’un morceau de contour correspond à sa frontière physique, l’autre à une occlusion, on constate deux choses : (1) La sous-région ayant la disparité la plus faible est représentative de la profondeur de la région d’origine et (2) la sous-région ayant la disparité la plus forte hérite, en cas d’occlusion, de la disparité d’une région voisine.

constituer le critère le plus pertinent. Les gradients sont cependant liés aux contours des objets segmentés, or certains de ces contours, pour un objet donné, peuvent ne pas correspondre à la frontière physique de l’objet mais à une frontière d’occultation. Il est donc nécessaire de tenir compte de cette observation dans l’attribution des disparités régionales.

Pour ce faire, chaque région \mathbf{R}_i de l’image de référence est découpée en deux sous-régions le long de son squelette vertical, soit \mathbf{R}_i^L et \mathbf{R}_i^R . On calcule leurs disparités régionales selon l’équation 5 :

$$d^*(\mathbf{R}) = \arg \min_d \sum_{(x,y) \in \mathbf{R}} |g_1[x, y] - g_2[x - d, y]| \quad (5)$$

où R est la région considérée, g_1 et g_2 les gradients des deux images stéréoscopiques.

Une heuristique déduite des observations de la figure 4 est appliquée afin de corriger les disparités issues des contours d’occlusion : lorsque $d^*(\mathbf{R}_i^R) \gg d^*(\mathbf{R}_i^L)$ et $d^*(\mathbf{R}_i^R) \simeq d^*(\mathbf{R}_j^L)$ tels que \mathbf{R}_i^R et \mathbf{R}_j^L sont voisins, la disparité $d^*(\mathbf{R}_i^R)$ est transférée à \mathbf{R}_i^L . Et vice-versa.

3.2 Relaxation des disparités de la segmentation fine

L’objectif du processus de relaxation est double. Premièrement, on souhaite obtenir une évolution des disparités de la partition fine qui soit suffisamment lisse tout en permettant les discontinuités au niveau des frontières d’objets. Pour cette raison on applique le processus de relaxation de manière indépendante sur chaque région grossière. Deuxièmement, les disparités issues du processus de relaxation doivent tenir compte des mesures réalisées sur la partition fine, c’est-à-dire des disparités régionales brutes obtenues pour la partition fine. Toutefois, certaines de ces mesures sont erronées et risquent de compromettre le processus de relaxation. Il faut donc un mécanisme permettant de les détecter.

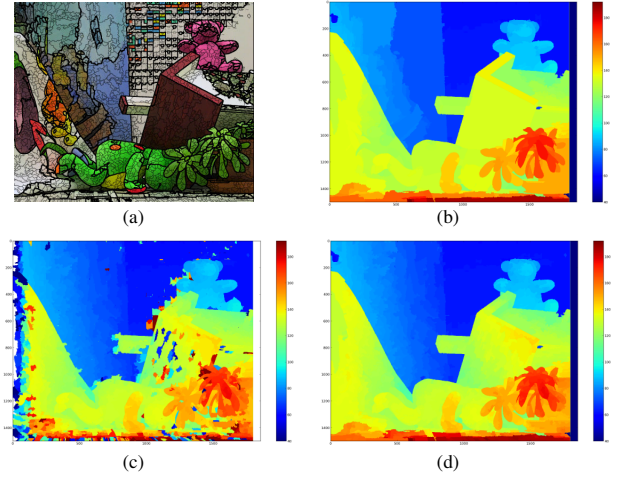


FIGURE 5 – Effet du processus de relaxation sur les disparités régionales de P_{F1} . (a) Segmentations grossières et fines emboîtées, (b) Disparités régionales de la partition P_{G1} servant à détecter les mesures erronées dans (c) Disparités régionales brutes de la partition P_{F1} . (d) Disparités régionales rectifiées suite au processus de relaxation sur la partition fine.

Dans cette optique un champ Markovien est défini pour chaque région grossière. Les nœuds du champ représentent les régions fines comprises dans la région grossière étudiée, tandis que les arêtes représentent les relations d’adjacence entre ces régions. À chaque champ Markovien est associée une fonction objective composée d’une somme d’énergies unaires pénalisant la prise d’une valeur de disparité qui serait trop éloignée de la mesure d’origine, et d’une somme d’énergies de voisinages pénalisant des valeurs de disparités trop éloignées entre régions voisines. Les disparités issues de l’algorithme de relaxation minimisent cette fonction objective. On utilisera pour cela l’algorithme de coupe minimale présenté dans [17].

Termes d’énergie. Soit U_i le terme d’énergie unaire associé à la région \mathbf{R}_i et $P_{i,j}$ le terme d’énergie de voisinage associé aux régions adjacentes \mathbf{R}_i et \mathbf{R}_j . On appelle $d^\circ(\mathbf{R})$ la disparité associée à la région \mathbf{R} suite au processus de relaxation.

Le terme de voisinage augmente plus la différence de disparités prises entre deux régions voisines est importante. Il est défini par :

$$P_{i,j} = (d^\circ(\mathbf{R}_i) - d^\circ(\mathbf{R}_j))^2$$

Le terme unaire doit tenir compte de la disparité initialement mesurée, sauf si celle-ci est erronée. On considère que la mesure de disparité régionale est peu fiable lorsque :

- une région de P_{F1} est fortement occultée. Les occlusions peuvent être estimées à partir de la carte de disparités régionale obtenue pour P_{G1} .
- la mesure de disparité pour une région de P_{F1} est manifestement très éloignée de la disparité régionale d_G de la région grossière concernée. On aura au préalable défini

une tolérance d'écart par rapport à la disparité régionale de la région grossière en fonction de son aire. Soit $\alpha_i = 1 - \bar{\alpha}_i$, une variable égale à 1 si la mesure $d^*(\mathbf{R}_i)$ est considérée comme étant fiable, 0 sinon. Le terme d'énergie unaire est alors défini comme suit :

$$U_i = \alpha_i |d^*(\mathbf{R}_i) - d^\circ(\mathbf{R}_i)| + \bar{\alpha}_i |d_G - d^\circ(\mathbf{R}_i)|$$

On peut voir l'effet du processus de relaxation sur les disparités régionales de la partition fine en figure 5.

4 Segmentations stéréo conjointes

Cette section présente un mécanisme générant des segmentations équivalentes entre deux images stéréoscopiques. De telles segmentations sont utilisées afin de faciliter l'appariement des points de contours présenté en section 5.

La co-segmentation morphologique est introduite dans [3] pour des images de microscopie à balayage électronique de fracture métallique. L'idée principale consiste à propager les marqueurs obtenus pour l'une des images à l'autre image de la paire stéréoscopique. Dans notre approche, le transfert des marqueurs est guidé par les disparités régionales calculées en section 3 et s'effectue en deux étapes :

Estimation de la partition équivalente. Étant donné le partitionnement de l'image de référence donné par la carte de labels \mathcal{L}_1 , tel que $\mathcal{L}_1[x, y] = i \Leftrightarrow R_i[x, y] = 1$, on estime la carte de labels \mathcal{L}_2 de la deuxième image en fonction des disparités régionales calculées précédemment, selon l'équation 6.

$$\mathcal{L}_2[x, y] = \arg \max_i \{d^*(\mathbf{R}_i) \times R_i[x + d^*(\mathbf{R}_i), y]\} \quad (6)$$

Littéralement, chaque région \mathbf{R}_i composant la partition de l'image de référence est décalée en fonction de sa disparité régionale $d^*(\mathbf{R}_i)$. Ces décalages n'étant pas de même intensité, certaines régions se retrouvent alors superposées. Dans la partition équivalente de la deuxième image, seule la région ayant la profondeur la plus faible sera visible au niveau de la superposition. Donc on attribue à $\mathcal{L}_2[x, y]$ le label de la région qui restera visible au point (x, y) . Les autres régions au même point seront occultées.

Segmentation équivalente. La segmentation équivalente s'obtient en calculant une ligne de partage des eaux contrôlée par des marqueurs préservant les labels issus de \mathcal{L}_2 , l'estimation de la partition équivalente recherchée. L'ensemble des marqueurs s'obtient en calculant les minima du gradient de la deuxième image, permettant de s'assurer que la segmentation équivalente soit bien incluse dans la LPE du gradient de la deuxième image. Chaque pixel (x, y) appartenant à l'ensemble des marqueurs se voit attribuer le label $\mathcal{L}_2[x, y]$. La préservation des labels au cours du processus de co-segmentation fournit de plus les appariements entre les régions des partitions obtenues comme le montre la figure 6.

Cette méthode suppose cependant que le déplacement entre les deux images stéréo ne soit pas trop important, c'est-à-

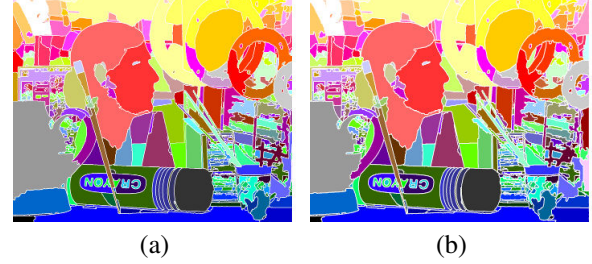


FIGURE 6 – Co-segmentation morphologique asymétrique de l'image Art de la base de Middlebury. (a) Segmentation de l'image de référence. (b) Co-segmentation équivalente de la deuxième image. Les labels attribués aux marqueurs des régions segmentées conduisent aux appariements régionaux.

dire qu'il n'y ait aucune région qui soit complètement occultée dans l'image de référence, faute de quoi, la segmentation conjointe de la deuxième image fusionnera la dite région avec une région non-occultée.

5 Lissage de cartes de profondeur

Le mécanisme de lissage de cartes de profondeur employé dans cette approche est basé sur le *krigeage ordinaire*. À partir d'un ensemble de points pour lesquels les valeurs de disparités sont connues et en admettant que la variabilité entre les disparités de deux points soit proportionnelle à leur distance euclidienne, le krigeage calcule un estimateur non biaisé dont la variance de l'erreur d'estimation est minimale.

Le but de cette section n'est pas de présenter la technique de krigeage en soit. Le lecteur trouvera un exposé détaillé du krigeage ordinaire et de son implémentation dans [7]. Nous nous concentrons ici sur l'extraction des points servant au krigeage, appelés *échantillons*.

5.1 Extraction des échantillons

Le but de l'opération est d'extraire et d'associer des points caractéristiques entre les deux images de la paire stéréo. Seules les associations de points conduisant à une disparité égale à la disparité régionale de la région fine les contenant sont finalement retenues. On dispose ainsi pour chaque région grossière de points d'ancrage permettant au mécanisme d'interpolation de localiser avec précision les disparités régionales calculées en section 3.

On distingue deux types de points : ceux qui sont internes à la région et ceux appartenant au contour d'une région.

Points internes. Les points internes aux régions grossières sont extraits au moyen d'une transformée haut-deforme permettant d'extraire les pics et le creux significatifs de la surface topographique représentant l'intensité de l'image. Les points détectés sont appariés en fonction d'un critère de corrélation et uniquement si la validation croisée [8] est vérifiée. Seuls les points ayant une disparité égale à la disparité régionale de leur région fine sont finalement retenus dans le processus de krigeage.

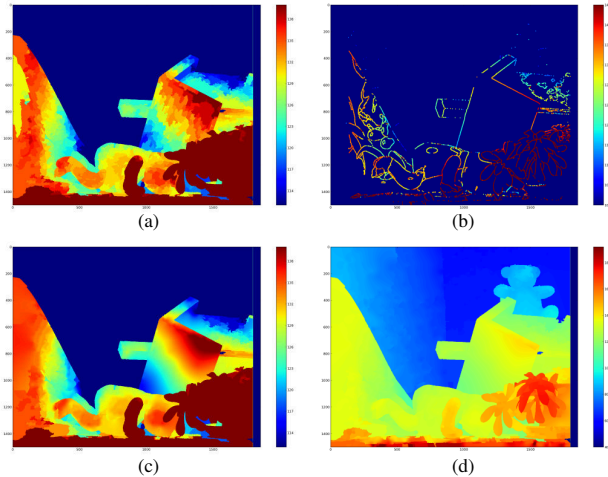


FIGURE 7 – (a) Disparités régionales de la partition fine. (b) Disparités des points de contours issus de la partition grossière et appariés à l’aide de la co-segmentation morphologique. (c) Disparités obtenues par krigeage des points remarquables ayant une disparité égale à la disparité régionale correspondante. (d) Disparités obtenues par krigeage (visualisation couvrant la gamme complète de disparités).

Points de contour. Les points de contour appartiennent aux LPE des deux images stéréo obtenues par les segmentations conjointes dont l’algorithme est donné en section 4. Une des qualités des segmentations conjointes est de pouvoir identifier et ordonner spatialement les régions qui sont adjacentes à un point d’une LPE apportant ainsi une forte contrainte dans l’appariement des points de contour.

Par construction, la LPE peut se trouver à cheval entre deux bassins versants, ce qui représente un inconvénient car le point de contour doit impérativement être attribué à la bonne région, celle dont la frontière physique correspond effectivement au contour. Une manière simple de procéder consiste à ne garder que les points dont la disparité est égale à la disparité régionale des régions fines correspondantes, à l’instar des points internes. Pour des applications partant de disparités régionales grossières et n’appliquant aucun seuillage sur les disparités des points détectés, l’approche consisterait à re-labelliser les partitions le long des LPE avec le label de la région adjacente ayant la plus forte disparité.

6 Évaluation

L’évaluation que nous proposons rapporte et analyse les résultats obtenus sur la base de Middlebury en terme de précision. Notre méthode est ensuite testée sur une paire d’images bruitée et contenant de nombreuses régions homogènes. Enfin, les perspectives de développement sont exposées tout au long de cet exposé.

Précision. L’application de cette approche sur la base stéréo v.2 de Middlebury [1] donne les résultats présentés en table 2. Les pixels considérés comme étant mal appariés

	Tolérance ± 0.5 px.			Tolérance ± 2 px.		
	<i>NonOcc</i>	<i>All</i>	<i>Disc</i>	<i>NonOcc</i>	<i>All</i>	<i>Disc</i>
Tsukuba	10.1	10.7	23.7	3.15	3.40	12.3
Venus	9.06	9.50	15.5	0.65	0.80	4.94
Teddy	14.2	20.2	28.6	4.25	7.76	10.7
Cones	13.5	18.8	23.0	3.80	8.04	8.36
Moyenne	16.4			5.38		

TABLE 2 – Pourcentages de pixels mal appariés selon différentes tolérances d’erreur sur la base de Middlebury

sont ceux dont la disparité diffère d’un certain seuil de la vérité terrain [18].

Il y a deux principales sources d’imprécision. La première est liée aux bords de l’image de référence qui, dans le cas de *Teddy* et *Cones* sont fortement occultés. Il est possible de détecter les régions qui disparaissent dans la deuxième image au moyen de la co-segmentation morphologique, puis de les fusionner avec les régions voisines les plus appropriées avant d’effectuer le krigeage, ce que l’on peut observer sur la figure 7(c). Cependant aucune hypothèse n’est actuellement formulée sur les objets composant la scène. La deuxième source d’imprécision provient des régions dont l’inclinaison est très importante, comme le sol dans l’image *Teddy*, cf. figure 7(d). La recherche d’une superposition optimale entre de telles régions demeure en effet instable, puisque la transformation entre ces régions homologues s’apparente de moins en moins à une simple translation. Il faudrait donc améliorer la robustesse du mécanisme d’appariement en tenant compte de transformations géométriques plus complexes.

On remarque cependant qu’à performances égales avec d’autres approches, la qualité des cartes en termes de perception apparaît comme étant supérieure, et ce même au niveau grossier. Notre sensibilité au contraste entre les différents plans de profondeur est en effet prédominante et cela ne ressort pas dans le critère de performance évaluant la précision pixel par pixel, de façon indépendante.

Micro-stéréoscopie. Nous comparons notre méthode à l’algorithme de Hirschmuller [11] sur une paire d’images stéréo ayant les caractéristiques suivantes : faible gamme de disparités, illuminations différentes, régions fortement homogènes, bruit, objets semi-transparents. Notre algorithme utilise un chaînage d’inondations volumiques et de contraste pour obtenir la partition de l’image de référence ainsi que le critère de corrélation basé sur le gradient pour les appariements régionaux. Les résultats sont présentés en figure 8. On remarquera la robustesse des disparités régionales face aux régions homogènes et à la préservation des contours.

Conclusion

Nous avons présenté une approche d’estimation de cartes de profondeur contrôlée par l’appariement et la superposition de régions. Les cartes de profondeur sont estimées à

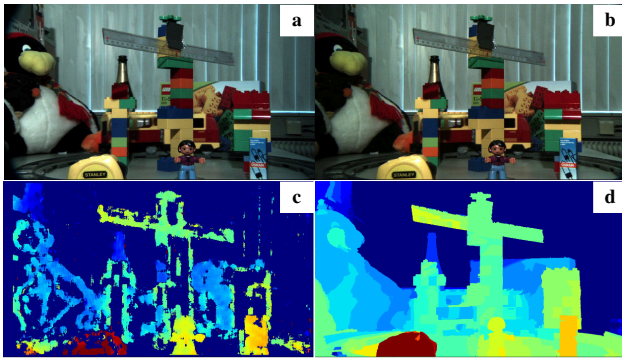


FIGURE 8 – Résultats obtenus en micro-stéréoscopie sur des images bruitées. (a) image de référence, (b) deuxième image, (c) Carte de disparités obtenue avec [11], (d) Carte de disparités régionales au niveau grossier.

deux échelles de segmentation, d'abord à un niveau grossier puis à un niveau fin. Les disparités finales sont déduites au moyen d'un procédé d'interpolation, le krigage, reposant sur des points émetteurs ayant une disparité égale à la disparité régionale obtenue sur la segmentation fine. Les points de contour sont également exploités une fois leur appartenance à la région adéquate établie. Enfin, la viabilité de cette approche repose sur l'utilisation de la segmentation morphologique, que ce soit pour l'obtention de partitions imbriquées ou encore le calcul de segmentations stéréoscopiques équivalentes.

Cette méthode offre également de bons résultats en terme de précision et de perception et ouvre la voie au traitement d'images stéréo non idéales. La possibilité de choisir le niveau de précision attendu est particulièrement intéressante pour des applications devant minimiser leur temps de calcul. On pourrait par exemple restreindre l'affinage des disparités aux régions ayant une faible profondeur. Nos travaux courants portent sur l'exploitation de ces mécanismes dans le traitement de séquences vidéo stéréoscopiques.

Acknowledgements

This work has been performed in the project PANORAMA, co-funded by grants from Belgium, Italy, France, the Netherlands, and the United Kingdom, and the ENIAC Joint Undertaking.

Références

[1] Middlebury stereo database. <http://vision.middlebury.edu/stereo/>.

[2] T. Aydin and Y. S. Akgul. Stereo depth estimation using synchronous optimization with segment based regularization. *Pattern Recognition Letters*, 31(15) :2389–2396, 2010.

[3] S. Beucher. *Segmentation d'Images et Morphologie Mathématique*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 1990.

[4] S. Beucher. Maxima and minima : a review. Technical report, Centre of Mathematical Morphology - Mines Paris-Tech, 2013.

[5] M. Bleyer and M. Gelautz. A layered stereo matching algorithm using image segmentation and global visibility constraints. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(3) :128–150, 2005.

[6] L. De-Maeztu, A. Villanueva, and R. Cabeza. Near real-time stereo matching using geodesic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2) :410–416, 2012.

[7] J.-P. Delhomme. *Applications de la théorie des variables régionalisées dans les sciences de l'eau*. PhD thesis, Université Pierre et Marie Curie, 1976.

[8] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine vision and applications*, 6(1) :35–49, 1993.

[9] A. Goshtasby. Similarity and dissimilarity measures. In *Image Registration, Advances in Computer Vision and Pattern Recognition*, pages 7–66. Springer London, 2012.

[10] M. J. Hannah. *Computer Matching of Areas in Stereo Images*. PhD thesis, Stanford, CA, USA, 1974.

[11] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2) :328–341, 2008.

[12] R. Horaud and O. Monga. *Vision par ordinateur : outils fondamentaux*, chapter Vision stéréoscopique. Hermes, 2nd edition, 1995.

[13] S. Huq, A. Koschan, and M. Abidi. Occlusion filling in stereo : Theory and experiments. *Computer Vision and Image Understanding*, 117(6) :688–704, 2013.

[14] P. Jaccard. Bulletin de la société vaudoise des sciences naturelles. Technical report, 1901.

[15] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–103. IEEE, 2001.

[16] Y. Ohta and T. Kanade. Stereo by intra-and inter-scanline search using dynamic programming. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2) :139–154, 1985.

[17] S. Prince. *Computer vision : models, learning, and inference*, chapter Models for grids. Cambridge University Press, 2012.

[18] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3) :7–42, 2002.

[19] K. Yamaguchi, T. Hazan, D. McAllester, and R. Urtasun. Continuous markov random fields for robust stereo estimation. In *Computer Vision–ECCV 2012*, pages 45–58. Springer, 2012.

[20] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(3) :492–504, 2009.

[21] M. F. Zanguera Tous. *Segmentation interactive d'images fixes et de séquences vidéo basée sur des hiérarchies de partitions*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 2001.

[22] C. L. Zitnick and S. B. Kang. Stereo for image-based rendering using image over-segmentation. *International Journal of Computer Vision*, 75(1) :49–65, 2007.