

Approche monodimensionnelle de la mise en correspondance stéréoscopique par corrélation

Sébastien LEFEBVRE¹, Sébastien AMBELLOUIS¹, François CABESTAING²

¹INRETS-LEOST

20, rue Elisée Reclus, 59666 Villeneuve d'Ascq, France

²LAGIS - UMR CNRS 8146. Université des Sciences et Technologies de Lille,
Cité scientifique, 59655 Villeneuve d'Ascq Cedex

sebastien.lefebvre@inrets.fr, sebastien.ambellouis@inrets.fr, fcab@ieee.org

Résumé – Dans ce papier, nous présentons une méthode originale de mise en correspondance stéréoscopique par corrélation de voisinages monodimensionnels. Pour chaque pixel de l'image de référence, un ensemble de courbes de corrélation est calculé pour des voisinages de taille variable et centrés ou non sur le pixel à appairer. Un ensemble de caractéristiques, extrait des courbes de corrélation, est analysé afin de déterminer, pour un couple d'images stéréoscopiques, une carte des disparités associée à une carte de confiance. La densité de la carte des disparités est fixée grâce à un post-traitement de la carte de confiance. Les résultats présentés montrent clairement l'intérêt de notre méthode notamment pour une application liée à la détection d'obstacles routiers.

Abstract – In this paper, we present an original correlation-based stereo matching approach which uses the information of brightness provided by line-shaped neighbourhoods. For each pixel of the reference image, a set of correlation curves is computed by comparing 1D neighbourhoods centred or not on the pixel to be paired and whose sizes vary. Features, extracted from the whole set of correlation curves, are analysed to yield a disparity and a confidence map. The density of the disparity map depends on a post-treatment step of the confidence map. The results presented in this paper show clearly the interest of our method especially for an obstacle detection application.

1 Introduction

La stéréovision permet de déterminer l'information de profondeur d'une scène à partir de deux images de cette scène (images gauche et droite) acquises au même instant mais depuis des points de vue légèrement différents. L'étape de mise en correspondance permet la détermination de paires de pixels homologues (un dans chaque image), lesquels sont les projections d'un même point de la scène réelle. Dans le cas d'images rectifiées, deux pixels homologues se situent sur la même ligne dans les images gauche et droite. Leurs coordonnées dans leur image respective ne diffèrent que par leurs abscisses. La différence entre ces deux abscisses est appelée « disparité ».

L'ensemble des techniques de mise en correspondance stéréoscopique par corrélation décrites dans la littérature est fondé sur l'utilisation de fenêtres de corrélation bidimensionnelles décalées le long des lignes épipolaires (les lignes de l'image dans le cas de couples d'images rectifiées). Ces méthodes font l'hypothèse que tous les pixels contenus dans la fenêtre ont la même disparité. Elles permettent de calculer une valeur de disparité dans des zones de l'image peu texturées en intégrant l'information contenue dans les lignes voisines. Toutefois, cette hypothèse est invalidée lorsque les images sont caractérisées par une forte perspective ou par des objets très proches situés à des profondeurs différentes. Afin de pallier cette difficulté, nous proposons une approche basée sur l'analyse de courbes de

corrélation obtenues à partir de fenêtres monodimensionnelles (1D). Nous préconisons de ne pas calculer systématiquement une disparité dans des zones de l'image où l'information peut s'avérer insuffisante mais plutôt d'affecter une valeur de confiance à chaque appariement possible. Cette méthode calcule ainsi, en chaque pixel de l'image de référence, une confiance pour chaque décalage de l'intervalle des disparités : dans les zones faiblement (resp. fortement) texturées, l'information conduit à une confiance faible (resp. élevée). Une étape de décision permet finalement d'obtenir une carte des disparités plus ou moins dense selon le niveau de confiance fixée.

Cette nouvelle approche est détaillée dans la deuxième section en abordant les trois points suivants : (1) les mesures de similarité 1D multi-configurations, (2) la modélisation de la confiance en chaque valeur de corrélation et (3) la prise de décision. Avant de conclure, la troisième section est consacrée à la présentation des résultats obtenus dans le cadre de l'évaluation proposée par Scharstein et Szeliski dans [1] et de ceux obtenus sur un couple d'images issues d'une application de détection d'obstacles routiers.

2 L'approche monodimensionnelle

2.1 Mesures de similarité

Des travaux [2, 3] ont montré l'apport de l'utilisation de fenêtres de corrélation de taille variable et pour lesquelles

le pixel à apparier est placé en différentes positions. Ces travaux déterminent la fenêtre optimale afin d'effectuer le meilleur appariement possible. Dans nos travaux, nous reprenons conjointement ces deux techniques en calculant, pour un même pixel, des valeurs de corrélation issues de plusieurs configurations de fenêtres 1D (taille et position variables).

Soit Ω un voisinage 1D de largeur $w_n = 2n + 1$ défini pour le pixel (x, y) dans l'image de référence, où n est un entier tel que $n \in [1, n_{max}]$. Soit p_i la position du pixel courant dans la fenêtre Ω avec $i \in [-n, n]$. Une fenêtre identique est placée dans l'autre image au pixel (x, y) et décalée d'une valeur entière s telle que $s \in [s_{min}, s_{max}]$.

En chaque pixel (x, y) , on obtient le volume de valeurs de corrélation présenté par la figure 1. Chaque cube de ce volume représente une valeur de corrélation calculée avec une fenêtre 1D de largeur w_n , décalée d'une valeur s et pour laquelle le pixel à apparier est en position p_i dans la fenêtre. De cette façon, chaque ligne de cubes suivant l'axe des s représente une courbe de corrélation pour une configuration de fenêtre donnée. L'équation $C(x, y, p, w_n, s)$ d'une telle courbe est donnée par :

$$C(x, y, p, w_n, s) = \sum_{i=-p+n}^{i=-p-n} |I_l(x+i, y) - I_r(x+i+s, y)|$$

avec

$$\begin{cases} w_n = 2n + 1 \\ n \in [0, n_{max}] \\ p \in [-n, n] \\ s \in [s_{min}, s_{max}] \end{cases}, \quad (1)$$

où I_g et I_d représentent respectivement les niveaux de gris des pixels dans l'image gauche et droite.

Dans la section suivante, nous détaillons le calcul de la confiance associée à chaque valeur de corrélation.

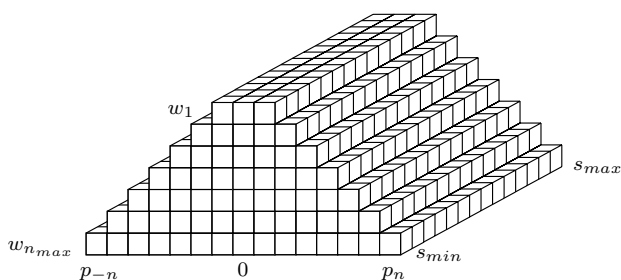


FIG. 1 – Volume des données de corrélation pour un pixel, toutes configurations de la fenêtre de corrélation confondues

2.2 Modélisation de la confiance

De nombreux critères peuvent être utilisés afin d'évaluer la confiance apportée à un appariement [4]. Nous proposons d'utiliser conjointement plusieurs d'entre eux :

- Le *nombre de points d'inflexion* c'est-à-dire le nombre de minima locaux. Nous affectons une confiance élevée à une courbe de corrélation ayant un seul pic.

- La *courbure* mesurée en chaque décalage de la courbe de corrélation. La confiance affectée à chaque décalage augmente avec la courbure.
- Le *rang* de chacune des valeurs de corrélation rangées par ordre croissant. La confiance augmente lorsque le rang diminue.

La modélisation et la combinaison de ces trois critères sont générées à l'aide de la logique floue. Chaque critère est représenté par une variable floue définie sur trois sous-ensembles flous (*mauvais*, *moyen*, *bon*), chacun caractérisé par une fonction d'appartenance de forme classique (triangulaire ou trapézoïdale). Un mécanisme d'inférence classique permet de mettre en œuvre 27 règles définies grâce à l'analyse du comportement des trois critères sur des voisinages typiques issus de plusieurs couples d'images stéréoscopiques. La sortie de chaque règle d'inférence est définie sur 5 sous-ensembles (*nul*, *mauvais*, *moyen*, *bon*, *excellent*). La confiance finale est obtenue en calculant leur centre de gravité. Dans la suite de l'article, cette confiance est notée $Conf_{FL}(p, w_n, s)$. Les variables x et y sont volontairement omises afin d'alléger les notations.

Ce processus est réitéré pour chaque pixel de l'image de référence. Un volume identique à celui de la figure 1 est ainsi obtenu pour les valeurs de confiance. L'analyse de ce volume permet alors de déterminer la carte finale des disparités et la carte de confiance associée. La section suivante présente la méthode d'analyse que nous avons mise en place.

2.3 Décision

La phase de décision doit garantir simultanément une densité élevée de la carte des disparités et un faible taux d'erreurs dans les zones de discontinuités et dans les zones non texturées. Afin d'y parvenir, le volume des valeurs de confiance est analysé selon les plans perpendiculaires à l'axe des w et dans l'ordre décroissant des tailles de fenêtres de corrélation. La procédure s'arrête lorsqu'une disparité est affectée à chaque pixel de l'image de référence ou lorsque tous les plans ont été analysés.

La stratégie de décision en chaque pixel (x, y) est résumée comme suit :

```

d(x, y) = 0
Do
  s*_{w_n}(p) = arg[max_s {Conf_{FL}(p, w_n, s)}]
  N_{w_n}(s) = Count_p[s = s*_{w_n}(p)]
  d*_{w_n} = arg[max_s (N_{w_n}(s))]
  If N_{w_n}(d*_{w_n}) ≥ T_{dec} × (2n + 1)
    d(x, y) = d*_{w_n}
  Else n - 1
While (n > 0) and (d(x, y) = 0)

```

Dans le synopsis précédent, $s*_{w_n}(p)$ est défini en chaque valeur de position du pixel courant dans le voisinage de taille w_n . Cette variable représente le décalage pour lequel la confiance est maximale. $N_{w_n}(s)$ est le nombre d'occurrences du décalage s lorsque la position du pixel courant varie dans la fenêtre de taille w_n .

Finalement la disparité $d*_{w_n}$ est définie comme la valeur de décalage qui obtient le maximum d'occurrences quelle

que soit la position du pixel courant dans le voisinage. Cette valeur de disparité est affectée au pixel (x, y) si la condition suivante est remplie :

$$\frac{N_{w_n}(d_{w_n}^*)}{2n+1} \geq T_{dec}, \quad (2)$$

où T_{dec} est un paramètre qui permet de fixer le rapport entre la densité de la carte des disparités et le taux d'erreurs. Si $T_{dec} = 1$, la relation (2) indique que la même valeur de disparité a été retenue pour toutes les positions du pixel courant dans la fenêtre de corrélation. Dans ce cas, la carte des disparités est dite quasi-dense et la confiance associée à chaque disparité est élevée. La valeur de confiance finale $Conf(x, y)$ associée à la disparité retenue $d(x, y)$ est la moyenne des valeurs de confiance calculée sur le sous-ensemble des valeurs de $Conf_{FL}(p, w_n, s)$ telles que $Conf_{FL}(p, w_n, s)$ est maximale en $d(x, y)$.

La section suivante présente les résultats obtenus par notre approche 1D. Notons que nous avons appliqué en fin d'algorithme la méthode classique de « Left-Right Consistency » qui permet la suppression de faux appariements [4].

3 Résultats

Dans cette partie, nous présentons tout d'abord les résultats obtenus sur l'évaluation proposée dans [1]. Ensuite, nous exposons les résultats de la méthode sur une paire d'images liée au contexte des transports.

3.1 Evaluation de Middlebury

L'évaluation proposée dans [1] a été largement utilisée dans un grand nombre de publications. Ici, nous utilisons l'image *Head and Lamp* de l'université de *Tsukuba*. Nous avons obtenu ces résultats avec des tailles de fenêtres de corrélation variant entre 3×1 et 21×1 . La figure 2 présente l'image gauche, la vérité terrain et les cartes des disparités et de confiance obtenues sur la paire d'images *Tsukuba*.

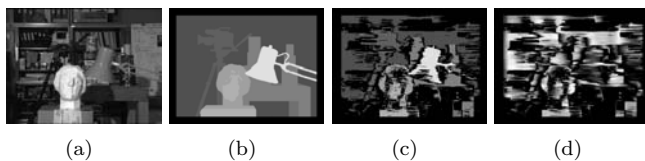


FIG. 2 – Image gauche *Tsukuba* (a), Vérité-terrain (b), Carte des disparités (c), Carte des confiances (d)

De manière qualitative, la carte des disparités obtenues semble assez proche de la vérité terrain avec un grand nombre de pixels affectés. Nous allons comparer de manière quantitative nos résultats à ceux obtenus par d'autres travaux relatés dans [1].

Nous avons comparé notre méthode à six méthodes « similaires », toutes basées sur la méthode locale de recherche exhaustive appelée « Winner-Take-All » (WTA). De telles méthodes sont utilisables pour nos systèmes de vision embarqués contrairement aux algorithmes utilisant des méthodes globales. Ces dernières sont basées sur une recherche itérative nécessitant des temps de calcul trop impor-

tants, mais surtout dont les points de convergence associés ne sont pas prévisibles.

Le tableau 1 présente les résultats obtenus sur *Tsukuba* pour différentes classes de pixels : les pixels non occultés $B_{\overline{0}}$, les pixels situés en région non-texturée $B_{\overline{T}}$, en zone de discontinuités $B_{\mathcal{D}}$ et en région texturée $B_{\mathcal{T}}$. Les taux d'erreurs sont calculés en effectuant la différence pixel à pixel entre la carte des disparités calculées et la vérité terrain. Si cette différence est supérieure à un seuil donné $|\Delta\varepsilon|$, la disparité est considérée comme erronée. Ici la condition est $|\Delta\varepsilon| > 0.5$.

TAB. 1 – Résultats obtenus à partir de l'évaluation de Middlebury avec $|\Delta\varepsilon| > 0.5$

Méthodes	Tsukuba			
	$B_{\overline{0}}$	$B_{\overline{T}}$	$B_{\mathcal{D}}$	$B_{\mathcal{T}}$
Densité	47.83	23.48	46.57	67.43
Realtime [5]	3.55 ⁶	2.73 ⁷	4.30 ⁴	4.15 ⁵
SSD+MF [1]	5.63 ⁷	2.71 ⁶	9.74 ⁷	7.77 ⁷
Improved Coop. [7]	3.12 ³	2.29 ⁴	3.78 ²	3.73 ⁴
Adapt. Weights [8]	2.41 ²	0.70 ¹	4.39 ⁵	3.65 ³
Comp. Win. [9]	3.28 ⁵	2.00 ³	4.07 ³	4.21 ⁶
MMHM colors [10]	3.14 ⁴	2.69 ⁵	5.40 ⁶	3.48 ²
Notre méthode 1D	2.12 ¹	1.17 ²	3.07 ¹	2.81 ¹

Comme le montre le tableau 1 la carte des disparités est quasi-dense. Elle a été obtenue pour $T_{dec} = 1$. La comparaison a été réalisée sur la base du même sous-ensemble de pixels affectés par notre méthode. On constate que notre méthode présente de bons résultats au regard des autres méthodes notamment pour les pixels en zones de discontinuités et texturées. L'exposant lié à chaque taux précise le classement de la méthode correspondante parmi les sept méthodes retenues. Ces résultats montrent bien que le choix d'utiliser une combinaison de plusieurs configurations de fenêtres 1D est pertinent.

Cette méthode laisse le choix à l'utilisateur de mettre en place, si nécessaire, une ultime tâche de régularisation de la carte quasi-dense en tenant compte de la carte de confiance obtenue afin d'augmenter la densité de la carte des disparités.

3.2 Application

Dans cette section, nous présentons les résultats obtenus dans le cadre du projet RaViOLi¹ sur des couples d'images stéréoscopiques issus d'un capteur situé à l'avant d'un véhicule automobile. Ces images sont généralement caractérisées par une forte perspective. Elles ne garantissent pas la véracité de l'hypothèse selon laquelle la disparité dans la fenêtre de corrélation est constante si cette dernière est bidimensionnelle.

Une application de détection d'obstacles nécessite de détecter avec robustesse et de localiser avec précision les objets situés sur la route. Dans [6], Labayrade *et al.* proposent d'analyser une carte des disparités à l'aide de l'algorithme de « v-disparité ». Afin de garantir une précision

¹RAдар, VIsion Orientable, LIdar : projet financé par l'Union Européenne, le gouvernement français et la région Nord-Pas de Calais sous le numéro de contrat OBJ2-2005/3-4.1-253-7820.

suffisante et un faible taux d'erreurs, l'algorithme est appliqué à des cartes éparées. Afin d'améliorer la qualité des résultats, et tout particulièrement l'estimation du modèle géométrique de la route, il est nécessaire d'augmenter la densité des cartes tout en conservant un faible taux d'erreurs et une précision élevée de la carte.

Le tableau 2 présente les résultats obtenus pour une paire d'images routières (image 95 de la séquence présentée figure 3) introduite dans [11] par Wannas van der Mark *et al.* et sur laquelle nous avons appliqué notre méthode.

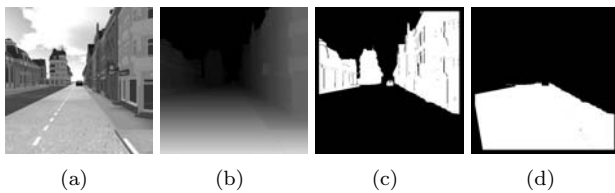


FIG. 3 – Image gauche 95 (a), Vérité-terrain (b), Pixels de fond (c), Pixels de la route (d)

Afin de mesurer uniquement l'influence de la perspective sur les performances, seuls les pixels appartenant à la route interviennent dans les calculs d'erreurs. Notre méthode est comparée à une méthode classique basée sur l'utilisation de fenêtres 2D et introduite dans [5].

Comme le montre la figure 3, à densité égale de carte des disparités, les disparités calculées pour les pixels de chaussée sont plus précises avec notre méthode 1D qu'avec une méthode 2D. Cette différence est particulièrement marquée lorsqu'on souhaite une précision sur la disparité inférieure à 0.5. Ceci est directement lié à la forme de la fenêtre 2D qui introduit un phénomène de moyennage en considérant systématiquement et sans précaution les pixels des lignes voisines. Il est important de noter que les paramètres de réglage de l'algorithme (i.e. les fonctions d'appartenance) sont identiques à ceux utilisés lors de la phase d'évaluation de Middlebury. Nous soulignons ainsi la bonne adaptabilité de notre méthode.

TAB. 2 – Comparaison des taux d'erreurs pour les pixels caractérisant la chaussée

Méthode	$ \Delta\varepsilon > 1$	$ \Delta\varepsilon > 0.5$	Densité
Hirschmüller (9 × 7)	0.27 %	16.96 %	44.35 %
Notre méthode 1D	0.12 %	0.84 %	44.35 %

4 Conclusion

Dans ce papier, nous avons présenté une méthode originale de mise en correspondance stéréoscopique par corrélation basée sur l'utilisation de fenêtres 1D multi-configurations. La méthode calcule une carte des disparités associée à une carte de confiance. Notre méthode permet d'obtenir une carte des disparités quasi-dense tout en maintenant une précision élevée et un faible taux d'erreurs. Ceci

montre clairement l'intérêt d'utiliser des fenêtres de corrélation 1D et plus particulièrement dans le contexte des scènes routières.

Nous travaillons actuellement sur une stratégie plus élaborée d'analyse des volumes de corrélation et de confiance afin d'augmenter la densité de la carte des disparités et de conserver un faible taux d'erreurs d'appariement.

Références

- [1] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. of Computer Vision*, 47(1) :7–42, April 2002. The evaluation page URL : www.middlebury.edu/stereo.
- [2] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window : theory and experiment. *IEEE Trans. on PAMI*, 16(9) :920–932, September 1994.
- [3] A. Fusiello, V. Roberto, and E. Trucco. Symmetric stereo with multiple windowing. *Int. J. of Pattern Recognition and Artificial Intelligence*, 14(8) :1053–1066, 2000.
- [4] G. Egnal, M. Mintz, and P. Wildes. A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and vision computing*, 22(12) :943–957, October 2004.
- [5] H. Hirschmüller. Improvements in real-time correlation-based stereo vision. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, Kauai, HI, USA, December 2001.
- [6] R. Labayrade, D. Aubert, and J. Tarel, “Real time obstacle detection in stereovision on non flat road geometry through V-disparity representation,” in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Versailles, France, June 2002, pp. 646–651.
- [7] H. Mayer. Analysis of means to improve cooperative disparity estimation. In *ISPRS Conf. on Photogrammetric Image Analysis*, pages 25–31, September 2003.
- [8] K.J. Yoon and I.S. Kweon. Locally adaptive support-weight approach for visual correspondence search. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 924–931, San Diego, CA, USA, June 2005.
- [9] O. Veksler. Stereo matching by compact windows via minimum ratio cycle. In *Int. Conf. on Computer Vision*, pages 540–547, Vancouver, Canada, July 2001.
- [10] K. Mühlmann, D. Maier, J. Hesser, and R. Männer. Calculating dense disparity maps from color stereo images, an efficient implementation. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 30–36, Kauai, HI, USA, June 2001.
- [11] W. van der Mark and D.M. Gavrila. Real-time dense stereo for intelligent vehicles. *IEEE Trans. on Intelligent Transportation Systems*, 7(1) :38–50, March 2006.