

**Une méthode rapide de recouvrement de la profondeur
à partir d'images stéréoscopiques d'inspiration biologique**

F. Maingreud¹, E. Pissaloux¹, C. Leroux² et A. Micaelli²

¹ Laboratoire de Robotique de Paris
Université Paris VI
CNRS FRE 2507
18 route du panorama
BP 61
F-92265 Fontenay-aux-Roses, France

² CEA, LIST
18 route du panorama
BP 61
F-92265 Fontenay-aux-Roses, France

maingreud@robot.jussieu.fr

Une méthode rapide de recouvrement de la profondeur à partir d'images stéréoscopiques d'inspiration biologique

Flavien Maingreud^{1,2}, Edwige Pissaloux¹, Christophe Leroux² et Alain Micaelli²

¹ Laboratoire de Robotique de Paris, Université Paris 6, CNRS FRE 2507

18 Route du Panorama, BP 61, 92265 Fontenay aux Roses, France

² CEA, LIST

18 Route du Panorama, BP 61, 92265 Fontenay aux Roses, France

maingreud@robot.jussieu.fr

Résumé – Cette communication propose une méthode, rapide et inspirée de la biologie, de recouvrement de la profondeur à partir des images stéréoscopiques. Deux principes sont mis en œuvre afin d'améliorer la phase d'appariement d'images : l'activation effective d'un neurone (génération d'un potentiel d'action) et le codage convenable d'un voisinage du neurone considéré (du pixel). Ces principes permettent de réduire le temps de calculs et augmentent leur robustesse.

La méthode proposée a été testée avec succès sur les stéréogrammes aléatoires de Julesz (taux d'erreur inférieur à 0.01%).

Tout système autonome tel que les robots, les déambulateurs, les systèmes électroniques d'aide à la mobilité (AEM) (pour les personnes âgées, déficients visuels, aveugles, ...) peuvent intégrer la méthode proposée à leur dispositif de détection d'obstacles.

Un nouvel dispositif d'AEM, appelé « Lunettes intelligentes » IG, d'aide au déplacement, indépendants et en sécurité, de déficients visuels dans les environnements non-coopérants, intégrera cette méthode. Ce dispositif, en conception et réalisation au LRP et au CEA, est également brièvement présenté.

Mots clés : détection de profondeur, perception humaine, fonctionnement d'un neurone, transformée d'ordre, stéréogrammes, appariement d'images, AEM (aides électroniques à la mobilité), Lunettes Intelligentes.

1. INTRODUCTION

Les résultats récents de la physiologie de la perception [Ber03] montrent que l'intégration par les différents sens des invariants de l'environnement nous permet de nous mouvoir. La procédure de préparation au déplacement peut se décomposer en deux sous-tâches :

- i. Construction d'une représentation cohérente et convenable de la scène,
- ii. Analyse de données sensorielles par tests/comparaisons de solutions possibles pour la tâche de déplacement, et sélection de la solution la plus pertinente pour la situation actuelle (en fonction de but et de la connaissance sur les obstacles ...).

Un tel modèle de procédure suppose que le système de perception fournit une information de retour pertinente pour la sous-tâche d'exécution du mouvement.

Dans le cas du système « Lunettes intelligentes » (IGS), système non-invasif, en cours de conception et de prototypage au LRP et au CEA, pour la bonne réalisation

de la première tâche, une représentation tactile de l'environnement 3D est fournie à l'utilisateur du système [Mai03].

Par conséquence, le recouvrement de la profondeur est une opération fondamentale pour les systèmes d'aide au déplacement. Ce papier propose une nouvelle méthode, inspirée par la biologie, pour y arriver.

Le papier est organisé comme suit : le Paragraphe 2 présente brièvement le système de Lunettes intelligentes et explique comment l'environnement 3D peut être représenté via le sens toucher. Le Paragraphe 3 fait un état de l'art sur les différents aspects de la perception et de l'extraction de la profondeur. Le Paragraphe 4 explique le principe de la méthode de recouvrement de profondeur proposée ; tandis que le Paragraphe 5 analyse les résultats de son évaluation sur des images de synthèse (les stéréogrammes de Julesz). Le Paragraphe 6 conclut le papier et propose les directions pour les recherches futures.

2. SYSTEME « LUNETTES INTELLIGENTES » (IGS)

Le système Lunettes intelligentes est une aide électronique au déplacement non-invasive. Sa conception vise à confirmer (et non à substituer) les données extraites par l'utilisateur final, à l'aide des autres sens, en implantant efficacement l'information pertinente au déplacement sur l'environnement 3D. En effet, la perception spécifique de l'environnement 3D par l'utilisateur final doit être intégrée à tous les AEM (Aides Electronique à la Mobilité), car elle constitue la condition nécessaire de son appropriation finale.

La figure 1 résume le principe de la perception de l'environnement via IGS, un système portable de petites dimensions physiques. Une scène observée (acquise avec 2 caméras stéréo) est traduite vers une carte d'obstacles (et un ensemble des zones sans obstacles), cette carte est affichée sur une surface à stimulation tactile qui est balayée avec une main.

Afin d'effectuer la traduction, l'IGS comprend 3 sous-systèmes :

- . un sous-système d'acquisition d'images : un système passif stéréovision ;
- . un analyseur de scènes : un sous-système de construction de la carte de profondeurs ;
- . une surface à stimulation tactile : une surface de 64x64 taxels (Tactile Element) [Vel 03].

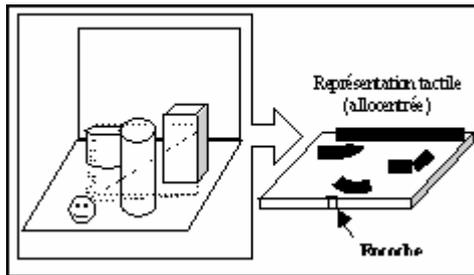


Fig. 1. Principe fonctionnel du système « lunettes intelligentes »

Le sous-système de vision acquiert les images et construit la carte d'obstacles : une carte de différentes parties de l'environnement classées en « zones sans obstacles » et « zones d'obstacles », avec les obstacles localisés au sol ou suspendus, ou encore localisés sous la surface de référence.

La carte d'obstacles peut être construite comme suit :

- . acquisition de deux images stéréo ;
- . construction de la carte de disparités et extraction de la profondeur (cf. §3) ;
- . simplification de la carte (ajustement de distances relatives et inclusion de l'information d'orientation) ;
- . classification des différentes zones de l'espace 3D ;
- . transmission de la carte de la représentation vers la surface à stimulation tactile.

Ainsi donc, une méthode simple, rapide et précise de calculs de profondeur est au centre du l'IGS.

3. PERCEPTION/EXTRACTION DE L'INFORMATION DE PROFONDEUR

L'homme, comme certains mammifères (les chats, les souris, les chiens, ...) reconstruit la profondeur (information 3D) en partie grâce à sa vision binoculaire (information 2D) ; aussi, notre système tentera de proposer une solution similaire (de part de ses principes).

A. Perception de la profondeur

La vision binoculaire utilise différents indices pour calculer l'information de profondeur : l'occlusion, la taille relative d'objets, la perspective, l'ombre, le flou, la vitesse relative, Cependant, la différence dans le contenu entre des images « gauche » et « droite » constitue sa principale différence avec la vision monoculaire (Figure-2) ; c'est la disparité binoculaire appelée aussi la stéréopsis [DeA00].

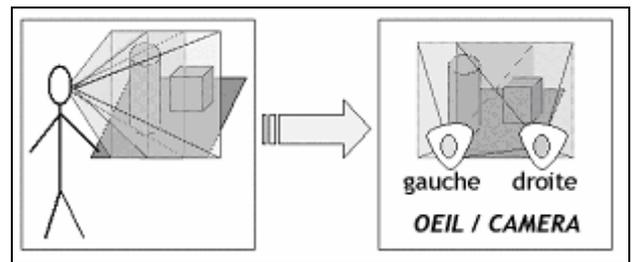


Fig. 2. Disparité (différence) obtenue par la projection d'une même scène dans deux yeux

En effet, l'information de profondeur est, en partie, directement corrélée avec les différences (phases, occlusion, ...) perçues par la même scène projetée dans les deux yeux.

Depuis le premier stéréoscope (Weatstone, 1838) jusqu'au les années 1960, la théorie de perception visuelle a considéré la fusion stéréoscopique comme un processus de perception visuelle de haut niveau (processus cognitif). Ce processus peut être résumé comme suit : tous les objets de la scène doivent être identifiés et reconnus dans chaque projection (dans chaque œil) ; l'étude des différentes positions de ces objets aboutit à la création d'une carte mentale de disparité [Man 98]. En 1960, Julesz, grâce à ses stéréogrammes (RDS – random dot stereograms) apporte la preuve que la fusion stéréoscopique est un processus de bas niveau et il permet de créer la perception de profondeur d'une scène sans la reconnaissance des objets la composant. Des expérimentations de Julesz on a induit que les connexions de bas-niveau entre les neurones permettent de détecter des zones identiques dans 2 projections de la même scène. Ces connexions renseignent sur l'information de disparité en réagissant positivement à la présence dans une image d'un stimuli lumineux jugé identique à celui de l'autre image.

B. Stéréopsis et géométrie épipolaire

La stéréopsis signifie qu'il existe deux représentations de la même scène acquises à partir de deux positions différentes. Du point de vue de la géométrie, il est possible de reconstruire/calculer la profondeur de chaque point d'une scène 3D à partir de couples de points homologues de deux images stéréoscopiques.

Si la position et l'orientation de deux récepteurs sont inconnues, un point d'une image peut être homologue de tout point de l'autre image (la recherche de points homologues se fait donc en deux dimensions). La connaissance de la transformation qui lie les deux récepteurs réduit le problème à une seule dimension : les points homologues se trouvent sur les lignes épipolaires obtenues par les intersections de deux plans caméras avec le plan épipolaire (le plan définit par les deux centres de projection et le point de la scène 3D considéré) ; c'est la stéréovision épipolaire.

Faugeras et al. [Fau 93] ont proposé une méthode pour réduire la stéréovision généralisée à sa forme simplifiée : la stéréovision épipolaire.

C. Extraction de la profondeur

L'extraction de la profondeur est souvent réalisée de deux façons : à partir de caractéristiques bien identifiées (les méthodes éparses) ou à l'aide de la technique d'appariement de blocs (les méthodes denses).

i. Extraction de la profondeur à partir de caractéristiques

Cette approche utilise différents extracteurs de caractéristiques. Parmi les caractéristiques considérées on retrouve : les points de contours, les contours, les segments de droites, ... [Har 88], [Tom 91], [Smi 95]. Les positions dans les deux images d'une même caractéristique permettent de retrouver la disparité.

Cette approche conduit à une carte de disparités éparses. Une carte de disparités dense peut être obtenue par extrapolation en utilisant les différentes propriétés de l'environnement choisies comme la continuité, la rigidité du corps, ... [Gri 85], [Can 97], [Sha 98].

ii. Technique d'appariement de blocs

Cette approche vise à construire une carte de disparités dense. Aussi, la caractéristique choisie doit avoir une forte occurrence dans l'image ; le niveau de gris ([Bir 96], [DiS 01]), le niveau de gris moyen dans un voisinage [Kan 95], le voisinage encodé [Dic 97], [Man 98], en sont quelques exemples. La programmation dynamique, la corrélation, l'appariement hiérarchique, ... sont les algorithmes le plus souvent utilisés pour la mise en correspondance des images [Zha 93], [Que 96], [Pis99], [Pis 00], [Sch 01], [Bro03]... La complexité temporelle et les erreurs de l'appariement peuvent être réduites grâce aux contraintes liées aux données-sources [Mar 82], [Pis 00bis], [Sun02].

Pour représenter le plus fidèlement possible une scène 3D, une carte de disparités dense devrait être établie.

4. APPROCHE INSPIREE DE LA BIOLOGIE POUR LE RECOUVREMENT DE LA PROFONDEUR

Le processus de codage neuronal de la luminosité est à la base de notre approche du recouvrement de la profondeur.

A. Quelques éléments de base sur la perception neuronale

La perception visuelle de l'homme se base sur la perception de luminosité de photorécepteurs de la rétine, les cônes et les bâtonnets [Tho 01], et la relation de voisinage spatiale créée par les liaisons inter-neurones de bas niveau.

i. Fonctionnement d'un photorécepteur

Un photorécepteur intègre l'information lumineuse reçue de l'environnement jusqu'à un niveau considéré comme pertinent pour être transmis vers le cerveau via le nerf optique. A ce moment on dit que le neurone a généré un potentiel d'action (Figure 3).

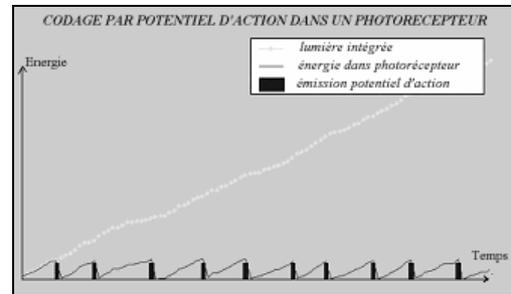


Fig. 3. Génération d'un potentiel d'action par un neurone.

ii. Fonctionnement d'un neurone et codage de l'information visuelle

Depuis le premier enregistrement de l'activité d'un neurone (1920), on pensait que c'était la fréquence de décharge d'un neurone qui encodait l'information sur l'environnement. Or il semble ([Tho 01], [Ger 02]), que l'ordre dans lequel les potentiels d'action de neurones voisins sont émis intervienne aussi dans ce codage.

Plus le voisinage considéré est grand plus l'information fournie est pertinente.

B. Principe de fonctionnement de la transformée d'ordre

On considère des images statiques à niveau de gris. Le niveau de gris est le résultat de l'intégration de la lumière par les photorécepteurs : plus la quantité d'énergie intégrée est importante plus le pixel est clair. Ainsi, le temps et la façon d'illuminer un voisinage du photorécepteur peuvent être simulés en classant les pixels par l'ordre de leur niveau de gris dans ce voisinage.

D'où le principe de notre algorithme afin d'établir le code (« étiquette ») d'un voisinage attaché au pixel central, appelé la transformée d'ordre :

Un photorécepteur/pixel blanc a intégré plus d'énergie, et a probablement généré un potentiel d'action plus rapidement que les autres photorécepteurs/pixels de son voisinage (Figure 4);

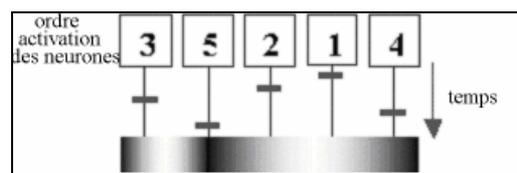


Fig. 4. Principe d'activation de neurones voisins, pour une distribution d'énergie lumineuse.

Par analogie, on fait la même analyse pour tous les pixels du voisinage considéré, on les ordonne dans l'ordre décroissant de leur activation, puis on remplace chaque pixel de l'image source par son numéro d'ordre dans le voisinage (Figure 5).

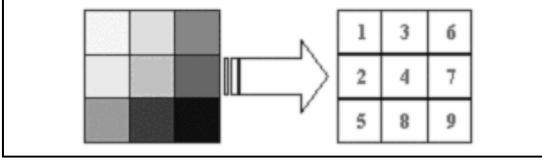


Fig. 5. Ordre d'activation dans un voisinage 3*3.

Il est possible de formaliser l'algorithme comme suit :

Soient :

T : la transformée d'ordre

$p_{i,j}$: pixel courant,

$Rank$: la transformation qui calcule l'ordre du pixel $p_{i,j}$ dans son voisinage 3x3 trié dans l'ordre décroissant du niveau de gris

$$Rank(p_{i,j}) \in [0,8] \text{ pour un voisinage de } 3*3$$

L'étiquette du pixel résultant, $T(p_{i,j})$, dans ce voisinage est définie par la séquence de nombres :

$$T(p_{i,j}) = (Rank(p_{i+k,j+l}))_{(k,l)} \quad (1)$$

avec $(k,l) \in \{(m,n)/(m,n \in \{-1,0,1\}) - (0,0)\}$

C. Recouvrement de la profondeur

Une fois les deux images stéréoscopiques épipolaires étiquetées, un algorithme d'appariement classique (la corrélation dans notre cas) suffit pour retrouver les points homologues dans les deux images et calculer les disparités ; en effet, les points homologues doivent avoir les mêmes étiquettes.

La disparité entre les deux images correspond à la différence entre les positions spatiales de pixels homologues. Cette observation peut être formalisée comme suit :

$R(i,j)$: la disparité du pixel (i,j)

$T(p_{i,j}^K)$: l'étiquette associée au pixel (i,j) par la transformée d'ordre de l'image considérée K

$$K = \begin{cases} G, image_gauche \\ D, image_droite \end{cases}$$

$Disp$: la disparité maximale autorisée.

$$\delta(x,y) = \begin{cases} 0, si x \neq y \\ 1, si x = y \end{cases}$$

$$R(i,j) = \sum_{k=0}^{Disp} (k * \delta(T(p_{i,j}^G), T(p_{i,j+k}^D))) \quad (2)$$

La fonction δ permet de sélectionner les pixels homologues de deux images ayant la même étiquette; pour simplifier, on suppose qu'il existe au maximum un couple de points homologues par ligne considérée dans nos images stéréo épipolaires.

5. EVALUATION DE LA METHODE PROPOSEE SUR DES IMAGES

La méthode proposée a été évaluée sur des stéréogrammes de Julesz (RDS). La figure 6 montre les résultats obtenus. Les RDS sont générés en intégrant un modèle de disparité (déphasage) à une image aléatoire. Après la transformation d'ordre, une corrélation est appliquée pour retrouver la carte de disparité.

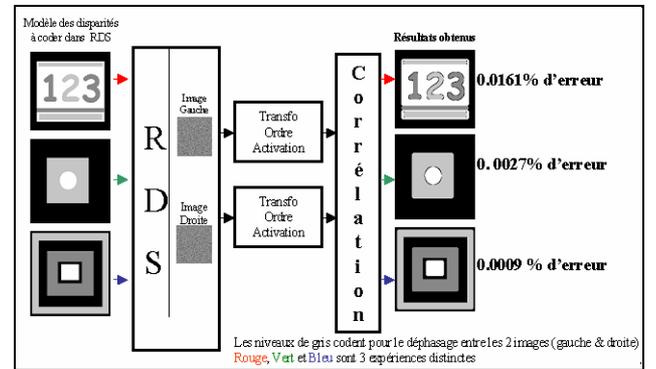


Fig. 6. Résultats de l'appariement de stéréogrammes de Julesz.

Sur les images testées, la carte de disparité obtenue par notre méthode affiche un très faible taux d'erreurs : 0,01% au maximum.

6. CONCLUSION

Ce papier propose une nouvelle méthode, d'inspiration biologique, pour le calcul de disparités entre deux images stéréo épipolaires. Le principe de cette méthode consiste à utiliser, pour l'appariement, non les images brutes mais des images de l'ordre d'activation de pixels d'un voisinage. Cette méthode a prouvé son efficacité sur les stéréogrammes de Julesz. Pour la généraliser à des images réelles, une inclusion des indices visuels forts, comme l'orientation, devra être faite au niveau de l'algorithme. Une implantation en parallèle devrait aussi être envisagée afin d'améliorer les performances temporelles.

7. REFERENCES

- [1] [Ber03] A.Berthoz « Le sens du mouvement », 1998, « La décision », 2003, Odile Jacob.
- [2] [Bro03] M.Z.Brown, D.Burschka, and G.D.Hager. « Advances in Computational Stereo ». IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8):993-1008, 2003.
- [3] [Bir96] S.Birchfield, C.Tomasi « Depth discontinuities by pixel to pixel stereo » Proc. 6th Intl Conf. Computer Vision, pp.1073-1080, Jan.1998.
- [4] [Can97] F.Candocia, M.Adjouadi « A similarity measure for stereo feature matching » IEEE Transactions On Image Processing, Vol.6, N°10, October 1997.
- [5] [DeA00] G C.DeAngelis « Seeing in three dimensions :the neurophysiology of stereopsis » Trends in cognitive Sciences, Vol.4, N°3, March 2000.
- [6] [Dic97] A.Dick « Multiresolution stereo image matching using wavelets » Report for Honours computer science, University of adelaide, November 1997.
- [7] [DiS01] L. Di Stefano, M. Marchionni « Fast Dense Stereo Based on the Uniqueness Constraint: Theory, Implementation and Results » Technical Report CSITE-10-01, Oct. 2001.
- [9] [Fau93] O. Faugeras, B.Hotz « Real time correlation-based stereo: algorithm, implementations and applications » rapport INRIA N°2013, août 1993.
- [10] [Ger02] W.Gerstner and W.M.Kistler « Spiking Neuron Models - Single Neurons, Populations, Plasticity » Cambridge University Press ,August 2002.
- [11] [Gri85] W. E. L. Grimson, “Computational experiments with a feature-based stereo algorithm,” IEEE Trans. Pattern Anal. Mach. Intell., vol.7, pp. 17-34, 1985.
- [12] [Har88] C.Harris, M.Stephens « A combined corner and edge detector », In Alvey Vision Conference, pp.147-151, 1998.
- [13] [Her01] J.Herault « De la rétine biologique aux circuits neuromorphiques » in « Les systèmes de vision » pp.55-96, Hermès Science, 2001.
- [14] [Jul60] B.Julesz, « Binocular Depth Perception of computer-generated patterns », Bell Systems Techn.J 39, pp.1125-1162
- [15] [Kan95] T.Kanade, H.Kano « Development of a video-rate stereo machine », Proceedings of Intl Robotics and Systems Conference, Pittsburgh PA, August 5-9, 1995.
- [16] [Mai03] F.Maingreud, E.Pissaloux, « Intelligent Glasses: a Vision System for Blind people displacements in 3D environment », Proc. Int. Conf. on Visual and Hearing Impairment, 2002 Granada, Espagne, pp. 121-130.
- [17] [Man98] J.Mansson, « Stereovision: a model of human stereopsis », Lund University Cognitive Studies, Internal report, 1998.
- [18] [Mar82] D.Marr « Vision » W.H.Freeman and Company.
- [19] [Pis 00] Pissaloux, E., E., Le Coat, F., Tissot, A., Durbin, F., *An adaptive parallel system dedicated to projective image matching*, IEEE ICIP2000 (Int. Conf. on Image Processing), Vancouver, Canada, 10—13 Septembre, 2000.
- [20] [Pis 00bis]Pissaloux, E., E., Le Coat, F., *A³C and its Application to Spatial Optimization of a Parallel Architecture dedicated to Image Matching*, 3rd. Int. Conf. on Computer Vision, Pattern Recognition & Image Processing, Atlantic City, USA, 27 février—3 mars, 2000, pp. 184—187
- [21] [Pis 99] Pissaloux, E., Le Coat, F., Bonnin, P., Tissot, A., Durbin, F., *Design and optimisation of a Parallel Heterogenous System for Aerial Image Matching*, IEEE Int. Conf. on Image Processing (ICIP'99), Kobe, Japon, Octobre 25—28, 1999, IEEE Cat. N° 99CH36348C, ISBN : 0-7803-5470-2
- [22] [Que96] G.M.Quénot « Image matching using dynamic programming: application to stereovision and image interpolation » Image Communication, Bordeaux, France, 20-22 Mai 1996.
- [23] [Sch01] D. Scharstein and R. Szeliski. « A taxonomy and evaluation of dense two-frame stereo correspondence algorithms ». Microsoft Research Technical Report MSR-TR-2001-81, 2001.
- [24] [Sha98] S. D. Sharghi, “Feature point-base stereo correspondence from stereo images using geometric relational matching,” Ph.D dissertation, Dep. Computer Science and Eng. Univ. Texas at Arlington, TX 1998.
- [25] [Smi95] S.M.Smith, J.M.Brady « SUSAN- A new approach to low level image processing », Int. Journal of Computer Vision, 23(1) pp.45-78, May 1997.
- [26] [Spi03] <http://www.spikenet-technology.com>
- [27] [Sun02] C.sun « Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques », Intl. Journal of Computer Vision, Vol.47, N°1/2/3, pp.99-117, May 2002.
- [28] [Tho01] S.Thorpe « Le système visuel humain » in « Les systèmes de vision » pp.23-54, Hermès Science, 2001.
- [29] [Tom91] C.Tomasi, T.Kanade, « Detection and tracking of point features », Technical report CMU-CS-91-132, Carnegie Mellon University, 1991.
- [30] [Vel03] R. Velázquez, F. Maingreud et E.Pissaloux « Conception d'un Système Electronique d'Aide à la Mobilité des Non-voyants Basé sur des Principes Visio-Tactiles » 17 èmes Journées des Jeunes Chercheurs en Robotique, Versailles, France, avril 2003 pp 102-108
- [32] [Zha93] Z.Zhang « Le problème de la mise en correspondance: L'Etat de l'art » rapport INRIA N°2146, décembre 1993.