

Reconstruction 3D temps réel dans un VSIP

Tarik GRABA¹, Bertrand GRANADO¹, Olivier ROMAIN¹, Thomas EA², Andrea PINNA¹, Patrick GARDA¹

¹Laboratoire des Instruments et Systèmes d'Ile de France (LISIF),
Université Pierre et Marie Curie (UPMC)
4 Place Jussieu, BC 252, 75252 Paris Cedex 05, France

²Institut Supérieur d'Electronique de Paris (ISEP)
28 Rue Notre Dame des Champs, 75006 Paris, France

tarik.graba@lisif.jussieu.fr, bertrand.granado@lisif.jussieu.fr,
olivier.romain@lisif.jussieu.fr
thomas.ea@isep.fr, andrea.pinna@lisif.jussieu.fr, garda@lisif.jussieu.fr

Résumé – Dans cet article sont décrits un système intégré pour la reconstruction du relief, les contraintes et les choix de conception, la structure du capteur de vision active et les méthodes et architectures utilisées pour obtenir une représentation 3D en temps réel.

Abstract – In this paper are described a integrated 3D system, the constraints and conception choices, the vision sensor structure and the algorithms and architectures used to have a realtime 3D reconstruction

1 Introduction

Dans de nombreux domaines, l'information apportée par l'imagerie classique est insuffisante. Par exemple, avec l'endoscopie, l'image apporte des information sur la couleur et sur la forme des lésions possibles mais très peu sur leurs dimensions ou leurs rugosités, obligeant dans de nombreux cas à pratiquer des biopsies. Pour faciliter les diagnostics et limiter le recours à des méthodes plus invasives, une information supplémentaire sur le relief apporterait au praticien des données quantitatives, lui permettant ainsi d'affiner ses diagnostics sans avoir recours à la chirurgie.

Pour obtenir une information du relief d'un objet ou d'une scène de nombreuses solutions ont été envisagées. Les recherches menées dans ce domaine depuis de nombreuses années ont mis en avant des techniques permettant de restituer le relief qui est perdu avec la prise d'image classique [1, 2, 3].

L'utilisation des système de vision est la moins contraignante et la plus commune. Ces systèmes sont basés sur l'utilisation de capteurs d'images pour obtenir l'information nécessaire à la restitution du relief. Ces caméras sont utilisées seules dans le cas des capteurs passifs ou associées à une source d'éclairage supplémentaire pour faciliter la reconstruction du relief dans le cas des capteurs actifs.

Les algorithmes de restitution du relief ont beaucoup évolué et permettent maintenant de retrouver de manière réaliste le relief d'une scène ou d'un objet. Ces algorithmes assez complexes font appel à des méthodes de calcul lourdes qui, en général, ne permettent pas d'avoir un résultat en temps réel, surtout sur des architectures où les ressources de calcul sont réduites et la consommation énergétique limitée telles celles destinées à l'embarqué.

Ces dernières années de nouveau axes de recherches sont apparus dans la vision intégrée. Des systèmes ont été conçus, intégrant sur une même puce le capteur d'images et un sys-

tème de traitement de l'information [4]. Ces VSOC (Vision System On Chip) permettent d'effectuer des traitements complexes (tels que la reconnaissance de forme, la classification, la compression vidéo...) tout en offrant un encombrement réduit et une consommation énergétique faible. Ces rétines intégrées peuvent ainsi répondre aux exigences des systèmes embarqués.

La miniaturisation et l'embarquabilité permettent d'envisager de nouvelles applications où les capteurs sont de très petite taille, autonomes et peuvent ainsi être dispersés en grand nombre dans des régions inaccessibles autrement, communiquer entre eux en des réseaux de capteurs et permettre la fusion de données. Le déploiement de capteurs de ce type, qui réalisent localement une reconstruction 3D puis communiquent leurs résultats, permettrait d'avoir rapidement une vue globale, sans zones d'ombres, de l'environnement.

Dans ce contexte, Nous proposons une nouvelle approche où la reconstruction 3D est envisagée non pas comme un problème isolé dans une chaîne de traitement mais dans le cadre plus global d'un système conçu dans sa totalité pour réaliser cette tâche de la manière la plus efficace possible. Ce système nous l'avons appelé "Cyclope", un système de vision en relief intégré basé sur le principe de stéréovision active. Il regroupe tous les organes nécessaires pour fonctionner en temps réel, c'est un système de type VSIP (Visual System In Package).

2 CYCLOPE

Cyclope est notre système de vision en relief intégré [5]. Nous avons opté pour un système de vision stéréoscopique actif. Les algorithmes de traitement pour la reconstruction 3D, dans les systèmes actifs, permettent de satisfaire les contraintes de temps réel et de précision dans la reconstruction [12]. De plus cela permet de limiter les ressources ce qui est indispensable pour un système embarqué si l'on faible consomma-

tion.

Cyclope est composé de trois parties principales :

- un bloc d'instrumentation : pour l'acquisition de l'image et la génération du motif structuré composé :
 - d'une source de lumière cohérente (LASER) couplée à une optique de diffraction. La lumière LASER étant produite par une diode à cavité verticale VCSEL (Vertical Cavity Surface Emitting Laser) émettant dans le proche infra-rouge.
- Les principaux avantages de cette structure par rapport aux diodes lasers conventionnelles, résident dans un faible encombrement permettant une forte intégration, une faible consommation et une émission par la face.
- d'une matrice de photodétecteurs de type CMOS. Elle permet l'acquisition d'une image de texture dans le visible et une image du motif projeté dans le proche infrarouge.
- un bloc de traitement et de contrôle composé d'une partie figée (microprocesseur) et d'une partie reconfigurable (FPGA).

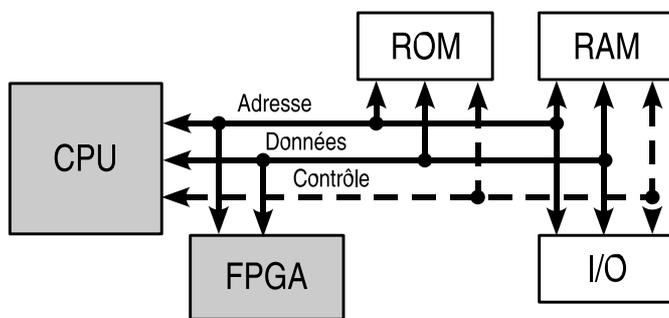


FIG. 1 – Schéma structural du bloc numérique

- un bloc de radio communication : pour la transmission des résultats et la reconfiguration dynamique du bloc numérique.

Ces trois parties sont réalisées grâce à des technologies différentes qui ne sont pas forcément compatibles entre elles, et l'intégration de celles-ci dans un circuit monopuce est de prime abord ardue. Le développement d'un boîtier intégré de type SIP-AMS (System In Package – Analog Mixed Signal) est actuellement la meilleure solution pour faire sauter les verrous technologiques qui se présentent tout en offrant des dimensions proches de celles d'un système monopuce. Cette intégration permet d'avoir des caractéristiques stables pour le couple caméra/projecteur et d'effectuer les traitements sur le circuit.

3 Vision active

Les systèmes de vision stéréoscopique actifs peuvent être vus comme des systèmes de stéréovision classiques où une des deux caméras est remplacée par une source d'éclairage projetant un motif structuré [6, 7, 8, 9, 10]. La position des points caractéristiques du motif dans l'image permet de calculer les profondeurs par triangulation. L'information contenue dans la structure du motif permet de lever une partie de l'ambiguïté d'appariement.

Le principe de restitution de la dimension manquante reste le même qu'en vision passive [11], sauf que nous avons une

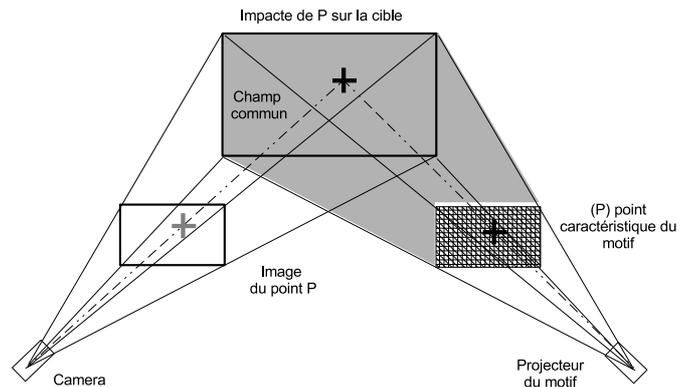


FIG. 2 – Principe de la stéréovision active

connaissance a priori du motif projeté et donc des primitives à rechercher dans l'image. Cette connaissance a priori permet de réduire grandement la complexité du problème réduisant ainsi les temps de calcul nécessaires pour effectuer les traitements, et permettant de répondre à des contraintes temporelles fortes [12].

4 Le motif structuré

Le choix du motif projeté est très important [13, 14]. Celui-ci doit apporter suffisamment d'information pour éviter les problèmes d'appariement tout en étant assez simple à générer de manière intégrée. L'option choisie est d'utiliser une diode LASER de type VCSEL couplée à un réseau de diffraction, nous sommes limités à un motif monochrome de géométrie simple et régulière. L'utilisation d'un faisceau laser monomode dans le proche infrarouge permet d'être plus robuste aux variations d'albédo et de colorimétrie. Nous avons opté pour un maillage régulier de points, ce maillage a l'avantage d'être simple à générer. L'imageur utilisé [15] est multi-spectral, il nous permet d'extraire facilement les points de ce motif dans l'image infrarouge indépendamment de la scène et de sa texture.

5 Modélisation du couple stéréoscopique

Les systèmes de stéréovision classiques sont composés en général, de deux caméras regardant le même objet avec deux angles différents. Calibrer le système revient à trouver la relation entre les coordonnées d'un point de l'espace réel (repère monde) et ses projections dans les deux images prises par les caméras. Ceci est réalisé en trouvant le modèle de chacune des caméras. Dans notre système le problème se présente différemment, en effet, pour faciliter l'appariement nous avons opté pour un système stéréoscopique actif où l'une des caméras est remplacée par un projecteur de motif structuré, projetant une image fixe. Chaque point de la matrice projetée peut être considéré comme un point d'une image que nous connaissons a priori. La projection de ce point dans la seconde image (celle de la vraie caméra) appartient forcément à une droite, c'est la contrainte épipolaire [11]. La connaissance de ces droites permet de faciliter grandement le problème d'appariement.

En se basant sur ce principe et en tenant compte des travaux [16, 17] nous avons décidé de caractériser le couple stéréo-

scopique (projecteur-caméra) non pas par un système d'équations décrivant toutes les transformations d'un repère à l'autre mais seulement par les équations de ces droites épipolaires. Les équations de ces droites sont obtenues après une étape de calibration analogue à celle décrite dans [18]. Plusieurs prises

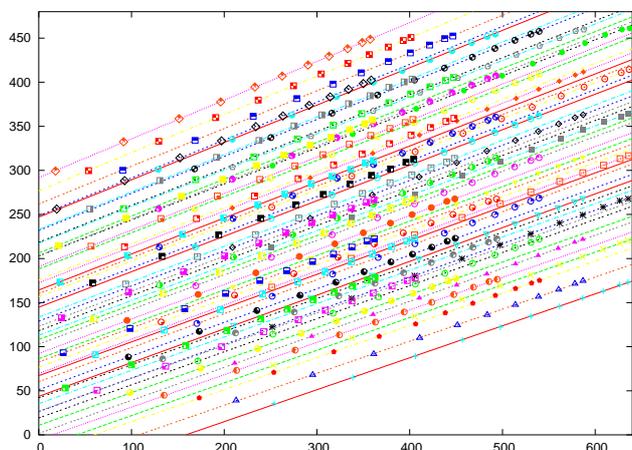


FIG. 3 – Lignes épipolaires

d'images ont été faite à différentes distances. Pour chaque image les coordonnées des impacts laser sont extraites. La trajectoire de chaque point en fonction de la profondeur donne une droite. Les paramètres de chaque droite sont obtenus par une simple approximation aux moindres carrés. La figure 3 montre les droites obtenues à partir de onze images d'un motif laser composé de 7×7 points.

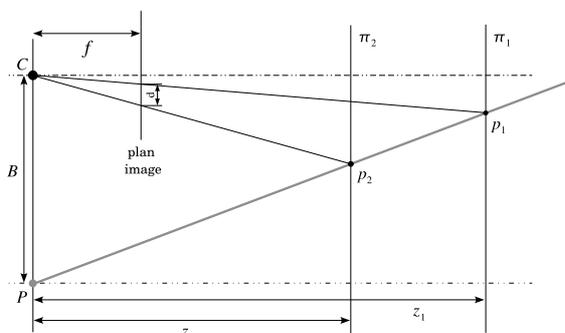


FIG. 4 – Déplacement d'un point en fonction de la distance

En plus du modèle des lignes épipolaires, nous avons dû modéliser la position de l'image des points du motif en fonction de la distance du couple stéréoscopique à l'impact de ces points dans la scène.

Si on considère un rayon laser projeté sur deux plans consécutifs se situant respectivement aux distances z_1 et z_2 (figure 4) alors le déplacement de l'image de ces impacts sur la ligne épipolaire est de la forme :

$$d = Bf \frac{(z_1 - z_2)}{z_1 z_2} \quad (1)$$

Où B est la base du stéréoscope et f la distance focale de la caméra.

A partir de là on montre que l'on peut lier la distance à laquelle se trouve un impact LASER à l'une des coordonnées image (i, j) si l'on connaît ce point et la ligne épipolaire à laquelle il appartient, par un modèle de la forme :

$$z = \frac{1}{\alpha i + \beta} \quad (2)$$

La même démarche expérimentale que celle utilisée pour la détermination des droites épipolaires nous a permis d'obtenir un modèle par rayon laser projeté. La figure 5 montre les courbes donnant z en fonction de i obtenues après approximation aux moindres carrés des paramètres du modèle.

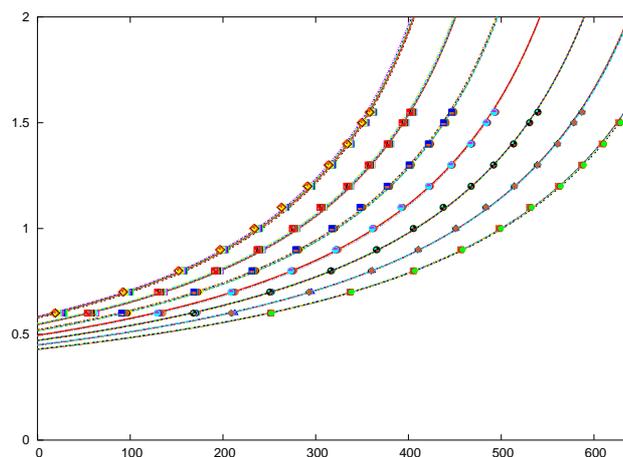


FIG. 5 – Distance en fonction de d'abscisse dans l'image

6 Architecture de reconstitution

Nous proposons (figure 6) une architecture massivement parallèle, qui à partir des coordonnées des points laser, trouve la droite épipolaire la plus proche. Cette dernière conditionne le choix du modèle de profondeur utilisé pour le calcul de la troisième dimension z .

Nous avons estimé le nombre d'opérations nécessaires pour effectuer ces opérations pour un motif de dimension N et grâce à une architecture de traitement séquentielle (un micro processeur standard).

Pour une image et un motif de N points projeté, $5N^2 + 2N$ opérations (en considérant les opérations d'addition, de multiplication, de division et de comparaison comme équivalentes) sont nécessaires.

Ce qui donne pour un motif de quatre cents points (ce qui correspond à un impact laser pour un carré de 8×8 pixel pour une image QCIF) à une cadence de 25 images/s une puissance de calcul nécessaire de 20 Mops par seconde.

7 Conclusion

Nous avons présenté dans cet article "Cyclope" un système temps réel, intégré et communicant pour la vision en relief. Nous avons présenté aussi une architecture pour la reconstitution du relief et nous avons estimé le nombre d'opérations

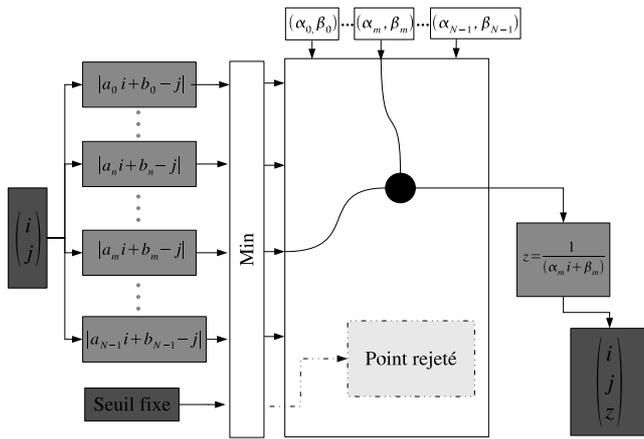


FIG. 6 – Architecture de traitement

nécessaire à la restitution du relief en temps réel et montré que ceci était compatible avec les contraintes d'exécution temps réel dans un système embarqué.

Références

[1] K. Ng & al. "An integrated multi-sensory system for photo-realistic 3d scene reconstruction", in Proceedings ISPRS Comm. V, pp. 356- 363, Hakodate, Japan, 1998

[2] V. Sequeira & al. "3D environment modeling using laser range sensing", in Robotics and Autonomous Systems, vol. 16, pp. 81-91, 1995

[3] O. Romain & al. "A multi spectral sensor dedicated to 3D spherical reconstruction", in Proceedings of IEEE ICIP 2001, Oct. 2001

[4] A. Pinna & al. "Connectionnist Retina : a neural network system integrated into an electronic retina", in the proceeding of the 25 th international congress on high speed photography and photonics, pp 218-223, 2002

[5] T. Graba & al. "Cyclope : An Integrated Real-time 3D Image Sensor", in Proc. DCIS 2004

[6] E. Cortizo & al. "Application of the structured illumination method to study the topography of the sole of the foot during a walk", in Optics and Laser in Engineering, Vol. 40

[7] F. Foster & al. "The hiscore camera – Areal time three dimensional color camera", in proceedings of IEEE ICIP 2001, Oct. 2001

[8] C. Guan & al. "Composite structured light pattern for three-dimensional video", in Optic express, Vol. 11, N. 5, pp. 406-417, 2003

[9] D. Scharstein & al. "High-accuracy stereo depth maps using structured light", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 195-202, 2003

[10] T. Ea & al. "Spherical stereoscopic sensor for 3D color imaging", in Proc. SPIE Vol. 4297, pp. 91-97, jun 2001

[11] J. Battle & al. "Recent progress in coded structured light as a technique to solve the correspondence problem : a

survey", in Pattern Recognition, Vol. 31, No. 7, pp. 963-982, 1998

[12] S. Boverie & al "Comparison of structured light and stereovision sensors for new airbag generation", in Control Engineering Practice, vol 11, pp. 1413-1421, 2003

[13] J. Salvi & al. "A robust-coded pattern projection for dynamic 3D scene measurement", in Pattern Recognition Letters, Vol. 19, pp. 1055-1065, 1998

[14] J. Salvi & al. Pattern "codification strategies in structured light systems", in Pattern Recognition, 2004

[15] A. Pinna & al. "Design of an integrated silicon connectionist retina", in the proceedings of the fifteenth annual IEEE international ASIC/SOC conference, pp 133-137, 2002

[16] A. Dipanda & al. "3-Dshape reconstruction in an active stereo vision system using genetic algorithms", in Pattern Recognition, vol. 36 pp. 2143-2159, 2003

[17] S. Woo & al. "Determination of an Optimal Configuration for a Direct Correspondence in an Active Stereovision System", in Proc. on 2nd IASTED INT. Conf. VIIP, Malaga, Espagne, pp. 596-601, September 2002

[18] F. Marzani & al. "Calibration of a three-dimensional reconstruction system using a structured light source", in Optical Engineering, Vol. 41 No. 2, pp. 484-492, February 2002