

Vers une reconstruction dense par zoom

Dense Reconstruction by Zooming

par Catherine DELHERM, Jean-Marc LAVEST, Michel DHOME, Nadine DAUCHER

Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA)
URA 1793 du CNRS, Université Blaise Pascal de Clermont-Ferrand
63177 Aubière cedex, France
delherm@lacos.iut.univ-lehavre.fr

résumé et mots clés

L'utilisation d'un objectif à focale variable permet d'inférer des informations tridimensionnelles mais les angles de triangulation mis en jeu sont alors très faibles. Pour obtenir des résultats de reconstruction corrects, le calibrage du système de prise de vues, la mise en correspondance des primitives ainsi que la phase de reconstruction doivent être aussi précis que possible. Cet article présente une nouvelle approche vers une reconstruction volumique dense à l'aide d'une caméra statique équipée d'un zoom. Notre algorithme se déroule en trois étapes :

- Tout d'abord, à partir d'une corrélation multi-échelle nous déterminons pour chaque point d'intérêt de la première image ses correspondants dans les images suivantes de la séquence avec une précision égale au pixel.
- La deuxième étape consiste, à partir des données fournies par la première étape, à affiner la recherche des points appariés par une méthode permettant d'obtenir une précision au sous-pixel.
- A partir des appariements, un algorithme de modélisation permet de reconstruire les points 3D. Cet algorithme ne nécessite aucun modèle explicite de calibrage et met en œuvre plusieurs images de grille régulières et précises formant un espace métrique particulier. Une expérimentation sur des données réelles est présentée et montre qu'il est possible de reconstruire des objets 3D à partir d'une séquence d'images obtenue à l'aide d'une caméra équipée d'un zoom.

Corrélation surfacique, reconstruction dense, stéréovision axiale, calibrage implicite.

abstract and key words

As it has been previously demonstrated, axial stereovision technics allows us to infer 3D information, but involves very small triangulation angles. Accurate calibration, data matching and reconstruction have to be performed to obtain satisfactory modelling results. In this paper, a new approach is proposed to realize dense reconstruction using a static camera with a zoom lens.

The proposed algorithm described in the following sections is divided in three major steps :

- First of all, the matching problem is solved using a correlation algorithm that explicitly takes into account the zooming effect through the images set. An intensity-based multiscale algorithm is applied to the feature points in the first image, in order to obtain unique point correspondences in all the other images.
- Then, using pixels matched by the previous method, an iterative process is proposed to obtain a sub-pixel accuracy.
- Finally, the 3D surface is reconstructed using image point correspondences. The modelling algorithm does not require any explicit calibration model and the involved computations are straightforward. As an implicit calibration process, this approach uses several images of accurate regular grids placed on a micrometric table. Experiments on real data are provided and show that it is possible to compute 3D dense information from a zooming image set.

Correlation, Dense Reconstruction, Axial Stereovision, Implicit Calibration.

1. introduction

Un zoom est en général utilisé pour obtenir des détails d'une scène observée par une caméra. Mais récemment [LRD93], [RA94] ont prouvé expériences à l'appui, qu'il était également possible

d'inférer des informations 3D à partir d'images prises à l'aide d'une caméra munie d'un zoom.

Dans les travaux précédemment cités, l'étape de calibrage est primordiale et doit être aussi précise que possible. Dans cet article, le problème de reconstruction est résolu par une technique de calibrage basée sur l'utilisation de grilles régulières et précises placées

sur une table micrométrique. Cette technique a déjà été validée par les résultats obtenus par [LPDD94], [DLPD95], [Del95]. Les auteurs ont ainsi montré qu'il est possible de reconstruire un objet 3D même si les données sont proches de l'axe optique. Les résultats présentés faisaient état de la reconstruction d'une grille placée à 1 mètre de la caméra avec moins de 1 mm d'erreur sur le paramètre de profondeur (le plus sensible). Un des avantages de cette méthode est de prendre en compte les phénomènes de distorsion (optiques et électroniques). Toutefois, dans les reconstructions de grille présentées, la mise en correspondance des points était gérée manuellement en utilisant une détection au sous-pixel des croix de la grille. L'article que nous proposons s'inscrit dans la suite logique de ces travaux. Nous invitons cependant les lecteurs désireux d'informations quant à la sensibilité et à la précision du processus de reconstruction à s'y référer.

Dans cet article, nous proposons une méthode de mise en correspondance de points par corrélation surfacique dans les images de luminance zoomées. A partir d'un point quelconque sélectionné dans la première image de la séquence, nous sommes capables de déterminer directement les points en correspondance dans les images suivantes de la séquence par analyse d'une fenêtre de corrélation. La méthode présentée ici se déroule en deux étapes. La première est une corrélation classique multi-échelles [Zha93]; elle nous fournit des points d'appariement avec une précision au pixel. Mais cette précision est insuffisante pour reconstruire correctement les points 3D. La seconde étape consiste à rechercher les appariements avec une précision inférieure au pixel. Cette étape fondée sur un processus d'optimisation itératif utilise les coordonnées déterminées précédemment comme initialisation.

Après avoir décrit en détail le processus de corrélation, nous décrivons brièvement la méthode de reconstruction utilisée. Enfin, des résultats de reconstruction dense d'objets réels seront présentés permettant ainsi la validation de la méthode.

2. état de l'art

Bien que peu répandue dans la littérature, la reconstruction par zoom présente des caractéristiques intéressantes. Le plus souvent, les études concernant de tels systèmes restent théoriques, comme dans [Car91], et ne sont pas suivies d'expériences sur des objets réels.

Les méthodes de calibrage classiques peuvent s'appliquer dans le cas d'une caméra munie d'un zoom, mais les propriétés d'un tel objectif permettent d'apporter quelques simplifications. La bibliographie existant dans ce domaine fait état d'un calibrage en deux temps [LT87], [Car91], [Lav92], [LL95] :

- détermination du point principal, ou centre de l'image,
- calcul des autres paramètres.

Willson [WS93] propose une nouvelle taxinomie des centres de l'image ainsi qu'un modèle ajustable : dans un premier temps, un calibrage classique du type de celui présenté dans [Tsa86] est effectué pour un ensemble de focales et de mises au point. Des polynômes sont ensuite ajustés aux modèles individuels en utilisant les données de calibrage. Le processus est répété jusqu'à l'estimation de tous les paramètres. Ces modèles ajustables s'avèrent très intéressants pour parvenir à un calibrage de qualité mais nécessitent un grand nombre de données de calibrage.

Dans le cadre de la vision dynamique, quelques auteurs [TTG92], [Li94b], [EVF94] ont étudié les problèmes de calibrage d'un tel objectif. [Li94a] propose un calibrage tenant compte à la fois des distorsions tangentielle et radiale selon deux modèles classiques de la photogrammétrie. Le problème est alors de déterminer un vecteur de 15 paramètres; la résolution s'effectue par une méthode de minimisation et les valeurs initiales sont données par les propriétés des points de fuite.

A notre connaissance seulement deux approches présentent des résultats tangibles de reconstruction par zoom [LRD93], [RA94]. Dans [LRD93] et [LRD95], les auteurs s'appuient sur les propriétés optiques d'un objectif de très bonne qualité. Par des exemples concrets et mesurables de reconstruction, ils ont montré la possibilité d'inférer des informations 3D par zoom. Mais seulement deux types d'objets sont reconstruits : les objets de révolution et les objets polyédriques. De plus, pour les objets polyédriques, les segments coupant l'axe optique ne sont pas reconstruits et ceux proches de l'axe optique donnent des segments 3D aberrants.

Dans [Rod93], l'auteur utilise une caméra équipée d'un zoom de qualité standard. Pour rendre le système axial, il applique des corrections, par une méthode dite de « rectification », à chaque couple de distances focales. Si la méthode paraît séduisante, elle s'avère complexe à mettre en oeuvre dès lors que l'on souhaite traiter une séquence d'images prises à de multiples distances focales. D'un autre côté, si l'on ne considère que deux images, les résultats de reconstruction seront moins précis et seuls les objets situés en bordure d'images seront modélisés de façon cohérente.

La corrélation est un outil souvent utilisé pour mettre en correspondance des primitives [Zha93]. Mais dans ces travaux, les images dont disposent les auteurs sont prises avec une caméra classique non équipée d'un zoom. A notre connaissance, seul [RA94] a utilisé une méthode de corrélation monodimensionnelle, dans un système de stéréovision axiale. Sa méthode est basée sur les propriétés des droites épipolaires dans le cas du zoom; l'étude des profils d'intensité le long de deux droites épipolaires qui se correspondent permet d'apparier des points caractéristiques (pics, vallées, points de fort gradients). A partir des points appariés une méthode de reconstruction 3D est proposée. Les résultats sont présentés sous la forme d'un maillage grossier et comportent quelques erreurs. L'erreur de reconstruction est, d'après [RA94], de moins de 1 cm pour un objet situé à environ 1 mètre de la caméra. Toutefois, il convient de noter que l'objet reconstruit est placé dans un coin de l'image; il est donc éloigné du centre de l'image autour duquel il est plus difficile de reconstruire.

3. calibrage implicite et reconstruction dans un espace métrique de grilles

3.1. Rappels

Considérons une grille réelle parfaite située devant la caméra. L'image de cette grille sur la matrice CCD est distordue (figure 1); cette distorsion peut provenir de phénomènes optiques, voire électroniques. Un calibrage classique, non linéaire, essaie de prendre en considération ces phénomènes dans un modèle mathématique de projection perspective, mais des erreurs subsistent entre le modèle mathématique manipulé et le contenu de l'image. Supposons qu'il existe une transformation \mathcal{T} , qui à tout point de la grille distordue fasse correspondre le point sur la grille de calibrage. Les phénomènes de distorsion seront ainsi automatiquement pris en compte. De cette façon, toute nouvelle image distordue peut être transformée dans la grille de référence, supposée aussi parfaite que possible.

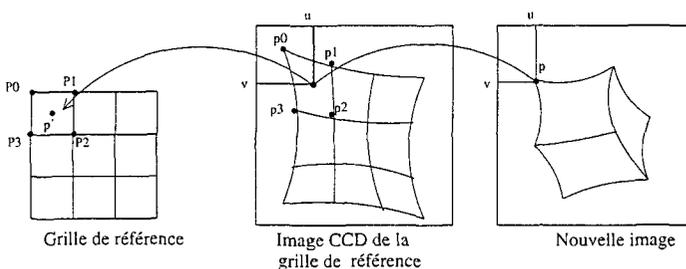


Figure 1. – Principe du recalage dans la grille de référence : approximation locale.

La technique d'étalonnage décrite dans cet article utilise deux images d'une même grille de référence pour des positions différentes (obtenues à l'aide d'une table micrométrique). Ces deux positions sont éloignées d'une distance D (voir figure 2).

Ce processus de calibrage appelé 'calibrage par bi-plans' n'est pas nouveau; [MBK81] a introduit le concept il y a quelques années

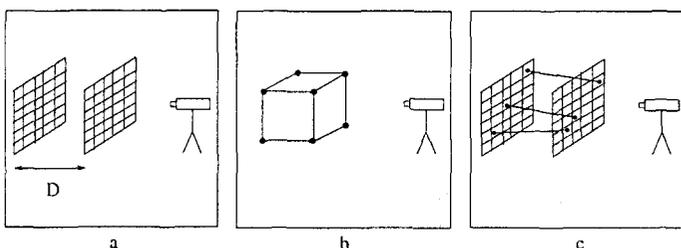


Figure 2. – Rayons optiques 3D dans l'espace métrique des grilles.

et [GTK88], [WM91], [Peu94] ont proposé depuis plusieurs méthodes de résolution.

Les différences entre les résultats présentés dans la littérature dépendent :

- des fonctions locales utilisées pour déterminer la relation entre l'image de la grille distordue et la grille de référence,
- mais également de la manière dont l'image des points de référence des grilles est détectée sur la matrice CCD.

Comme nous venons de le préciser, une des difficultés majeures est de déterminer précisément les positions des points de calibrage (p_0, p_1, p_2, p_3 de la figure 1) sur l'image de la grille distordue. [Peu94] a développé un algorithme original qui cherche directement dans l'image de luminance les paramètres d'un modèle mathématique caractérisant la réponse photonique de la caméra au voisinage d'un point de calibrage. Des travaux similaires ont été réalisés par [Ack84], [BMB94], [DG93] et [Bey92]. Ainsi, la localisation d'un point (intersection de deux droites de la grille) est obtenue avec une précision inférieure au centième de pixel. Une étude détaillée du modèle utilisé ainsi que du choix des méthodes d'interpolation dans une maille élémentaire de la grille est fournie dans [DLPD95].

3.2. interpolation par rapport anharmonique

Soit un point $p(u, v)$ (figure 1) dans une nouvelle image distordue. Connaissant ses coordonnées (u, v) , il est possible de déterminer dans l'image distordue de la grille, ses 4 points voisins (p_0, p_1, p_2, p_3). Une fonction d'interpolation par bi-rapport permet alors de passer de l'image distordue (p_0, p_1, p_2, p_3, p) à la grille de référence (P_0, P_1, P_2, P_3, p') où les points (P_0, P_1, P_2, P_3) sont respectivement les points de la grille de référence correspondant à (p_0, p_1, p_2, p_3) et le point p' le point interpolé.

Soient deux grilles parallèles dont la position est parfaitement connue (figure 2a). Une table micrométrique est utilisée durant la phase de calibrage et un contrôle mécanique permet de positionner précisément les grilles. Les schémas ci-dessus nous permettent de comprendre l'idée de base du processus de calibrage.

Considérons l'image d'un objet prise par la caméra : (2b), il est possible de déterminer pour chaque point de cette image, par une interpolation locale, ses points correspondants dans la première et la seconde grille (figure 2c). Nous pouvons ainsi estimer le rayon optique 3D du point image considéré, dans l'espace métrique défini par les grilles de référence, sans connaître les paramètres de la caméra. Bien entendu, il est possible d'utiliser plus de deux grilles pour estimer le rayon optique; les calculs seront plus précis mais le temps de calcul ainsi que le stockage en mémoire seront d'autant augmentés.

La méthode présentée ici permet donc de calibrer sur une large gamme de distances focales sans changer d'objet de calibrage en

cours de séquence. De plus, elle ne nécessite aucun traitement préalable du système et peut tout aussi bien être utilisée avec un zoom de qualité standard.

3.3. processus de reconstruction

L'algorithme de modélisation ne présente pas de difficulté. Pour chaque distance focale utilisée durant la séquence de prise de vue avec le zoom, deux grilles de calibrage sont préalablement digitalisées et les coordonnées de chaque point d'intersection du maillage sont mémorisées. Cette étape constitue le processus de calibrage. La stabilité de l'axe optique et l'ensemble des phénomènes de distorsion sont automatiquement pris en compte dès lors que le problème est transposé dans l'espace métrique des grilles.

Soient $p^i, i \in [1, n]$, les projections dans une séquence d'images d'un même point tridimensionnel P_{3d} (figure 3), (i représente l'indice de la focale utilisée pendant les prises de vue). A partir de l'interpolation dans la première puis dans la seconde grille, il est possible d'estimer les rayons 3D relatifs à chaque image et exprimés dans l'espace métrique des grilles (cf paragraphe 3.2).

Il apparaît, d'après la figure 3, que le point tridimensionnel observé sur la séquence d'images P_{3d} , se situe à l'intersection des rayons 3D définis dans l'espace métrique des grilles.

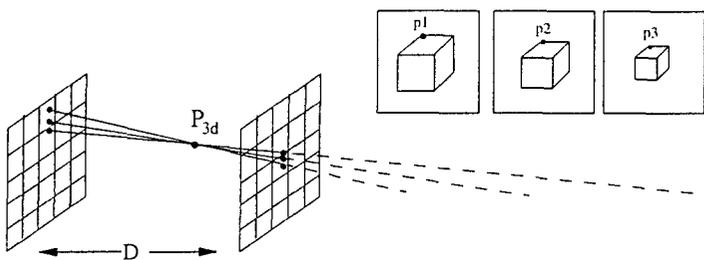


Figure 3. – Processus de modélisation.

4. appariement de points par corrélation

4.1. principe de la corrélation

La corrélation est une technique couramment utilisée en traitement du signal qui permet d'étudier la ressemblance entre deux signaux. Dans le cas de la vision par ordinateur, plusieurs formulations sont possibles [Zha93]. Nous utiliserons ici la définition de la corrélation normalisée [Law82]. Soient i et j deux indices parcourant le même domaine discret $[-n..n]$. Soient I_1 et I_2 les

fonctions définissant le niveau de gris de deux images. Le coefficient de corrélation entre les pixels p_1 et p_2 respectivement de coordonnées (u_1, v_1) et (u_2, v_2) dans la première et la seconde image considérées, défini sur une fenêtre carrée centrée de taille $(2n + 1)^2$, est donné par la formule suivante :

$$C = \frac{\sum_i \sum_j I_1(u_1+i, v_1+j) I_2(u_2+i, v_2+j)}{\sqrt{\sum_i \sum_j I_1(u_1+i, v_1+j)^2} \sqrt{\sum_i \sum_j I_2(u_2+i, v_2+j)^2}}$$

La corrélation permet une mesure simple de la ressemblance entre deux voisinages définis dans deux images. L'idée de base est d'associer au point $p_1(u_1, v_1)$ de la première image, le point $p_2(u_2, v_2)$ de la seconde image pour lequel le coefficient de corrélation est maximum. Selon [Mor77] et [Law82], la corrélation est sensible au bruit. Pour pallier ce problème nous utiliserons une technique de corrélation multi-résolutions proposée par [Zha93]. Une image est en fait une matrice d'intensités lumineuses. A partir de cette image, nous pouvons construire de nouvelles images par diminution progressive de la résolution. A chaque étape, l'image est filtrée et échantillonnée [BB82]. Classiquement la technique de corrélation est utilisée pour établir la mise en correspondance de points de contour dans deux images issues d'une caméra standard. Notre approche est différente dans le sens où nous utilisons la corrélation sur une séquence d'images prises avec des distances focales différentes sans déplacement physique de la caméra. De plus, notre algorithme ne s'applique pas obligatoirement sur des points de contour mais sur tout type de point issu de l'image de luminance. Comme nous l'avons précédemment mentionné, l'algorithme se déroule en deux étapes. La première que nous appellerons « corrélation simple » travaille depuis l'image la moins zoomée (champ large) jusqu'à l'image la plus zoomée. La mise en correspondance d'un même point est réalisée avec une précision au pixel. La seconde partie de l'algorithme, que nous appellerons « corrélation au sous-pixel », travaille en traitant les images en sens inverse, c'est à dire depuis l'image la plus zoomée vers l'image la moins zoomée. Cette étape est nécessaire pour obtenir des mises en correspondance les plus précises possibles afin de minimiser les erreurs de triangulation dans le processus de reconstruction axial.

4.2. corrélation simple

Ce processus cherche à appairer par corrélation les points homologues dans une séquence en partant de l'image acquise avec la distance focale la plus courte (champ de vision le plus large). Cette recherche se fait de proche en proche entre deux images successives I^k et I^{k+1} de la séquence. L'idée de base est d'associer au point $p_i^k(u_i^k, v_i^k)$ de l'image I^k , le point $p_i^{k+1}(u_i^{k+1}, v_i^{k+1})$ de l'image I^{k+1} pour lequel le coefficient de corrélation normalisé est maximum. A priori, il est nécessaire de rechercher le point p_i^{k+1} dans une zone de l'image I^{k+1} suffisamment grande pour englober le déplacement existant entre les points homologues des deux images. Pour réduire le coût algorithmique, nous utilisons

un processus de recherche pyramidal travaillant à différentes résolutions :

* Paramètres de l'algorithme

n : demi-taille de la fenêtre de corrélation

t : demi-taille de la zone de recherche dans l'image I^{k+1}

r : nombre de résolutions considérées

* Principe

1. initialisation

$$x_1 = u_i^k, y_1 = v_i^k, x_2 = x_1, y_2 = y_1, C_{max} = 0$$

2. recherche du coefficient de corrélation maximum pour la résolution $dl = 2^{r-1}$

pour dx variant de $-t$ à $+t$

pour dy variant de $-t$ à $+t$

$$C = \frac{\sum_{i=-n}^n \sum_{j=-n}^n I^k(x_1 + i \cdot dl, y_1 + j \cdot dl)}{\sqrt{\sum_{i=-n}^n \sum_{j=-n}^n I^k(x_1 + i \cdot dl, y_1 + j \cdot dl)^2}} \cdot \frac{I^{k+1}(x_2 + (i + dx) \cdot dl, y_2 + (j + dy) \cdot dl)}{\sqrt{\sum_{i=-n}^n \sum_{j=-n}^n I^{k+1}(x_2 + (i + dx) \cdot dl, y_2 + (j + dy) \cdot dl)^2}}$$

si $C > C_{max}$ alors $C_{max} = C$,

$$x_{2max} = x_2 + dx \cdot dl, y_{2max} = y_2 + dy \cdot dl$$

3. si $dl > 1$ alors

$$x_2 = x_{2max}, y_2 = y_{2max}, r = r - 1, \text{ retour à l'étape 2}$$

sinon

$$u^{k+1} = x_{2max}, v^{k+1} = y_{2max}, \text{ fin}$$

Remarque : Pour les expériences décrites dans cet article, nous utilisons les paramètres suivants : $n = 8, t = 6, r = 3$ ce qui correspond à une recherche dans une zone totale, à la résolution la plus fine, de taille 65×65 (4225 pixels) pour un calcul de seulement 507 coefficients de corrélation.

4.3. corrélation au sous-pixel

Le processus précédent, travaillant par couples d'images successives, peut subir une dérive progressive. Pour pallier cet inconvénient, nous allons travailler avec une image de référence unique pour rechercher les points homologues avec une précision sous-pixel.

Soit $\{p_i^1, \dots, p_i^l, \dots, p_i^m\}$ un ensemble de points homologues obtenus à l'étape précédente. Le point p_i^m et son voisinage vont nous servir

de référence durant tout le processus. Cette démarche semble logique puisque l'image I^m est celle qui contient le plus de détails (obtenue avec la distance focale la plus grande). Bien entendu, pour trouver deux points homologues entre les images I^m et I^l ($l \in [1..m - 1]$), il nous faut d'une part filtrer et sous-échantillonner l'image I^m et d'autre part placer au mieux la fenêtre de corrélation dans la seconde image considérée I^l . Pour cela, nous cherchons à minimiser le critère suivant à l'aide d'un processus itératif :

$$C(I^m, I^l, dl, df, dx, dy) = \sum_{i=-n}^{+n} \sum_{j=-n}^{+n} (I^m(u_i^m + dl \cdot i, v_i^m + dl \cdot j, df) - I^l(u_i^l + dx + i, v_i^l + dy + j, 1))^2$$

où les paramètres dl, df, dx, dy prennent des valeurs réelles et représentent respectivement la distance de sous-échantillonnage, la taille de la zone de filtrage, le déplacement horizontal et vertical de la fenêtre de corrélation dans la seconde image.

Dans la formule précédente, $I^k(x, y, \rho)$ représente une fonction définissant le niveau de gris filtré au point de coordonnées (x, y) dans l'image I^k . Le filtrage est réalisé sur une fenêtre de taille $(\rho \times \rho)$ centrée en (x, y) en pondérant la valeur des niveaux de gris de chaque pixel de l'image I^k par un coefficient proportionnel à la surface du pixel réellement incluse dans la fenêtre considérée.

Le minimum du critère est obtenu en recherchant la valeur des coefficients dl, df, dx, dy minimisant le critère C . Le processus d'optimisation retenu est fondé sur la technique itérative de Levenberg-Marquardt [PFTV86]. Les dérivées premières du critère sont estimées simplement par différence finie, contrairement à l'utilisation classique de la méthode précédente où une forme analytique est fournie.

Bien entendu, comme pour tout processus itératif, les conditions initiales sont importantes. La connaissance du facteur de grossissement obtenu à partir de la corrélation simple nous permet de calculer approximativement la valeur de dl, f est initialisé à la valeur de dl, dx et dy sont initialisés à zéro.

Nous disposons donc en sortie de notre processus d'une liste de points appariés entre les images de la séquence. Les coordonnées de ces points sont d'une précision inférieure au pixel. Chaque point, par une interpolation dans les grilles d'étalonnage, permet

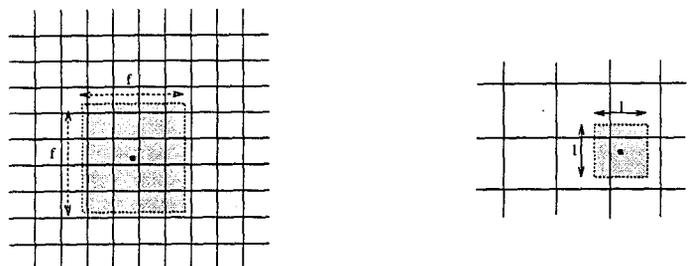


Figure 4. – Zone de calcul de $I^m(x, y, f)$ Zone de calcul de $I^l(x, y, 1)$.

de remonter jusqu'à la trace 3D du rayon optique correspondant. L'algorithme de reconstruction, détaillé dans [DLPD95], calcule le point de l'espace le plus proche de l'ensemble du faisceau de rayon optique.

5. expérimentations

5.1. conditions expérimentales

Les expériences présentées ici sont toutes réalisées à partir d'images réelles. Le zoom utilisé est de type Angénieux T14 × 9; la gamme de focales s'étend de 9 à 124 mm. Le processus se déroule en deux étapes : le calibrage puis l'appariement des points et la reconstruction 3D.

• Calibrage

Pour seize focales (de 25 à 32 mm en notation objectif), deux grilles régulières sont digitalisées pour réaliser l'espace métrique. L'ouverture et la mise au point restent constantes tout au long de la séquence. Leurs valeurs sont choisies de façon à obtenir une profondeur de champ suffisante pour l'expérience. La taille d'un maillon élémentaire de la grille est de 6.2×6.2 mm. Les deux grilles sont distantes de 15 cm. La première grille est située à environ 90 cm de la caméra. Elles sont prétraitées et chaque point d'intersection est stocké en mémoire.

• Mise en correspondance et Reconstruction

Après le traitement des grilles, l'objet à reconstruire est positionné à environ un mètre de la caméra. Cette distance correspond à l'espace défini par les deux grilles. Cependant, la localisation de l'objet entre les grilles n'est pas obligatoire. En effet, dans [DLPD95], les auteurs montrent qu'une précision satisfaisante est obtenue même si l'objet est situé en dehors de l'espace des grilles; bien entendu si l'objet s'éloigne trop de cet espace, la précision de reconstruction diminue.

Seize images de l'objet à reconstruire sont acquises pour les mêmes focales que celles utilisées lors de la phase de calibrage. Un dispositif mécanique sur le zoom nous permet de retrouver de façon précise les positions des focales.

Sur la première image de luminance de la séquence considérée, nous sélectionnons manuellement un certain nombre de points répartis sur toute la surface de l'image. Les points sont choisis au hasard et le processus de sélection pourrait tout à fait être automatisé [HS88]. L'algorithme de corrélation est ensuite appliqué sur ces points. Il nous fournit les coordonnées des points appariés sur toute la séquence d'images. Bien entendu, un certain nombre de points situés en bordure d'images vont rapidement disparaître par effet de zoom. Certains ne seront appariés que dans très peu d'images. Afin d'obtenir une bonne précision de reconstruction, seuls les points apparaissant dans cinq images au moins seront pris en considération pour la reconstruction. Ceci signifie que les coordonnées 3D des points correspondent à la meilleure intersection,

au sens des moindres carrés, d'au moins cinq rayons optiques 3D définis dans l'espace métrique des grilles de référence.

5.2. résultats de corrélation

Comme nous l'avons mentionné dans le paragraphe 4, la première partie de l'algorithme de corrélation se déroule, de proche en proche, selon un processus multi-échelles. Afin de visualiser le résultat de la procédure de corrélation, nous avons sélectionné un point (croix noire) sur la première image de la séquence qui en comporte 14 (figure 5). Sur les images suivantes sont représentés les points mis en correspondance respectivement sur les images I_6 , I_{10} et I_{14} .

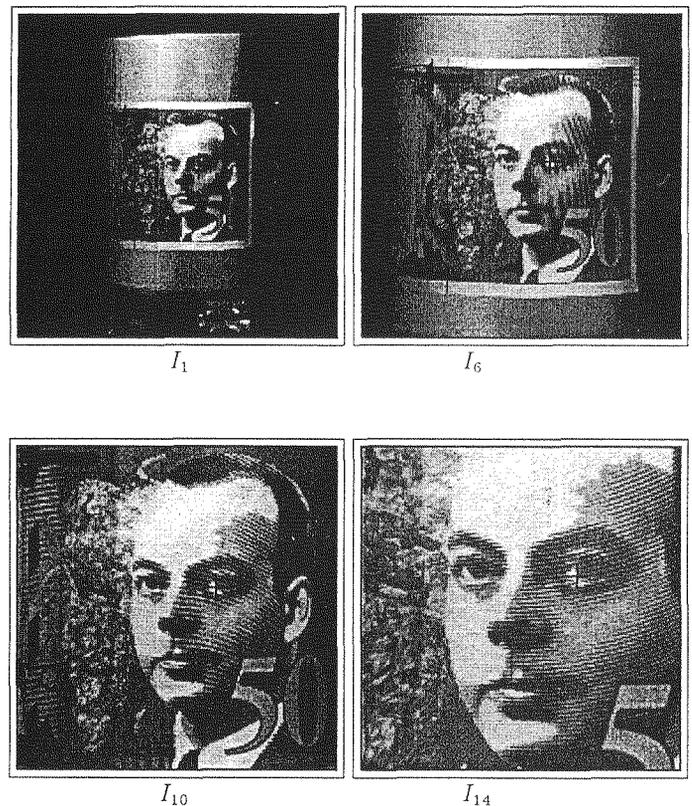


Figure 5. – Appariements par Corrélation simple (croix noire).

A partir des données ainsi obtenues (appariements au pixel), la procédure sous-pixel est appliquée; le résultat obtenu est représenté en figure 6. L'image de référence, pour le point considéré, est I_{14} .

Afin de visualiser dans de bonnes conditions les résultats, la taille du masque de corrélation est de 41×41 pixels. Nous avons représenté pour les treize corrélations $C(I_{14}, I_l)$, $l \in [1..13]$, à la solution de l'algorithme itératif :

- sur la première ligne, la position dans l'image I_l minimisant le critère $C(I_{14}, I^l, dl, df, dx, dy)$,

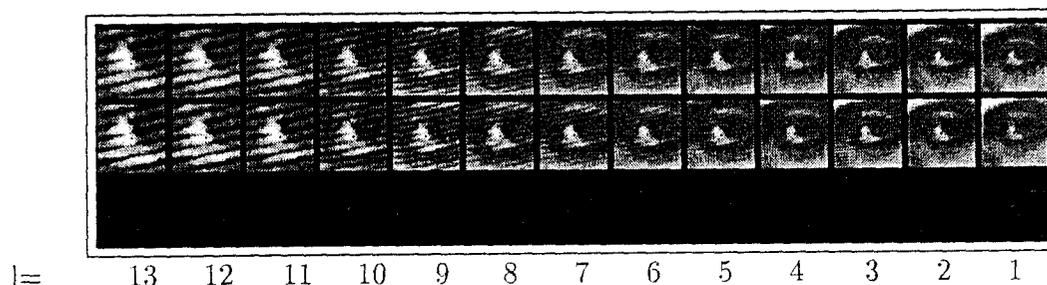


Figure 6. – Résultats de la corrélation sous-pixel.

- sur la seconde ligne, l'image de référence I_{14} sous-échantillonnée,
- sur la troisième ligne, la différence, pixel par pixel, entre les deux imagerie.

Nous pouvons noter l'effet de zoom sur la séquence, où une fenêtre de corrélation (41×41) dans l'image numéro 1 représente un champ important de l'image initiale. Notons également qu'à la solution, le sous-échantillonnage-filtrage de I^{14} permet d'obtenir une imagerie dont la fonction de luminance est très proche de I^1 .

Le point sélectionné manuellement dans la première image de la séquence a pour coordonnées (285, 240) pixels dans le repère lié à l'image. L'algorithme de corrélation simple permet de déterminer les coordonnées en pixels des points appariés dans les images suivantes de la séquence (I_2 à I_{14}). L'image de référence, pour le point considéré, est la dernière image de la séquence (I_{14}); en considérant le point correspondant de cette image de coordonnées (341, 203), l'algorithme de corrélation sous-pixel est appliqué de l'image I_{13} à l'image I_1 . Les coordonnées sous-pixels sont données dans le tableau 1 ainsi que les différents paramètres de l'algorithme : le déplacement (dx, dy), les facteurs de grossissement (dl) et de filtrage (df).

D'une façon générale, les paramètres de filtrage et de grossissement augmentent lorsque l'on se dirige vers les premières images de la séquence (à partir de l'image de référence). Ces paramètres

permettent la reconstitution précise d'imagerie (voir figure 6). Nous avons obtenu expérimentalement de meilleurs résultats en découplant les paramètres dl et df . La fonction de lissage df bien qu'initialisée à la même valeur que dl ne subit pas tout à fait la même loi de variation. Nous pouvons également noter que le déplacement (dx, dy) appliqué aux coordonnées entières peut atteindre 1.2 pixels pour dy sur les deux premières images de la séquence. La corrélation sous-pixel a donc conduit à un très fort réajustement de la corrélation simple.

5.3. représentation des objets reconstruits

Nous présentons trois exemples d'objets reconstruits : un ensemble de boîtes présentant deux faces parallèles (de taille $15 \times 10 \times 4$ cm), un cube parfaitement régulier (de 10 cm de côté) et un galet (d'envergure approximative $10 \times 8 \times 5$ cm). L'algorithme de reconstruction nous donne les coordonnées 3D des points reconstruits; or il est très difficile de se rendre compte de la forme d'un objet à partir d'un nuage de points. Selon l'objet reconstruit, nous allons donc effectuer un certain nombre d'opérations afin de pouvoir évaluer la précision de la reconstruction.

5.3.1. représentation des boîtes

L'objet est constitué de deux boîtes collées; nous obtenons ainsi deux faces grossièrement parallèles, distantes d'environ 40mm.

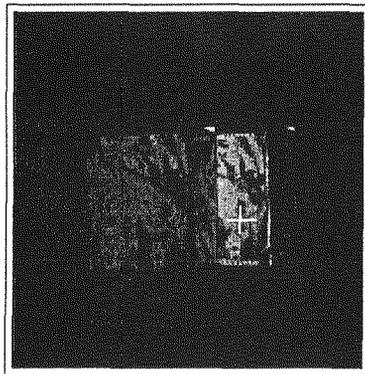
Tout d'abord, 410 points sont sélectionnés sur la face gauche et 220 sur la face droite.

La figure 7 montre un point sélectionné sur la première image de la séquence (la moins zoomée) ainsi que les points déterminés par la corrélation simple sur une image intermédiaire puis sur la dernière image de la séquence dans laquelle le point considéré est encore visible (la 13ème image constituant ici l'image de référence).

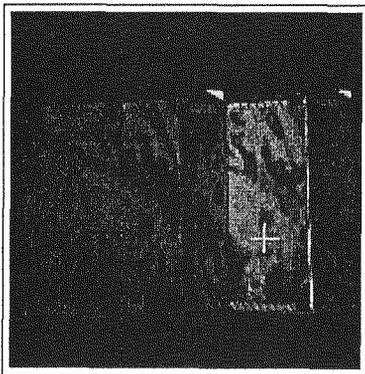
A partir des points 3D obtenus (figure 8), nous déterminons, pour chaque face, les paramètres des plans passant par les points au sens des « moindres carrés ». Nous pouvons ainsi quantifier l'erreur de planarité des deux faces reconstruites (tableau 2) ainsi que la distance séparant les deux plans. La distance entre les plans est de

Tableau 1 : Coordonnées des points appariés par corrélation.

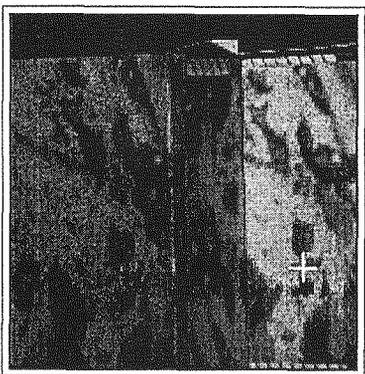
	coordonnées après corrélation simple (pixels)	coordonnées après corrélation sous-pixel (pixels)	dx (pixels)	dy (pixels)	dl (pixels)	df (pixels)
I_1	(285,240)	(284.83,241.21)	-0.17	1.21	3.76	3.25
I_2	(286,239)	(286.36,240.21)	0.36	1.21	3.50	3.12
I_3	(288,238)	(288.05,238.89)	0.05	0.89	3.26	2.76
I_4	(290,237)	(289.96,237.59)	-0.04	0.59	3.01	3.16
I_5	(292,236)	(292.01,236.37)	0.01	0.37	2.77	2.95
I_6	(294,234)	(294.42,234.56)	0.42	0.56	2.55	2.83
I_7	(297,232)	(297.21,232.65)	0.21	0.65	2.33	2.51
I_8	(300,230)	(300.39,230.51)	0.39	0.51	2.12	2.32
I_9	(304,227)	(304.21,227.89)	0.21	0.89	1.92	1.56
I_{10}	(314,221)	(313.89,221.49)	-0.11	0.49	1.55	1.56
I_{11}	(328,212)	(327.73,211.87)	-0.27	-0.13	1.21	0.74
I_{12}	(333,209)	(332.27,209.04)	-0.73	0.04	1.13	0.60
I_{13}	(337,206)	(336.61,205.96)	-0.39	-0.04	1.06	0.55
I_{14}	(341,203)	(341,203)	-	-	-	-



I_1



I_9



I_{13}

Figure 7. – Trois images de la séquence des boîtes.

42 mm. L'angle formé par les deux faces est d'environ 15 degrés. Ces résultats sont à replacer dans le contexte d'une stéréovision axiale où les angles de triangulation restent faibles.

Le paramètre σ_{di} représente l'écart-type de la distance des points à chaque plan. Nous avons ensuite éliminés les points reconstruits dont la distance au plan « moyen » était supérieure à σ_{di}

Sur la figure 9, nous avons représenté la reconstruction des deux faces de la boîte. Sur la première image les points 3D de chacune des faces sont reprojétés sur le plan correspondant aux moindres

Tableau 2 : Erreurs de planarité des boîtes

	face gauche	face droite
<i>nbPts Rec</i>	410	220
σ_{di} (mm)	4.67	5.79
<i>seuil = σ_{di}</i>		
<i>nbPts Rec</i>	280	147
σ_d (mm)	2.38	3.26

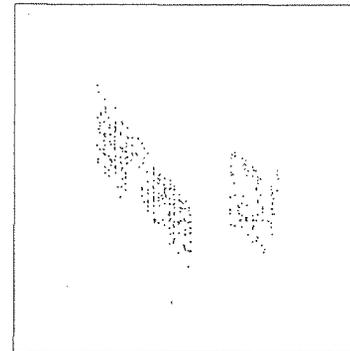


Figure 8. – Reconstruction des boîtes (nuage de points 3D bruts).

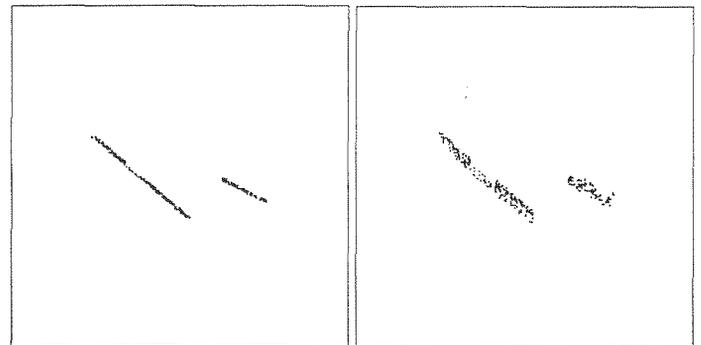


Figure 9. – Reconstruction des boîtes (vue de dessus).

carrés. Sur la seconde image, nous avons représenté les points 3D réels bruts. Le point d'observation est placé au dessus de l'objet pour mettre en évidence les éventuelles erreurs de profondeur.

5.3.2. Représentation du cube

De manière à réaliser une reconstruction plus propice à une validation terrain, nous avons essayé de reconstruire un cube de dimension parfaitement connue. Le cube est disposé de façon à observer trois de ses faces (figure 10). Pour la reconstruction, nous procédons comme pour les boîtes; 194 points sont sélectionnés sur la face supérieure, 187 points sur la face gauche et 181 points sur la face droite. Les points 3D (figure 11) nous permettent de déterminer les trois plans correspondant aux trois faces du cube.

- Planarité : Comme pour l'exemple précédent, l'estimation de chacun des plans aux moindres carrés nous renseigne sur la planarité de la reconstruction de chaque face. Les résultats

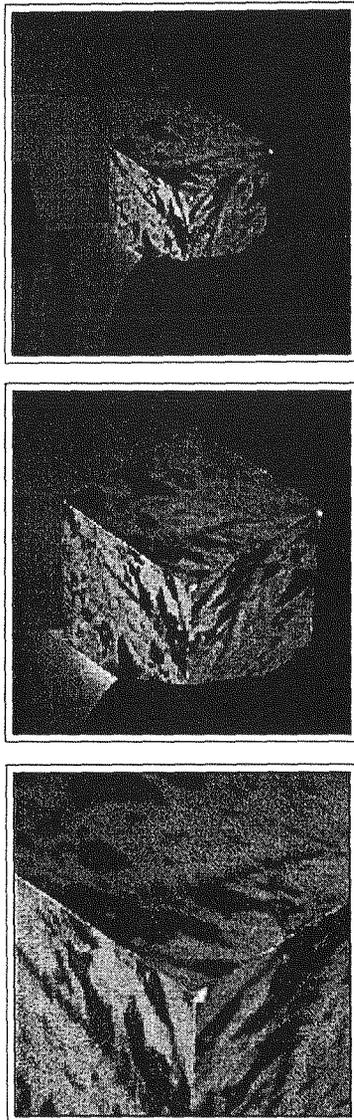


Figure 10. – Trois images de la séquence du cube.

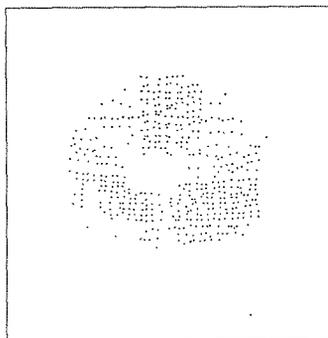


Figure 11. – Reconstruction du cube (nuage de points 3D).

quantifiant les erreurs de planarité sont regroupés dans trois tableaux 3, 4 et 5. Pour la détermination des paramètres des plans, nous avons instauré un système de filtrage des points

3D. Nous considérons dans un premier temps tous les points 3D reconstruits; nous déterminons le plan au sens des moindres carrés passant par ces points ainsi que l'écart-type correspondant (premières lignes des tableaux). Ensuite, par élimination des points dont la distance est supérieure à un seuil, nous déterminons de nouveaux plans passant par les points filtrés ainsi que les écarts-type correspondants. Ces calculs sont effectués pour différents nombres d'appariements entre les points :

Tableau 3 : Reconstruction de la face supérieure du cube

	$nbApp > 4$	$nbApp > 9$	$nbApp > 15$
$nbPts Rec$	194	194	67
σ_i (mm)	4.16	4.16	5.79
$seuil = \sigma_i$			
$nbPts Rec$	164	164	59
σ (mm)	1.98	1.98	1.78
$seuil = 2\sigma_i$			
$nbPts Rec$	184	184	62
σ (mm)	2.44	2.44	2.35
$seuil = 3\sigma_i$			
$nbPts Rec$	191	191	65
σ (mm)	2.97	2.97	3.61

Tableau 4 : Reconstruction de la face gauche du cube

	$nbApp > 4$	$nbApp > 9$	$nbApp > 15$
$nbPts Rec$	187	168	32
σ_i (mm)	7.61	4.18	2.04
$seuil = \sigma_i$			
$nbPts Rec$	155	137	22
σ (mm)	2.59	1.99	1.03
$seuil = 2\sigma_i$			
$nbPts Rec$	180	158	31
σ (mm)	3.93	2.52	1.95
$seuil = 3\sigma_i$			
$nbPts Rec$	182	164	32
σ (mm)	4.51	2.84	2.04

Tableau 5 : Reconstruction de la face droite du cube.

	$nbApp > 4$	$nbApp > 9$	$nbApp > 15$
$nbPts Rec$	181	164	31
σ_i (mm)	2.04	1.99	1.03
$seuil = \sigma_i$			
$nbPts Rec$	127	114	22
σ (mm)	1.16	1.01	0.49
$seuil = 2\sigma_i$			
$nbPts Rec$	173	156	28
σ (mm)	1.69	1.61	0.65
$seuil = 3\sigma_i$			
$nbPts Rec$	180	162	31
σ (mm)	1.97	1.86	1.03

Vers une reconstruction dense par zoom

- la première colonne considère les points appariés dans au moins quatre images ($nbApp > 4$),
- la seconde colonne prend en compte les points visibles dans au moins 9 images ($nbApp > 9$),
- la dernière colonne tient compte uniquement des points visibles sur toute la séquence.

La précision de reconstruction de la face droite du cube semble meilleure que celle des autres faces; en effet l'écart-type correspondant pour $nbApp > 4$ sans filtrage est seulement de 2.04 mm alors que pour la face gauche l'écart-type vaut 7.61 mm. Toutefois, dès lors qu'un filtrage est effectué sur les données brutes, l'écart-type diminue de façon significative. Bien entendu, plus le nombre d'appariements pris en compte augmente, plus le nombre de points diminue; les écarts-type obtenus avec $nbApp > 15$ sont en général meilleurs mis à part pour la face supérieure du cube. Etant donné le faible nombre de points mis en jeu, si quelques points sont mal reconstruits, les résultats sont faussés.

- Orthogonalité : Afin de vérifier la précision de reconstruction, nous calculons les angles formés par les faces prises deux à deux, respectivement les faces Droite-Gauche, Droite-Dessus et Gauche-Dessus (voir tableau 6).

Tableau 6 : Angles en degrés entre les faces du cube

	Dr-Ga	Dr-Des	Ga-Des
$nbApp > 4$			
$seuil = \sigma_i$	92.46	91.94	88.07
$seuil = 3\sigma_i$	92.38	92.50	87.37
$nbApp > 9$			
$seuil = \sigma_i$	89.14	93.50	83.43
$seuil = 3\sigma_i$	89.35	93.37	84.55
$nbApp > 15$			
$seuil = \sigma_i$	87.39	97.95	64.25
$seuil = 3\sigma_i$	87.60	97.13	62.63

Les angles ne sont pas parfaitement égaux à 90 degrés, les plans reconstruits ne sont donc pas parfaitement orthogonaux. Les angles sont moins proches de 90 degrés pour $nbApp > 15$; en fait le nombre de points mis en jeu est considérablement réduit et se concentre près de l'axe optique où les angles de triangulation sont très faibles. Cependant certaines erreurs proviennent en fait d'appariements erronés. En effet, le choix des points au départ est arbitraire; certains points sont difficiles à appairer s'ils sont situés dans une zone très peu texturée.

La figure 5.3.2 montre trois vues de points reconstruits par couple de faces pour une appréhension plus facile des résultats.

La première série d'images est la représentation des points projetés sur les plans aux moindres carrés; la seconde série montre les points réellement reconstruits sous les mêmes points de vue.

5.3.3. représentation du galet

Dans les deux exemples de reconstruction précédents, les objets appréhendés présentaient des surfaces planes, propices à des mesures de qualité de reconstruction. Nous présentons dans ce paragraphe la reconstruction d'une surface courbe, par l'analyse d'une séquence d'images représentant un galet.

Les figures ci-après présentent trois vues du galet reconstruit sans traitement de données (nuage de points 3D bruts).

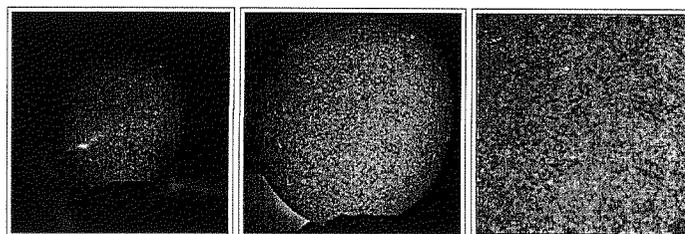
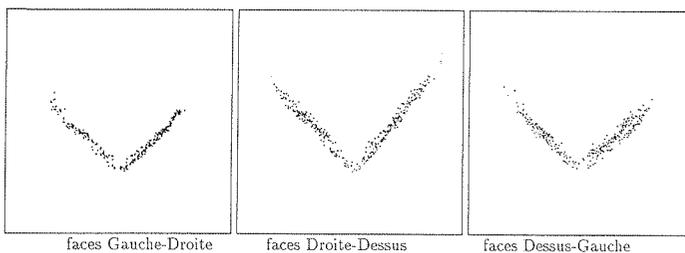
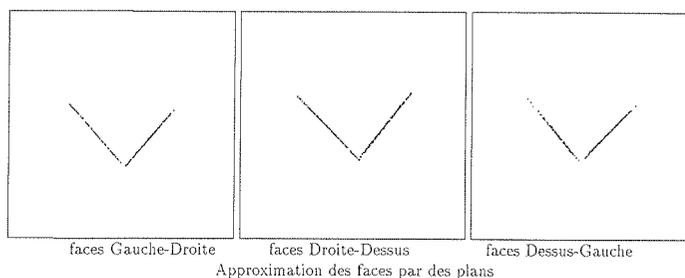


Figure 12. – Trois images de la séquence du galet.

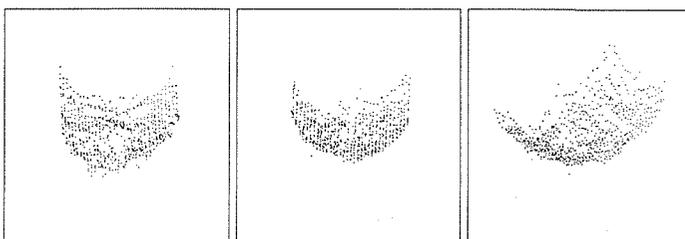


Figure 13. – Reconstruction du galet (nuage de points).

Malgré la texture de l'objet, des erreurs de reconstruction subsistent, dues en majeure partie à des imprécisions de mise en correspondance. Notons cependant que la forme générale du galet est conservée.

6. conclusion et discussion

Cet article présente une méthode de reconstruction dense d'objets à partir d'images prises à l'aide d'un zoom. C'est une méthode complète dans le sens où elle inclue le calibrage, l'appariement des points et la reconstruction 3D. Le problème de la mise en correspondance des points est résolu par une méthode de corrélation surfacique. La précision d'appariement est inférieure au pixel; une telle précision est nécessaire étant donné les faibles angles de triangulation mis en jeu lors de la reconstruction. La modélisation est réalisée dans un espace métrique défini par deux grilles de calibrage de très grande qualité. Une approximation locale basée sur le rapport anharmonique de quatre points permet d'établir une transformation entre l'image de la grille distordue et la grille réelle parfaite située devant la caméra; les phénomènes de distorsion sont ainsi automatiquement pris en compte.

Nous présentons plusieurs expériences, toutes réalisées à partir d'images réelles. Les résultats de reconstruction sont satisfaisants dans le sens où ils permettent une représentation cohérente des objets. Certains points sont toutefois mal reconstruits; ces erreurs proviennent en fait de mauvais appariements. En effet, dans le cas d'appariements très précis (au centième de pixel), la précision de reconstruction est inférieure au millimètre sur la coordonnée en z, donc la plus sensible [DLPD95]. Ici, le choix des points sur la première image est arbitraire. La méthode peut cependant être améliorée si l'on effectue au préalable un filtrage de ces points.

Il est notamment possible de ne conserver que les points dont « l'activité » est suffisamment grande, c'est à dire les points au voisinage desquels la fonction de luminance varie fortement.

Cette méthode de reconstruction pointue, porte en elle ses limites. Qu'en est-il si la méthode de calibrage par bi-plans nécessite des grilles de grande taille, ou que la profondeur de champ ou la définition des grilles dans l'image requiert un changement de grille? Les objectifs zoom que nous utilisons sont-ils mécaniquement suffisamment stables pour donner des mesures reproductibles? Autant de questions qui dépassent le cadre de cette application et auxquelles on se doit de réfléchir pour l'utilisation de ce type d'objectif en Vision.

BIBLIOGRAPHIE

- [Ack84] F. Ackermann. Digital Image Correlation : performance and potential application in Photogrammetry. *Photogrammetric Record*, 11(64) : 429-439, October 1984.
- [BB82] D.H. Ballard, C.M. Brown, Computer Vision, Prentice Hall
- [bey92] H.A. Beyer. Accurate Calibration of CCD Cameras, in *Proc. of Conference on Computer Vision and Pattern Recognition, Urbana Champaign, USA*, pages 96-101, 1992.
- [BMB94] P. Brand, R. Mohr, P. Bobet. Distorsions optiques : correction dans un modèle projectif. *9ème congrès AFCET RFIA*, pages 87-98, Paris, Janvier 1994.
- [Car91] C. Carmona, Etude de la stéréovision axiale. Modélisation mathématique et résolution algorithmique. *Thèse de Doctorat de l'Institut National Polytechnique de Toulouse*, 1991.
- [Del95] C. Delherm. Reconstruction volumique dense par stéréovision axiale. *Thèse de Doctorat de l'Université Blaise Pascal de Clermont-Ferrand, (754)*, Novembre 1995.
- [DG93] R. Deriche, G. Giraudon. A Computational Approach for Corner and Vertex Detection. *International Journal of Computer Vision*, 10(2) : 101-124, 1993.
- [DLPD95] C. Delherm, J.M. Lavest, B. Peuchot, N. Daucher, Reconstruction Implicite par Zoom. *Traitement du Signal*, 12(4) : 303-315, 1995.
- [EVF94] R. Enciso, T. Vieville, O. Faugeras. Approximation du changement de focale et mise au point par une fonction affine à trois paramètres. *Traitement du Signal*, 11(5) : 361-372, 1994.
- [GTK88] K.D. Gremban, C.H. Thorpe, T. Kanade, Geometric Camera Calibration using Systems of Linear Equations, *Proc. of IEEE Robotics and Automation*, pages 562-567, 1988.
- [HS88] Harris, M. Stephens. A combined corner and edge detector. in *proc. of Fourth Alvey Vision Conference*, pages 189-192, August 1988.
- [Law82] D.T. Lawton, Motion Analysis via Local Translational Processing, *IEEE Workshop on Computer Vision : Representation and control*, 1982.
- [Li94a] M. Li. Camera Calibration of the KTH Head-Eye System for Active Vision. *Proc. of the 3rd European Conference on Computer Vision*, pages 543-554, May 1994.
- [Li94b] M. Li. Camera Calibration of a Head-Eye System for Active Vision, *Proc of 3rd European Conference on Computer Vision, ECCV, Stockholm, May 1994*.
- [LL95] M. Li, J.M. Lavest, Some Aspect of Zoom Lens Camera Calibration, *To appear in PAMI, Technical Report CVAP 172, Computational Vision and Active Perception Laboratory, RIT Sweden February 95*.
- [LPDD94] J.M. Lavest, B. Peuchot, C. Delherm, M. Dhome. Reconstruction by Zooming from Implicit Calibration. *Proc of 1st IEEE International Conference on Image processing*, 2 : 1012-1016, Austin, November 1994.
- [LRD93] J.M. Lavest, G. Rives, M. Dhome. 3D Reconstruction by Zooming. *IEEE Transactions on Robotics and Automation*, 9 (2) : 196-208, April 1993.
- [LRD95] J.M. Lavest, G. Rives, M. Dhome. Modeling an Object of revolution by Zooming. *IEEE Trans. on Robotics and Automation*, 11(2) : 267-271, April 1995.
- [LT87] R.K. Lenz, R.Y. Tsai. Techniques for calibration of the Scale Factor and Image Center for High Accuracy 3D machine Vision Metrology. *IEE Int. Conf. on Robotics and Automation*, pages 68-75, March 1987.
- [MBK81] H.A. Martins, J.R. Birk, R.B. Kelley, Camera Models Based on Data from Two Calibration Planes. *Computer Graphics and Image Processing*, 17 : 173-180, 1981.
- [Mor77] H.P. Moravec, Towards Automatic Visual Obstacle Avoidance. in *Proc of 5th International Joint Conference on Artificial Intelligence*, page 584, August 1977.
- [Peu94] B. Peuchot, Utilisation de détecteurs sub-pixels dans la modélisation d'une caméra. *9ème congrès AFCET RFIA*, pages 691-695, Paris, Janvier 1994.
- [PFTV86] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling. Numerical Recipes. *Cambridge University Press*, 1986.
- [RA94] V. Rodin, A. Ayache. Stéréovision axiale : Modélisation et calibrage d'un système de prise de vue, reconstruction 3d d'objets naturels. *traitement du signal* 11 (5) : 373-392, 1994.
- [Rod93] V. Rodin. Stéréovision axiale : calibrage, appariement et reconstruction tridimensionnelle. Application en coelochirurgie. *Thèse de Doctorat de l'Institut National polytechnique de Toulouse*, Octobre 1993.
- [Tsa86] R.Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D machine Vision. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 364-374, June 1986.
- [TTG92] K. Tarabanis, R.Y. Tsai, D.S. Goodman. Modeling of Computer-Controlled Zoom Lens. in *Proc. of IEEE Conference on Robotics and Automation, Nice*, pages 1545-1551, 1992.

Vers une reconstruction dense par zoom

[WM91] G.Q. Wei, S.D. Ma. Two plane Camera Calibration : a Unified Model. in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 133-138, June 1991.

[WS93] R.G. Willson, S.A. Shafer. What is the Center of the Image? in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, New-York*, pages 670-671, June 1993.

[Zha93] C.S. Zhao. Reconstruction de surfaces tridimensionnelles en vision par ordinateur. *Thèse de doctorat, Institut National Polytechnique de Grenoble, France* 1993.

Manuscrit reçu le 20 mai 1996.

LES AUTEURS

Catherine DELHERM



Catherine Delherm a soutenu son Doctorat (Option Vision pour la Robotique) en novembre 1995, au LASMEA, URA 1793 du CNRS. Son sujet traitait de la reconstruction Dense par Stéréovision Axiale. Lauréate des bourses Régionales d'Excellence elle a travaillé durant 6 mois en collaboration avec le CERN sur les problèmes d'étalonnage fin de caméras. Elle occupe actuellement un poste d'ATER à l'Université du Havre et continue ses activités de recherche au LACOS (Lab. d'Analyse et de Commande des Systèmes.)

Nadine DAUCHER



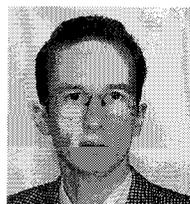
Nadine Daucher est Maître de Conférences à l'Université Blaise-Pascal de Clermont-Ferrand, spécialité Automatique. Ses activités de recherche concernent la vision monoculaire et la commande pour la robotique au sein du groupe GRAVIR du LASMEA.

Michel DHOME



Michel Dhome est Directeur de Recherche au CNRS et exerce au sein du LASMEA, URA 1793 du CNRS à Clermont-Ferrand. Il a obtenu une thèse de l'Université Blaise-Pascal en 1984 avant de séjourner un an comme boursier post-doctoral à Ottawa. Ses activités de recherche concernent le domaine de la Vision Artificielle et plus particulièrement la localisation, la modélisation et la reconnaissance d'objets volumiques.

Jean-Marc LAVEST



Jean-Marc Lavest est Maître de Conférences à l'Institut des sciences pour l'Ingénieur CUST de Clermont-Ferrand. Après un Séjour post doctoral au Royal Institute of Technology de Stockholm chez le Professeur Jan-Olof Eklundh il rejoint le groupe GRAVIR du LASMEA, URA 1793 du CNRS. Ses activités de recherche concernent l'utilisation de zoom en Vision Artificielle et les problèmes liés au calibrage précis des systèmes de vision.