

Dense Reconstruction by Zooming

C Delherm, JM Lavest, M Dhome, JT Lapresté

Université Blaise-Pascal de Clermont-Ferrand,
Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique
URA 1793 of the CNRS, 63177 Aubière Cedex, France

Abstract. Reconstruction by zooming is not an unachievable task. As it has been previously demonstrated, axial stereovision technics allows 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 equipped 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, 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 matching.
- Finally, the 3D surface is reconstructed using image point correspondances. The modelling algorithm does not require any explicit calibration model and the computations involved are straightforward. This approach uses several images of accurate regular grids placed on a micrometric table, as a calibration process [1]. Complete experiments on real data are provided and show that it is possible to compute 3D dense information from a zooming image set.

Key-words : Correlation, Dense Reconstruction, Axial Stereovision, Implicit Calibration.

1 Introduction

Zoom-lens are commonly used to capture precise details of a global scene but more recently [2], [3] have shown that is also possible to infer 3D information from a zoom-lens mounted on a static camera.

In this paper, using high quality calibration grids, the modelling problem is transferred into a metric space defined between two grids. In [1], experimentations on real data show that this new calibration approach leads to a much more accurate reconstruction than in previous experimentations. The accuracy achieved is better than 1 mm in z-coordinate for a planar grid located 1 meter in front of the camera. Even data close to the optical axis lead to relevant reconstruction. However, in experiments described in [1], points matching observed

through the whole images set have been carried out manually, by a sub-pixel cross detection.

In order to improve this step, this article proposes a matching algorithm based on a surfacic correlation method that takes into account the zooming effect.

Then, in order to valid this approach, dense reconstructions on real objects are presented and some statistical analysis performed to measure the modelling reliability.

2 Calibration Algorithm Using Two Reference Grids

Let us suppose we have a distorted grid image (Figure 1b); distortion phenomena can be produced by optics, electronics or sampling problems. A classical non linear pin-hole calibration tries to insert these phenomena into the mathematical model of projection but this kind of model can never be perfectly accurate. If it is possible, as shown in Figure 1a, to compute a transformation from the distorted grid to the regular one, then distortion effects can be corrected. In this way, any new image taken with the camera can be mapped in the reference grid, which is assumed to be as perfect as possible.

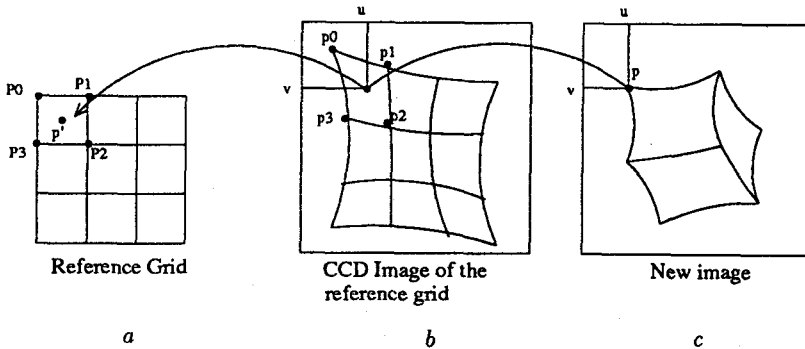


Fig. 1. Two Grids Calibration Principle: Local approximation

Let us suppose now that two parallel grids are used, for which relative space positions can be accurately determined. A micrometric table is used during calibration and an optical control can ensure these assumptions. These two simple images correspond to the basis of the calibration process and define a metric space. The grid closest to the camera defines the origin of this metric space, x and y coordinates correspond respectively to the row and column of the grid, z coordinate is defined along the grid axis of translation.

If a new object is digitized with the camera, it is possible to compute, for each image data, its corresponding point in the first and the second grid. By this way, we have estimated the 3D optical ray of the considered image data in the metric space defined by the two reference grids, without knowing any camera parameter. Of course, it is possible to use more than two grids. The accuracy of the computed 3D optical ray will be improved; nevertheless it will require more computation time and storage capacity.

This calibration process called 'Two Grids Calibration' is not new; [4] introduced it few years ago and [5], [6] have proposed several ways to solve this problem. Two distinct principles can be found:

- the first one tries to compute a global transformation from the distorted image content to the reference point located on each reference grid,
- the second one estimates a local approximation that maps an image point to its corresponding point on the reference grids.

From a computational point of view, global approximation is economical in execution time and storage capacity but errors are averaged over all points. Obviously, local interpolation provides more accurate results than global interpolation. In the local model, the closest calibration points (at least three) are searched around each point detected in the new image, and a local function is used to perform an interpolation over each local region.

Differences between results presented in literature depend on:

- local functions used to establish the relationship between the distorted image and the reference frame,
- but also, on the way the reference points are detected on the CCD matrix.

As indicated above, one of the main difficulties is to compute accurately each point position (p_0, p_1, p_2, p_3 in Figure 1b) in the distorted grid image. [7] has developed an original algorithm that fits directly into the grey level image a mathematical model of the photonic point response. Similar works have been realized by [8], [9] et [10]. In this way, the point location is obtained with high accuracy (more than 1/100 pixel) [7].

Let us consider a point $p(u, v)$ (Figure 1c) in a new distorted image. Knowing its coordinates (u, v) , it is possible to find in the stored CCD grid image, the 4 points (p_0, p_1, p_2, p_3 Figure 1b) around it. [11] has proposed several interpolation models and proved that a bi-cubic function leads to better results in the case of high distortion phenomena (X-Ray image), but for a standard camera, linear approximation leads to the same order of accuracy.

In the following experiments, a local function using the anharmonic ratio will be used to interpolate the coordinates of a new point p relatively to the local base (p_0, p_1, p_2, p_3). In order to avoid instability problem, calculus are realized in an homogeneous coordinates system.

3 Points Matching by Correlation

Correlation technics are commonly used to establish the correspondences between edge points in two images from a standard camera device. In case of images taken with a zoom-lens (i.e different magnification), correlation scores need to be modified to take into account the scale effect through the image set. Furthermore, the triangulation angles involved in axial stereovision require an accurate matching between features observed in the images set. The matching algorithm, proposed in this article, is done in two steps:

- The first one is called "basic correlation"; The correlation is applied from the less zoomed image (wide-angle field) to the most zoomed one (narrow-angle field). Matching is performed with an accuracy at least equals to one pixel.
- The second one is called "sub-pixel correlation". Using previous matched data, an iterative process is applied. Images are considered in the opposite way, it means from the most zoomed to the less zoomed image. Scale effect is explicitly introduced in the correlation parameters to obtain a sub-pixel accuracy.

3.1 Basic Correlation

The aim of this first algorithm is to follow a feature point (not necessarily an edge point) along all an images set taken with a zoom-lens. The correlation is performed from the image taken with the shortest focal-length (wide-angle field) to the longest one. Matching is realized step by step between two successive images I^k and I^{k+1} . The point $p_i^k(u_i^k, v_i^k)$ belonging to the image I^k is associated to the point $p_i^{k+1}(u_i^{k+1}, v_i^{k+1})$ of I^{k+1} that maximizes the correlation coefficient.

In order to look for point p_i^{k+1} in a sufficiently large area in the image I^{k+1} (without spending too many time), a pyramidal research process [12] has been implemented. It works with different images extracted from (I^k, I^{k+1}) but at a lower resolution.

3.2 Sub-Pixel Correlation

Results from the previous correlation algorithm, which is working with two consecutive images $((I_1, I_2), (I_2, I_3), \dots (I_l, I_m))$, can diverge along the images set. In order to avoid this phenomena, the sub-pixel matching will be done according to the same reference image along all the sequence. Let us define $\{p_i^1, \dots, p_i^k, \dots, p_i^m\}$ a set of matched point p_i , viewed from image I^1 to image I^m and computed by the first correlation algorithm. The last point p_i^m and its grey-level neighbourhood will be used as the reference point during all the sub-pixel process. This assumption seems to be logical, because most information is contained in the image I^m (taken with the longest focal length). Of course, to find accurately two matched points between images I^m and I^l ($l \in [1 \dots m - 1]$), zooming effect has to be taken into account. Therefore, the image I^m will be under-sampled and smoothed automatically, and the algorithm will compute the best location

of the correlation window in I^l that fits the data. Parameters are estimated using an iterative process; the following criterion is minimised:

$$C(I^m, I^l, \Delta, f, du, dv) = \sum_{i=-n}^{+n} \sum_{j=-n}^{+n} (I^m(u_i^m + \Delta.i, v_i^m + \Delta.j, f) - I^l(u_i^l + du + i, v_i^l + dv + j, 1))^2 \quad (1)$$

where parameters (Δ, f, du, dv) are real values (we need sub-pixel accuracy) and take into account respectively the undersampling Δ (i.e. the zooming effect), the smoothing f (due to undersampling), and the subpixel location du, dv of the correlation window in the second image.

In equation [1], $I^k(u, v, \rho)$ is a function that defines the grey-level value according to real point coordinates (u, v) in the image I^k . As image I^k can be undersampled, a smoothing effect is realized on a window $(\rho \times \rho)$ centred in (u, v) . It gives a ponderation to the grey level information (of each pixel) contained in the image I^k , by a coefficient proportionnal to the pixel surface (greater than 1 due to the zooming phenomena) actually included in the considered window.

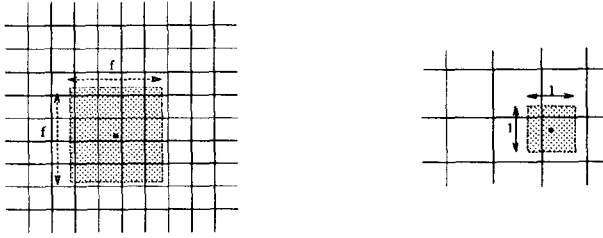


Fig. 2. *Calculus area associated with $I^m(u, v, f)$*

Calculus for $I^l(u, v, 1)$

Solution of equation [1] is found by searching the value of coefficients (Δ, f, du, dv) that minimises criterion C . We have to solve a non linear optimisation problem ; as it is not possible to obtain an analytical expression of the derivatives, the first derivatives of the criterion C are numerically estimated.

After the completion of the sub-pixel algorithm, data matched along the images set, are available. Each point coordinates are estimated with a sub-pixel accuracy. Using this information, the reconstruction process will provide the 3D coordinates of all the feature points, computed in the metric space of calibration grids.

4 Modelling Process

Using considerations developped in section 2, the modelling algorithm does not involve any difficulties. For each focal-length position used during the zooming

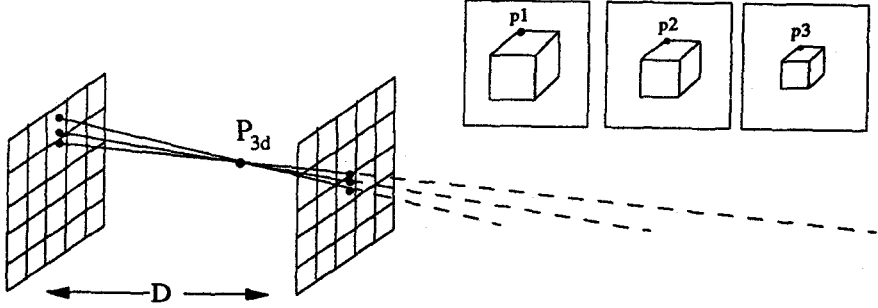


Fig. 3. Modelling Process.

sequence, two calibration grids are previously digitized and each calibration point accurately detected and memorized. This step represents the calibration process.

Given a same physical point p^i , matched in all the new images set (Figure 3), (i represents the number of different focal-length used during the zooming sequence), it is possible to estimate the 3D optical rays relative to each image and expressed in the metric space of the grids.

Let (P_1^i, P_2^i) be the interpolated points of p^i relative to the i_{th} image and respectively defined in the first and second grid.

Let V_i be the vector between (P_1^i, P_2^i) :

$$V_i = \begin{pmatrix} P_2^i.x - P_1^i.x \\ P_2^i.y - P_1^i.y \\ P_2^i.z - P_1^i.z \end{pmatrix} \quad (2)$$

Let \mathcal{P}_H^i be the quasi horizontal plane and \mathcal{P}_V^i be the quasi vertical plane that contain V_i .

The plane equations are given as follows:

$$\begin{cases} A_1^i x + B_1^i y + C_1^i z + D_1^i = 0 \\ A_2^i x + B_2^i y + C_2^i z + D_2^i = 0 \end{cases} \quad (3)$$

Obvioulsy, the three-dimensional coordinates of the observed point $P_{3d}(x, y, z)$ correspond to the intersection of all the optical rays (Figure 3) and are computed by solving the overdetermined linear system :

$$A_j^i x + B_j^i y + C_j^i z = -D_j^i \quad j = 1 \dots 2, i = 1 \dots n \quad (4)$$

5 Experiments

5.1 Experimental conditions

All experimental results presented in this section are taken from real images. We have used a *Angénieux* zoom lens T14 \times 9;

The reconstruction process is realized in three steps: first of all calibration, then matching by correlation, finally 3D reconstruction.

– Calibration

Given 16 focal-lengths (between 25 à 90 *mm* according to the zoom length parameters), two accurate patterns are digitized, to create the grid metric space. Aperture and focus remain constant during images acquisition. The size of a unit square of the grid is equal to 6.2 \times 6.2 *mm*. The distance between the two grids used during the calibration is set to 15 *cm*. The first one is located at approximatively 90 *cm* from the camera. Each grid is preprocessed and each calibration point (row-column intersection) located, matched and stored in memory.

– Matching and reconstruction using a new set of images

After the calibration step, the new object to be reconstructed is located in front of the camera at approximatively 1 meter. This distance corresponds to the metric space defined between the two grids. Of course, it is not necessary to put the object accurately between the two grids. As it is described in [13], it is possible to achieve good reconstruction results even if the object is not exactly located in the metric space defined by the grids. However, if the object is too far from this space, the reconstruction accuracy is decreasing. Sixteen images are digitized using the same focal length as those used during the calibration step. A mechanical device makes possible to accurately recover the different focal length position on the zoom.

On the first images of the new set, a given number of points to be reconstructed are manually selected. They cover all the image surface. As they are randomly selected, this step could easily be automated.

The correlation algorithm is performed for all the points. After convergence, it gives the coordinates of all paired points along the image set. As it can be noticed, some points located near the image border, quickly disappear due to the magnification effect and are only matched in a few images. In order to achieve accurate reconstruction results, only points that appear in at least five images will be taken into account for reconstruction purpose. It means that the 3D coordinates of a reconstructed point correspond to the best intersection, using a least square approximation, of at least five optical rays defined in the grids metric space.

In the following sections, two experimentations of differents reconstructed objects will be presented : a 3D cube (size 10 \times 10 \times 10 *cm*) and a pebble (whose the size is approximately 10 \times 8 \times 5 *cm*). 3D coordinates are obtained for each point taken into account in the first image. As it is really difficult to recognize the object shape from a set of independant 3D points, the reconstruction results

will be presented on different ways according to the object, in order to evaluate the modelling accuracy.

5.2 Cube reconstruction

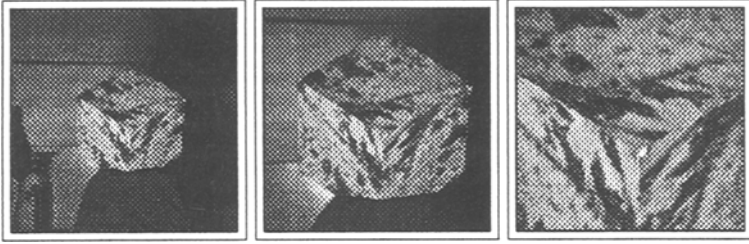


Fig. 4. *Three views of the cube image set*

The cube is placed in front of the camera in order to observe three faces (Figure 4). To perform statistical studies, points belonging to each face are independently selected (approximately one hundred points per face). By doing so, it is possible to estimate the reconstructed points face after face. Angle and planarity between them are estimated using a least square approximation.

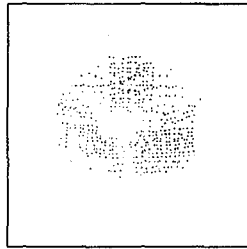


Fig. 5. *Cube Reconstruction (dense map of 3D points)*

Figure 5 shows the dense reconstructed map obtained with data collected on the three faces.

- Orthogonality : To verify the modelling accuracy, the angles between each plane are estimated (Table 1):
Reconstructed planes are not perfectly perpendicular. Some points are badly reconstructed and introduce major errors in the statistical study. As the initial selection of point to be reconstructed is randomly done, some of them are located in area where grey level information is poor; these points matchings are inaccurate and lead to bad reconstruction results.

Table 1. Angles in degrees between the cube faces

| | <i>Right-Left</i> | <i>Right-Top</i> | <i>Left-Top</i> |
|--------------------------------|-------------------|------------------|-----------------|
| <i>nbMatch</i> > 4 | | | |
| <i>threshold</i> = σ_i | 92.46 | 91.94 | 88.07 |
| <i>threshold</i> = $3\sigma_i$ | 92.38 | 92.50 | 87.37 |
| <i>nbMatch</i> > 9 | | | |
| <i>threshold</i> = σ_i | 89.14 | 93.50 | 83.43 |
| <i>threshold</i> = $3\sigma_i$ | 89.35 | 93.37 | 84.55 |
| <i>nbMatch</i> > 15 | | | |
| <i>threshold</i> = σ_i | 87.39 | 97.95 | 64.25 |
| <i>threshold</i> = $3\sigma_i$ | 87.60 | 97.13 | 62.63 |

In Table 1, parameter (*nbMatch*) gives the minimum number of matched data to reconstruct a given point; for example (*nbMatch* > 9) means that a point will be reconstructed if and only if it appears at least in 9 images. The threshold value σ_i is used to suppress some bad reconstructed points before computing angles between planes.

It can be noticed that angles are close to 90 *degrees*. Mean error increases when statistical studies are realized with (*nbMatch* > 15). But in this case, only points visible during all the images set (it means points close to the optical axis) are taken into account. In this case the triangulation angles are the smallest.

- **Planarity** : As for previous example, a least square estimation allows to establish the planarity error of each faces. Results are presented in (Tables 2, 3) and expressed in millimeters.
 - (*nbMatch*) gives the minimum number of images where a point should be visible to be reconstructed.
 - (*nbPtsRec*) gives the number of points taken into account according to (*nbMatch*) and the threshold values.
 - (*threshold*) allows to suppress points too far from the mean plane.

In Figure 6 different points of view of the reconstructed object are presented. To have an easier representation of dense data, 3D reconstructed points are presented using couple of faces. As it can be noticed, results are quite good, the real shape is almost estimated even for data close to the image center. A complete study of the triangulation angles influence is realized in [1].

5.3 Pebble reconstruction

In previous experiments, the object to be reconstructed was composed by a set of plane surfaces. Statistical studies could easily be performed to estimate the

Table 2. Cube Reconstruction : Top Face

| | $nbMatch > 4$ | $nbMatch > 9$ | $nbMatch > 15$ |
|-------------------------|---------------|---------------|----------------|
| $nbPts Rec$ | 194 | 194 | 67 |
| $\sigma_i(mm)$ | 4.16 | 4.16 | 5.79 |
| $threshold = 3\sigma_i$ | | | |
| $nbPts Rec$ | 191 | 191 | 65 |
| $\sigma(mm)$ | 2.97 | 2.97 | 3.61 |
| $threshold = 2\sigma_i$ | | | |
| $nbPts Rec$ | 184 | 184 | 62 |
| $\sigma(mm)$ | 2.44 | 2.44 | 2.35 |
| $threshold = \sigma_i$ | | | |
| $nbPts Rec$ | 164 | 164 | 59 |
| $\sigma(mm)$ | 1.98 | 1.98 | 1.78 |

Table 3. Cube Reconstruction : Left Face / Right Face

| | $nbMatch > 4$ | $nbMatch > 9$ | $nbMatch > 15$ | $nbMatch > 4$ | $nbMatch > 9$ | $nbMatch > 15$ |
|-------------------------|---------------|---------------|----------------|---------------|---------------|----------------|
| $nbPts Rec$ | 187 | 168 | 32 | 181 | 164 | 31 |
| $\sigma_i(mm)$ | 7.61 | 4.18 | 2.04 | 2.04 | 1.99 | 1.03 |
| $threshold = 3\sigma_i$ | | | | | | |
| $nbPts Rec$ | 182 | 164 | 32 | 180 | 162 | 31 |
| $\sigma(mm)$ | 4.51 | 2.84 | 2.04 | 1.97 | 1.86 | 1.03 |
| $threshold = 2\sigma_i$ | | | | | | |
| $nbPts Rec$ | 180 | 158 | 31 | 173 | 156 | 28 |
| $\sigma(mm)$ | 3.93 | 2.52 | 1.95 | 1.69 | 1.61 | 0.65 |
| $threshold = \sigma_i$ | | | | | | |
| $nbPts Rec$ | 155 | 137 | 22 | 127 | 114 | 22 |
| $\sigma(mm)$ | 2.59 | 1.99 | 1.03 | 1.16 | 1.01 | 0.49 |

reconstruction quality. In the following experiment, a pebble reconstruction is performed. We present the modelling results obtained with a curved surface, using a set of pebble images.

Figures 8 shows three different views of the reconstructed object. In spite of the object texture, reconstruction errors still remain, due to a lack of accuracy during the matching algorithms. However, the global object shape is well estimated.

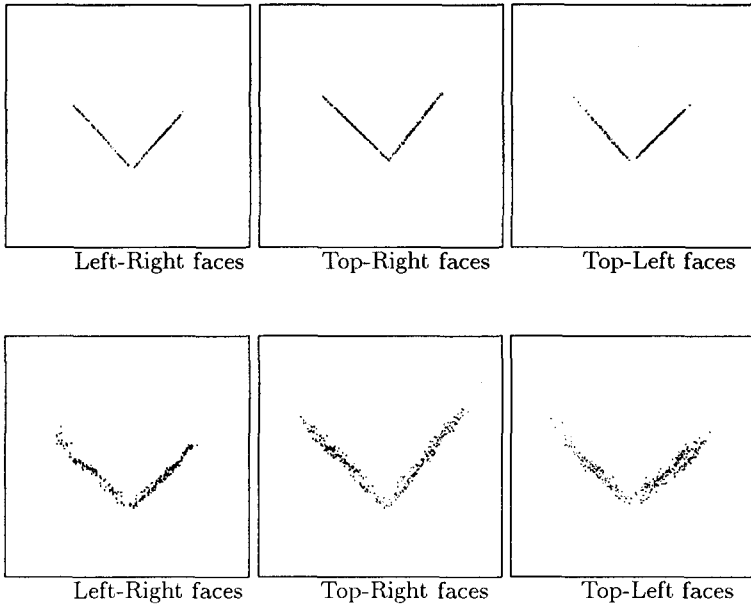


Fig. 6. *Cube Reconstruction*

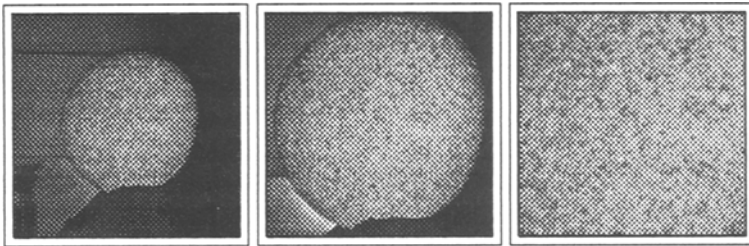


Fig. 7. *Three images of the pebble set*

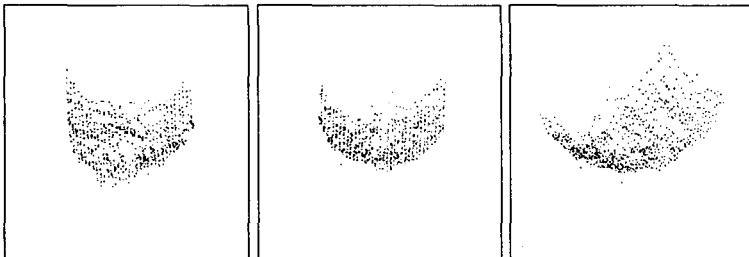


Fig. 8. *Pebble Reconstruction (dense map of 3D points)*

6 Conclusion

This article describes a dense reconstruction method from images taken with a zoom-lens. Calibration, matching and reconstruction steps are detailed. The matching problem between points viewed along the zooming set is performed with an iterative correlation algorithm. Zooming effect is explicitly introduced in the correlation parameters in order to achieve a sub-pixel accuracy.

Such an accuracy is essential due to the small angles involved in zooming reconstruction. The modelling algorithm is realized in a particular metric space composed by two accurate calibration grids. A local approximation allows to transform the distorted image content into the grids space which is supposed distortion free. By doing so, all distortion phenomena (optical and electronics) are automatically taken into account.

Several experiments on real images are presented. Promising results are obtained ; the global object shapes are well estimated, even for data close to the image center.

References

1. JM Lavest, C Delherm, B Peuchot, and N Daucher. Implicit Reconstruction by Zooming. *To appear in Computer Vision, Graphics and Image Processing*, 1995.
2. JM Lavest, G Rives, and M Dhome. 3D Reconstruction by Zooming. *IEEE Transactions on Robotics and Automation*, 9(2):196–208, April 1993.
3. JM Lavest, G Rives, and M Dhome. Modeling an Object of Revolution by Zooming. *IEEE Trans. on Robotics and Automation*, 11(2):267–271, April 1995.
4. HA Martins, JR Birk, and RB Kelley. Camera Models Based on Data from Two Calibration Planes. *Computer Graphics and Image Processing*, 17:173–180, 1981.
5. KD Gremban, CH Thorpe, and T Kanade. Geometric Camera Calibration using Systems of Linear Equations. *Proc. of IEEE Robotics and Automation*, pages 562–567, 1988.
6. GQ Wei and SD Ma. Two Plane Camera Calibration: a Unified Model. *in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 133–138, June 1991.
7. B Peuchot. Utilisation de détecteurs sub-pixels dans la modélisation d'une caméra. *9ème congrès AFCET RFIA*, pages 691–695, Paris, January 1994.
8. P Brand, R Mohr, and P Bobet. Distorsions optiques : correction dans un modèle projectif. *9ème congrès AFCET RFIA*, pages 87–98, Paris, January 1994.
9. R Deriche and G Giraudon. A Computational Approach for Corner and Vertex Detection. *International Journal of Computer Vision*, 10(2):101–124, 1993.
10. HA Beyer. Accurate Calibration of CCD Cameras. *in Proc. of Conference on Computer Vision and Pattern Recognition, Urbana Champaign, USA*, pages 96–101, 1992.
11. B Peuchot. Camera Virtual Equivalent Model, 0.01 Pixel Detector. *Computerized Medical Imaging and Graphics*, 17(4-5):289–294, 1993.
12. CS Zhao. Reconstruction de surfaces tridimensionnelles en vision par ordinateur. *Thèse de doctorat, Institut National Polytechnique de Grenoble, France*, 1993.
13. C Delherm, JM Lavest, B Peuchot, and N Daucher. Reconstruction implicite par zoom. *To appear in Traitement du signal*, 1995.