Global 3D Planar Reconstruction with Uncalibrated Cameras and Rectified Stereo Geometry

Jean-Philippe Tarel
INRIA
Domaine de Voluceau, Rocquencourt,
B.P. 105, 78153 Le Chesnay Cedex, France.
http://www-syntim.inria.fr/~tarel/
E-mail: Jean-Philippe.Tarel@inria.fr

Abstract

This article presents a seldom studied 3D reconstruction approach, and focus on its numerous advantages. It is a global approach, in which the algorithm attempts to reconstruct geometric high level features by using only global attributes of the image projections of the feature. Our elementary feature is a 3D planar patch, and geometric moments are the global attributes. The proposed method of reconstruction has the following advantages:

- *the use of the geometric moments of image region does not require a pixel-to-pixel matching and yields robustness to small errors of segmentation on region edges,*
- *the rectified stereo geometry constraint allows reconstruction with uncalibrated or calibrated cameras when epipolar geometry is known,*
- *valid matched regions are selected and thus the probably occluded planar patches are not reconstructed,*
- *interpolation of views between the left and right cameras can be performed to produce synthetic intermediate views of the observed scene.*

Experiments on real and synthetic stereograms are presented to illustrate the advantages of the approach.
Keywords: *3D, Stereo, Reconstruction, Planar Assumption, Uncalibrated Camera, Rectified geometry, Moments of Inertia, Geometric Attributes Affine Invariants, View morphing.*

1 Introduction

Since the introduction of early stereo techniques, steady efforts have been made to robustly recover depth from stereo images. Many studies have focused on the classical goal of autonomous robot navigation, but new applications in scene analysis for virtual environment

construction and synthesis require a relatively high level of understanding, in terms of where visible *surfaces* are located, as opposed to less structured information such as pointwise features or segments. Thus, the planar surface assumption (which holds reasonably well in indoor static scenes) has received most attention so far.

The planar assumption is often used as a constraint for other techniques, such as stereo contour reconstruction [1, 2] or MRF field-based correlation [3]. This approach can be seen as a local approach for planar patch reconstruction. It is worth noting that most studies do not reconstruct 3D planar patches in a global manner.

In [4] for the first time, the global approach is presented. While it is proved than no local matching is needed for planar patch reconstruction from two projections, the technique presented assumes a *calibrated and rectified* stereoscopic pair. Later, the global approach was extended to the case where camera geometry is calibrated and the stereo geometry is near to the rectified geometry, by using an approximation of the perspective projection [5].

In this article, a technique for the 3D global reconstruction of planar patches from a single rectified stereo pair is proposed, which is valid for *both calibrated and uncalibrated cameras*. In the case of uncalibrated cameras, only the epipolar geometry is assumed known. The method requires that region segmentation has been performed on both images, as well as the matching between 2D primitives, with methods such as described in [6, 7]. Notice that a further advantage of the proposed closed-form solution for planar patches is that it also valid for planar curves and discrete sets of coplanar points.

Our procedure fundamentally relies on the linearity of the pixel matching relationship of a point in a planar patch. In section 2, such equations are used for 3D planar reconstruction with simple computations in a transformed coordinate system which we introduce and call the *disparity space*.

Section 3 deals with using the obtained equations to compute the plane parameters which give the best *global* correspondence between the pair of projected regions. Our emphasis is on avoiding any local feature matching. We propose a scheme that is solely based on the regions shape and uses both center of gravity and centered moments of inertia as geometric attributes. Experiments on real data are presented in section 4.

It is well known that occlusion presents the main challenge to geometric region-based stereo algorithms. Furthermore, segmentation is often prone to errors. Yet very few attempts have been made to detect and quantify these effects precisely. Therefore, in section 5, we propose a way of checking the stereo data is consistent or not, and this enables us to decide whether the reconstruction is valid or not. It is shown that, affine invariants computed on the matched regions are robust enough to detect unreliable data, i.e. when a region's geometry is *inconsistent* with the hypothesis of the projection of a planar patch.

Finally, section 6 considers view morphing and synthesizing images of the scene from other points of view by taking advantage of the *disparity space* with uncalibrated cameras.

2 Problem Formulation

In this section, we first define rectified stereo geometry in a general sense. Then, we consider advantages of using the rectified geometry for planar patch reconstruction with calibrated

and uncalibrated stereo geometry. Finally, the concept of *disparity space* is introduced and its properties are explained.

2.1 The Need for a Rectified Stereo Geometry

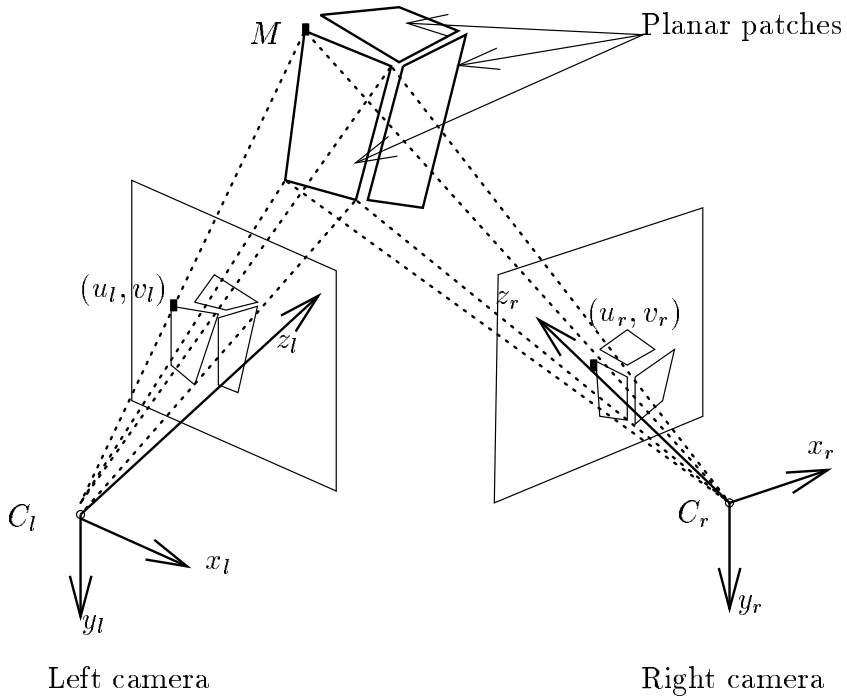


Figure 1: 3D stereo reconstruction of planar patches.

Consider the set-up in Figure 1, where a stereo system (or stereo rig) is looking at a scene consisting of planar patches. One can establish [8] the equation linking the two projections (u_r, v_r) and (u_l, v_l) of a given 3D point M lying on a plane P is a homographic (or projective) transform. By introducing homogeneous coordinates, this equation can be written:

$$\begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} \equiv \mathcal{H} \begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix}$$

where $\mathcal{H}$ is a 3×3 matrix and its components are functions of the plane parameters. The symbol $\equiv$ denotes equality up to a scalar. This relationship is nonlinear and as such of little practical use in a global approach, where the algorithm attempts to reconstruct 3D planar patches by using only geometric moments of its image projections. Geometric moments can only be used when the homographic transform $\mathcal{H}$ is degenerated and is in fact an affine transform. It can be shown [9] that the transform is affine for all planes only when image planes (or retinal planes) are parallel to the line defined by the two cameras' centers C_l and C_r. This kind of stereo geometry is the more general rectified stereo geometry. We named

it *uncalibrated cameras and rectified stereo geometry*. Indeed, the constraint to have parallel image planes applies even if cameras calibration parameters (intrinsic parameters of the two cameras and relative position of the right camera to the left one) are unknown.

When camera calibration is known, it is possible to be more specific. The definition usually used [10] is to assume that the stereo geometry is rectified if and only if epipolar lines are corresponding raster lines in both images. This defines the *calibrated cameras and rectified stereo geometry*.

In practice, an intermediate to the two above stereo geometries proves useful. In this stereo geometry, called *oriented cameras and rectified stereo geometry*, the constraint is that epipolar lines are raster lines. With such definition, rectified stereo geometry is less constrained than the one usually used because the two cameras can have different focal lengths and unknown intrinsic parameters. Contrary to the uncalibrated cameras and rectified stereo geometry, in the intermediate stereo geometry, the orientation of both cameras is aligned with the line joining centers C_l and C_r. An example of this kind of stereo geometry is shown in Figure 11(a) and (b) where epipolar lines in the left and right images are parallel but not aligned.

2.2 Disparity Space

In the case of a rectified stereo system, the equations linking a 3D point M to its two projections are drastically simplified. Intermediate reference systems of cameras are defined as original reference systems eventually rotated along the view axis to align the line joining C_l and C_r. If we denote (x_l, y_l, z_l) and (x_r, y_r, z_r) the coordinates of point M in the intermediate reference systems of the left and right cameras, we have:

$$\left.\begin{array}{rcrcl} y_l & = & y_r & = & y \\ z_l & = & z_r & = & z \\ x_l & = & x_r + \delta & = & x \end{array}\right\}$$

Parameter δ is the distance between the two centers C_l and C_r (see Figure 2). In other words, the intermediate left camera coordinates system is chosen as the reference system.

2.2.1 Calibrated Cameras

When the two cameras are calibrated and the stereo geometry is rectified, the intermediate left and right camera reference systems can be chosen to be the same as the intermediate ones. Since intrinsic calibration parameters are known, image coordinates are directly in mm, and thus the left and right projections are given by:

$$\left.\begin{array}{rcl} u_l & = & \dfrac{x}{z} \\ u_r & = & \dfrac{x-\delta}{z} \\ v_l & = & v_r = \dfrac{y}{z} \end{array}\right\} \qquad (1)$$

As expected, we have the epipolar relation $v_r = v_l$, since epipolar lines are raster lines aligned in the left and right images. This set of equations (1) is easily inverted to obtain coordinates

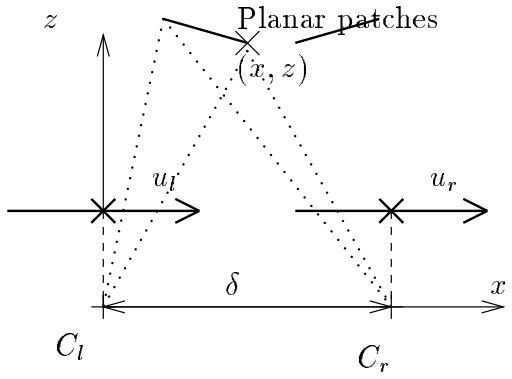


Figure 2: 2D section of a rectified camera stereo system.

(x, y, z) of point M as a function of images coordinates (u_l, v_l) and (u_r, v_r):

$$\left.\begin{array}{rcl} x & = & \dfrac{u_l \delta}{u_l - u_r} \\ y & = & \dfrac{v_l \delta}{u_l - u_r} \\ z & = & \dfrac{\delta}{u_l - u_r} \end{array}\right\} \tag{2}$$

At this point, it is convenient to map the Cartesian 3D space (x, y, z) into the coordinate space (u_l, u_r, v_l). The second bisector of this new coordinate space is the disparity axis (since disparity is $u_l - u_r$, see also Figure 13), thus we call this space the *disparity space*. A similar space is introduced by some authors and named "search space for intra-scanline search" [11], or "cross correlation volume" [12], but without taking advantage of its properties for planar reconstruction with uncalibrated cameras. The most useful advantage of the *disparity space* for reconstruction is that, as shown in Figure 3, the projection is exactly orthogonal on each image and not projective as in real 3D space.

2.2.2 Oriented Cameras

When camera calibration is not completely known, lets us introduce pixel coordinates, and thus the original reference systems of the left and right cameras.

In the oriented camera case, epipolar lines are raster lines. Therefore, both original camera reference systems are aligned with the intermediate ones, and the left and right projections are:

$$\left.\begin{array}{rcl} u_l & = & \alpha_{ul} \dfrac{x}{z} + u_{0l} \\ u_r & = & \alpha_{ur} \dfrac{x - \delta}{z} + u_{0r} \\ v_l & = & \alpha_{vl} \dfrac{y}{z} + v_{0l} \\ v_r & = & \alpha_{vr} \dfrac{y}{z} + v_{0r} \end{array}\right\} \tag{3}$$

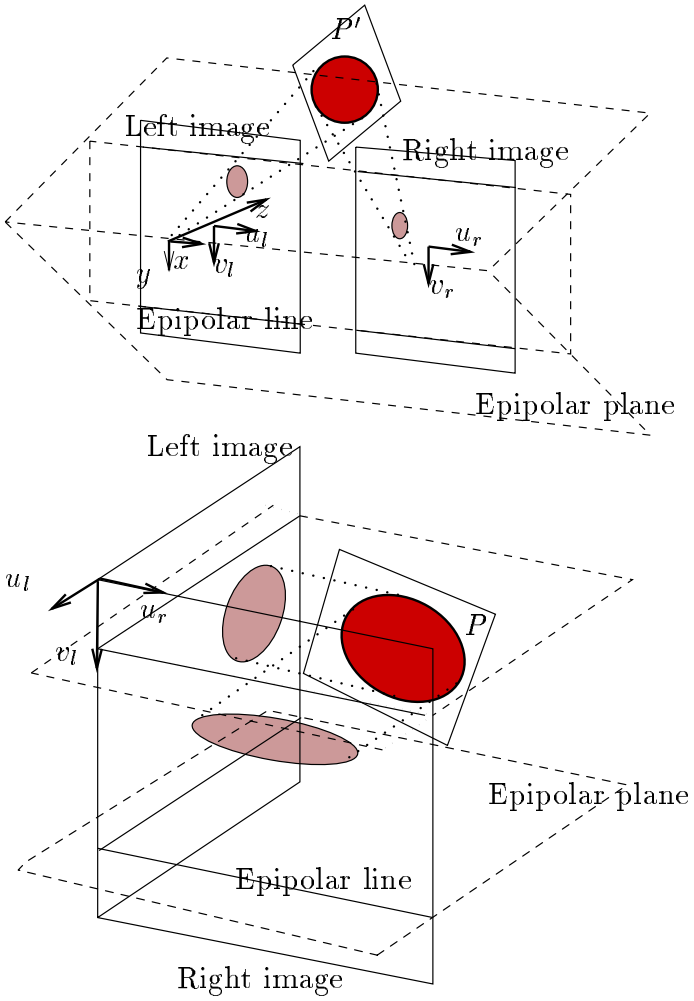


Figure 3: 3D stereo reconstruction of planar patches in the physical 3D space (above) and in the disparity space (below).

The image center (u_0, v_0), expressed in pixels, is defined as the intersection between the optical axis and the CCD matrix. Vertical and horizontal sample sizes on image are $1/\alpha_u$ and $1/\alpha_v$, respectively. These previous four parameters define the intrinsic parameters of a camera. Notice that using the last two equations of (3), we deduce the epipolar relation:

$$\alpha_{vr}(v_l - v_{0l}) = \alpha_{vl}(v_r - v_{0r})$$

As in calibrated case described previously, this set of equations is inverted to obtain coordinates (x, y, z) of point M as a function of image coordinates (u_l, v_l) and (u_r, v_r):

$$\left.\begin{array}{rcc} x & = & \dfrac{(u_l - u_{0l})\delta\alpha_{ur}}{\alpha_{ur}u_l - \alpha_{ul}u_r + \alpha_{ul}u_{0r} - \alpha_{ur}u_{0l}} \\ y & = & \dfrac{(v_l - v_{0l})\delta\alpha_{ur}\alpha_{ul}}{\alpha_{vl}(\alpha_{ur}u_l - \alpha_{ul}u_r + \alpha_{ul}u_{0r} - \alpha_{ur}u_{0l})} \\ z & = & \dfrac{\delta\alpha_{ur}\alpha_{ul}}{\alpha_{ur}u_l - \alpha_{ul}u_r + \alpha_{ul}u_{0r} - \alpha_{ur}u_{0l}} \end{array}\right\} \quad (4)$$

And, the Cartesian 3D space (x, y, z) can also be mapped into the coordinate space (u_l, u_r, v_l). Consequently, the introduction of the *disparity space* is still valid and can be used with the same advantages with oriented cameras than with calibrated cameras.

In particular, the mapping between the real and *disparity space* is homographic, as is proved by equations (3) or (4). Consequently, a 3D plane P': $a'x+b'y+c'z = d'$ is transformed into another plane P in the *disparity space*: $au_l + bu_r + cv_l = d$. The relationship between the equations of the two planes is obtained by substituting equations (3) in plane equation P:

$$\left.\begin{array}{rcc} a' & = & a\ \alpha_{ul} + b\ \alpha_{ur} \\ b' & = & c\ \alpha_{vl} \\ c' & = & a\ u_{0l} + b\ u_{0r} + c\ v_{0l} - d \\ d' & = & b\ \alpha_{ur}\ \delta \end{array}\right\} \quad (5)$$

2.2.3 Uncalibrated Cameras

With oriented cameras and rectified stereo geometry, camera orientation is assumed aligned with the line joining camera centers. This is not necessarily true in the uncalibrated case: an additive orientation parameter ϕ of the image plane must be introduced for each camera. Thus, the equations describing image coordinates in pixels as functions of 3D coordinates become:

$$\left.\begin{array}{rcc} u_l & = & \alpha_{ul}\dfrac{\cos\phi_l x - \sin\phi_l y}{z} + u_{0l} \\ u_r & = & \alpha_{ur}\dfrac{\cos\phi_r(x-\delta) - \sin\phi_r y}{z} + u_{0r} \\ v_l & = & \alpha_{vl}\dfrac{\sin\phi_l x + \cos\phi_l y}{z} + v_{0l} \\ v_r & = & \alpha_{vr}\dfrac{\sin\phi_r(x-\delta) + \cos\phi_r y}{z} + v_{0r} \end{array}\right\} \quad (6)$$

While equations (6) are more complex in comparison to equations (3), it is still possible to rewrite this set of 4 equations as:

- 3 equations where (x, y, z) is mapped to the disparity space (u_l, u_r, v_l) in an homographic way,
- and the linear epipolar relation which is independent of 3D point coordinates (x, y, z):

$$v_l = eu_l + fu_r + gv_r + h \tag{7}$$

The reader can derive equations similar to (4) and (5) when cameras are uncalibrated by following the scheme described in the previous section.

Consequently, with calibrated, oriented, and uncalibrated cameras, the *disparity space* can be used and it is *equivalent to recovering planar patch equations in the disparity space or in the real 3D space*. Of course, we focus on global reconstruction in this article, but the concept of *disparity space* is useful for all kinds of 3D reconstruction.

2.3 Affine Mapping

From the plane equation $au_l + bu_r + cv_l = d$ and with the epipolar constraint (7), we verify, as claimed in section 2.1, that the relative displacement of the matching points along the epipolar lines is an affine function of u_l, v_l when point M is in plane P and when the stereo geometry is rectified.

For each epipolar geometry, the equation of the affine mapping is:

- with calibrated cameras ($e = f = h = 0$, $g = 1$):

$$\begin{pmatrix} u_r \\ v_r \end{pmatrix} = \begin{pmatrix} -\frac{a}{b} & -\frac{c}{b} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_l \\ v_l \end{pmatrix} + \begin{pmatrix} \frac{d}{b} \\ 0 \end{pmatrix} \tag{8}$$

- with oriented cameras ($e = f = 0$):

$$\begin{pmatrix} u_r \\ v_r \end{pmatrix} = \begin{pmatrix} -\frac{a}{b} & -\frac{c}{b} \\ 0 & g \end{pmatrix} \begin{pmatrix} u_l \\ v_l \end{pmatrix} + \begin{pmatrix} \frac{d}{b} \\ h \end{pmatrix} \tag{9}$$

  where $g = \frac{\alpha_{vr}}{\alpha_{vl}}$, and $h = v_{0r} - gv_{0l}$,

- and with uncalibrated cameras:

$$\left.\begin{array}{c} au_l + bu_r + cv_l = d \\ v_l = eu_l + fu_r + gv_r + h \end{array}\right\} \tag{10}$$

  where e, f, g and h are only functions of the unknown calibration parameters.

The affine mapping equation (10) between left and right images is very useful since it allows, among other things, the global planar reconstruction when the epipolar geometry is known. In general, epipolar geometry defines the epipolar relation between the projections of a point in the left and right images. Epipolar geometry is in general described by the epipolar matrix (or the fundamental matrix). For rectified geometry, the epipolar relation is only determined by e, f, g and h.

From an unrectified stereo geometry, techniques exist to rectify a given stereogram if the cameras are calibrated [13, 10], or if only the epipolar geometry is known [14]. These kind of techniques can be applied preliminary to the proposed reconstruction algorithm. This extends its usefulness to unrectified geometry as shown in section 4.2. Moreover, as soon as the epipolar geometry is known, it is possible to map stereo images in a way where epipolar lines are raster lines corresponding in both images. But for a large set of stereo geometry, this rectification with a strong requirement on the epipolar geometry, has the drawback of creating rectified images that present a loss of pixel information along and between epipolar lines. It is why we are discussing three kinds of epipolar geometry (with calibrated, oriented, and uncalibrated cameras) where epipolar lines are less or not constrained to take care of all practical cases without loss of accuracy.

The coefficients e, f, g and h in the epipolar relation (7) are only functions of the calibration parameters, and not of plane equation P. Therefore, if the epipolar relation is not known before the reconstruction process, and a rectified stereo geometry is assumed, the epipolar relation can be estimated by using couples of left and right regions. Indeed, the epipolar relation applies for the centers of gravity $(\overline{u_l}, \overline{v_l})$ and $(\overline{u_r}, \overline{v_r})$:

$$\overline{v_l} = e\overline{u_l} + f\overline{u_r} + g\overline{v_r} + h$$

since the mapping is affine. For each couple of matched regions, this equation gives a linear constraint on the four parameters e, f, g and h. Consequently, four couples are enough to determinate the coefficients of the epipolar relation. But to achieve a better robustness, this computation has to be done on whole region couples with a least square criterion. This way of computing the epipolar relation is a region-based alternative to generic self-calibration techniques which is specific to rectified stereo geometry.

If we assume that the left and right images are rectified, no geometric camera calibration is needed for global planar patch reconstruction in the *disparity space*, as is derived in the next section. Planar patch reconstruction requires only the knowledge of the epipolar geometry. This fact is important because it means that the *disparity space* is the natural space where uncalibrated global reconstruction has to be done.

3 3D Plane Computation

Planar reconstruction consists in computing the 3D plane support ($P : au_l + bu_r + cv_l = d$ in the uncalibrated case, and $P' : a'x + b'y + c'z = d'$ in the real 3D space) of a patch from geometric attributes of its left and right projections.

In the proposed planar patch reconstruction scheme, we assume that two images of a stereo pair have been segmented into regions. Many segmentation methods are available [6], and we are currently using an algorithm based on region splitting [7]. Segmented regions are subsequently matched, using epipolar constraints and image intensity based characteristics. Geometric invariants are also used as described in section 5.1. The advantage of this approach comes from the fact that it is relatively easy to reliably extract and globally match extended 2D features, if the planar surfaces approximation holds, whereas obtaining exact contour-to-contour matching is much trickier [15]. Therefore we will not try to use the left to right matching equation (10) at a local level, but to integrate it over all the regions.

Starting from the plane equation P, we integrate this equation on the whole surface of left R_l and right R_r regions to obtain:

$$a\overline{u_l} + b\overline{u_r} + c\overline{v_l} = d \tag{11}$$

where $(\overline{u_l}, \overline{v_l})$ and $(\overline{u_r}, \overline{v_r})$ are the left and the right centers of gravity, respectively.

Thus, we can compute the plane distance d in terms of plane normal (a, b, c). Moreover, in the left and right image reference systems centered at $(\overline{u_l}, \overline{u_r})$ and $(\overline{u_r}, \overline{v_r})$ respectively, the affine mapping (10), is reduced to:

$$\left.\begin{array}{l} au_l + bu_r + cv_l = 0 \\ v_l = eu_l + fu_r + gv_r \end{array}\right\}$$

Given a region R, we define *centered* second order moments $\overline{uu}, \overline{uv}, \overline{vv}$ (with $(u, v) \in R$) as the inertia terms in the center of gravity coordinate system. These geometric moments are defined for each left and right region. Using second order moments leads to the following equations set:

$$\left.\begin{array}{rcrcrcl} a\ \overline{u_l v_l} & + & b\ \overline{u_r v_l} & + & c\ \overline{v_l v_l} & = & 0 \\ a\ \overline{u_l u_l} & + & b\ \overline{u_l u_r} & + & c\ \overline{u_l v_l} & = & 0 \\ a\ \overline{u_l u_r} & + & b\ \overline{u_r u_r} & + & c\ \overline{u_r v_l} & = & 0 \end{array}\right\} \tag{12}$$

But unknown second order *cross-moments* between the two matching regions R_l and R_r are introduced, such as:

$$\overline{u_l u_r} = \sum_{(u_l, v_l) \in R_l, (u_r, v_r) \in R_r} (u_l - \overline{u_l})(u_r - \overline{u_r})$$

To eliminate these cross-moments, we have to use the epipolar relation.

3.1 Calibrated or Oriented Cameras

In the center of gravity coordinate system, the epipolar relation is reduced to $v_l = gv_r$ with oriented cameras. Furthermore, in the calibrated camera case, we have $g = 1$. Therefore, we rewrite (12) by introducing this relation to eliminate several cross-moments:

$$\left.\begin{array}{rcrcrcl} a\ \overline{u_l v_l} & + & g\ b\ \overline{u_r v_r} & + & c\ \overline{v_l v_l} & = & 0 \\ a\ \overline{u_l u_l} & + & b\ \overline{u_l u_r} & + & c\ \overline{u_l v_l} & = & 0 \\ a\ \overline{u_l u_r} & + & b\ \overline{u_r u_r} & + & g\ c\ \overline{u_r v_r} & = & 0 \end{array}\right\} \tag{13}$$

In the previous equation, we still have one cross-moment $\overline{u_l u_r}$, which is unknown since we have not computed any pixel matching. It can be linearly eliminated by using the second and third equations only, and we get:

$$c = -\frac{a^2\overline{u_l u_l} - b^2\overline{u_r u_r}}{a\ \overline{u_l v_l} - g\ b\ \overline{u_r v_r}} \tag{14}$$

Introducing c into the first equation of system (13), we deduce:

$$\frac{a}{b} = \pm\sqrt{\frac{g^2\overline{u_r v_r}^2 - \overline{u_r u_r}\ \overline{v_l v_l}}{\overline{u_l v_l}^2 - \overline{u_l u_l}\ \overline{v_l v_l}}} \tag{15}$$

Since a normal vector (a, b, c) is being normalized, there are two solutions, but a negative $\frac{a}{b}$ ratio will be rejected as it corresponds to the situation where the two cameras would look at different sides of the 3D plane (see Figure 3), which is unlikely for real scenes.

With oriented cameras, the reconstruction is done in the *disparity space* up to an a priori unknown homographic transform. But if intrinsic camera parameters are available, the corresponding real 3D space plane equation is computed via equations (5). In particular for calibrated cameras, equations (5) giving the components of plane equation P' in the real 3D space are specialized to:

$$\left.\begin{array}{rcl} a' & = & a + b \\ b' & = & c \\ c' & = & -d \\ d' & = & b\ \delta \end{array}\right\}$$

3.2 Uncalibrated Cameras

For uncalibrated cameras, it is still possible to eliminate all the unknown cross-moments and compute the plane equation P in the disparity space, even though the equations are a more complex. After computations and simplifications, it turns out that $\frac{a}{b}$ ratio is one of the square roots of second degree polynomial $A\left(\frac{a}{b}\right)^2 + B\frac{a}{b} + C = 0$, where:

$$\begin{array}{rcl} A & = & (\overline{u_l v_l}^2 - \overline{u_l u_l}\,\overline{v_l v_l})(e^2\overline{u_l u_l} - 2e\overline{u_l v_l} + \overline{v_l v_l}) \\ B & = & -2e(\overline{u_l v_l}^2 - \overline{u_l u_l}\,\overline{v_l v_l})(g\,\overline{u_r v_r} + f\,\overline{u_r u_r}) \\ C & = & -\overline{v_l v_l}(g^2\overline{u_r v_r}^2 + 2fg\overline{u_r u_r}\,\overline{u_r v_r} + f^2\overline{u_r u_r}^2) + \overline{u_r u_r}(e^2\overline{u_l v_l}^2 - 2e\overline{v_l v_l}\,\overline{u_l v_l} + \overline{v_l v_l}^2) \end{array}$$

This equation extends (15) in the uncalibrated case. The extended version of (14) is:

$$c = \frac{a(\overline{u_l v_l} - e\,\overline{u_l u_l}) + b(f\,\overline{u_r u_r} + g\,\overline{u_r v_r})}{e\,\overline{u_l v_l} - \overline{v_l v_l}}$$

Therefore, with uncalibrated cameras, the reconstruction can be also done in the *disparity space* up to an unknown homographic transform.

In summary, with each stereo geometry, global geometric reconstruction is a very fast process: only equations (11), (14) and (15) (or equivalent equations in the uncalibrated case) have to be used to compute the plane support. However, we have assumed a scenario with no occlusions. So after presenting some reconstruction results, we shall deal with this limitation in the section 5.

4 Experimental Results

4.1 Reconstruction Accuracy

The suggested reconstruction method is first tested with synthetic data, where the whole environment is controlled. Synthetic scenes also make it possible to isolate errors produced by the reconstruction process, without interferences induced by the previous analysis processes: segmentation, matching and possibly calibration.

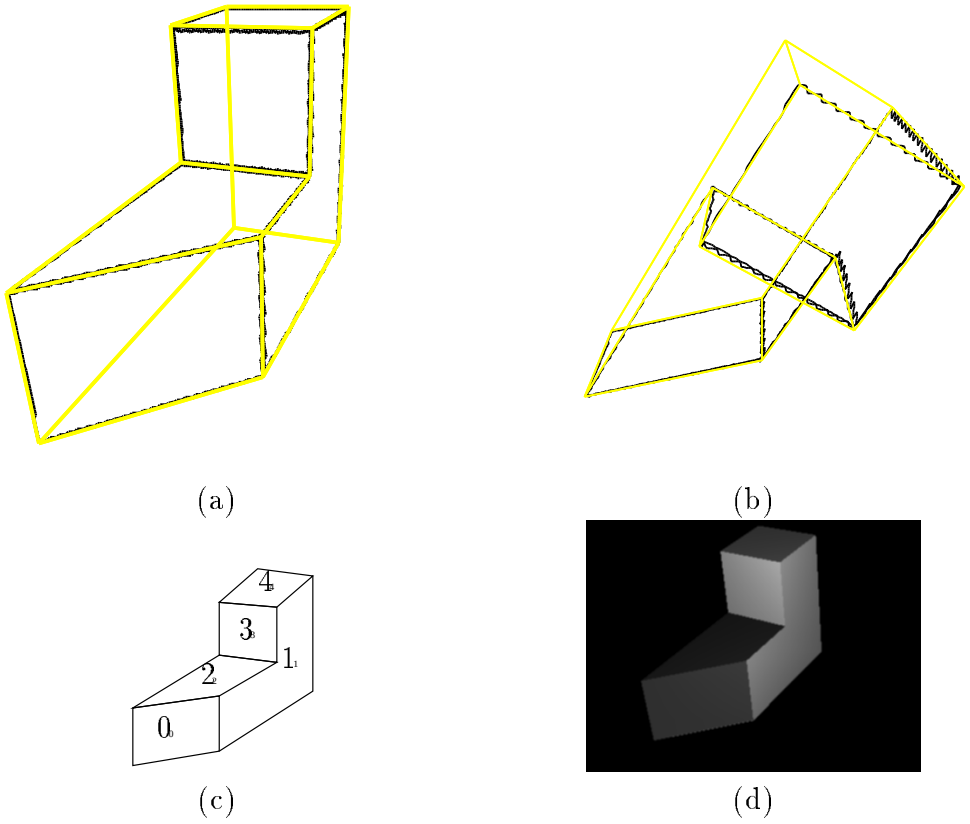


Figure 4: Front (a) and top (b) views of the synthetic original model and its reconstruction obtained from image (d). The reconstruction and the model are well superimposed. (c) shows patch numbers used in Table 1.

Facets pair	actual	estimated
0–1	135.0°	134.93°
0–2	90.0°	89.93°
0–3	45.0°	45.41°
0–4	90.0°	89.87°
1–2	90.0°	89.87°
1–3	90.0°	90.33°
1–4	90.0°	89.83°
2–3	90.0°	89.28°
2–4	0.0°	0.06°
3–4	90.0°	89.23°

Table 1: Actual normal angle of pair of patches, and estimations with the proposed reconstruction algorithm in rectified geometry. Facets numbers are shown in Figure 4(c).

We verify that geometric reconstruction gives accurate results. The geometric reconstruction obtained from the image of Figure 4(d), which is one of the two images in the original stereo pair, is shown in Figure 4(a)(b).

The orientations of the patches relative to one another are shown in Table 1, where the actual angles are compared with the values obtained from the geometry reconstruction. The average of the error of orientation is only 0.29°.

This experiment, along with others not presented here [16], shows that the proposed reconstruction is as accurate than a local sub-pixel edge-based technique [15], however no sub-pixel edges are extracted with the global approach. But in its present implementation, the algorithm does not produce the whole reconstruction, but only the closed planar patches.

4.2 Reconstruction Capabilities

Figure 5(d) shows the left image of an original stereogram of a calibration set-up, and Figure 5(a)(b)(c) are views of the reconstructed result. Notice that the algorithm allows the reconstruction of curved border planar patches. This property is a consequence of the use of geometric moments. Thus, we are not constrained by any specific representation (segments, algebraic curves, snakes, etc.) of region edges.

Figure 6(d) shows the left image of an original stereogram of a football, and Figure 6(a)(b)(c) the reconstructed result. Note that the method is not disturbed by the fact that the patches – hexagons and pentagons – are only approximately planar. This example illustrates the robustness of the approach.

Our reconstruction technique can be used in more general epipolar geometry than rectified one by using a rectification preprocessing. For example in Figure 7(a)(b), the angle between the two camera optical axis is 10°. After image rectification, a 3D Euclidean reconstruction of this scene consisting of two planes is obtained. The two planes geometry is correctly reconstructed as shown in Figure 7(c)(d).

The above experiments show that the algorithm presented here is suited for planar patch reconstruction, and furthermore that it allows a wide variety of data to be processed. With this said, we must acknowledge the critical dependence of planar patch reconstruction on segmentation preprocessing.

5 Checking Data Consistency

Before planar patch reconstruction, it is possible to test the relevance of its application to a given region couple.

5.1 Geometric Moments Invariants

In section 2, we described the affine mapping equation between projections in the right and left images of a point in a given plane. It follows that affine moment-based invariants exist between corresponding regions of a stereo pair [17, 18]. As in the previous sections, $\overline{u^p v^q}$ is the centered moment of order (p, q) in u and v, respectively. S is the surface of the region,

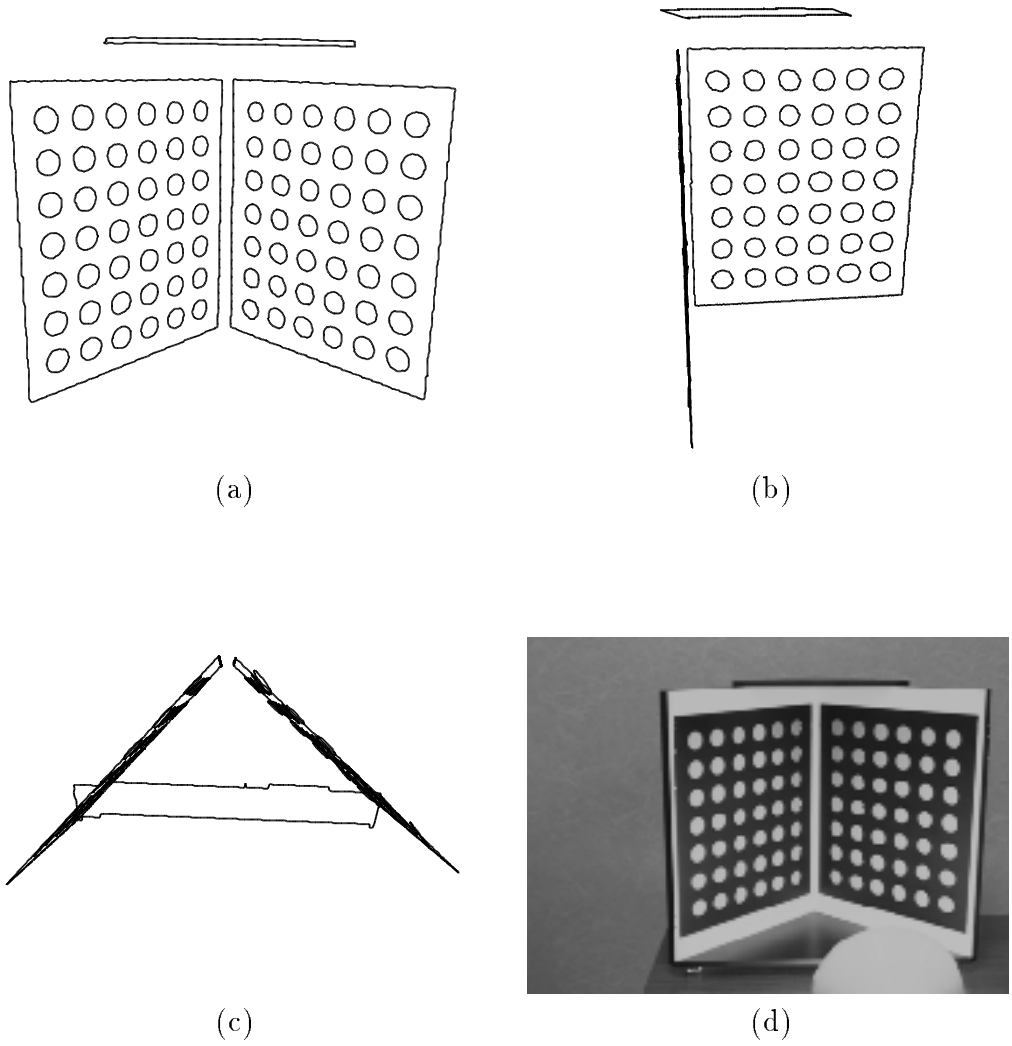
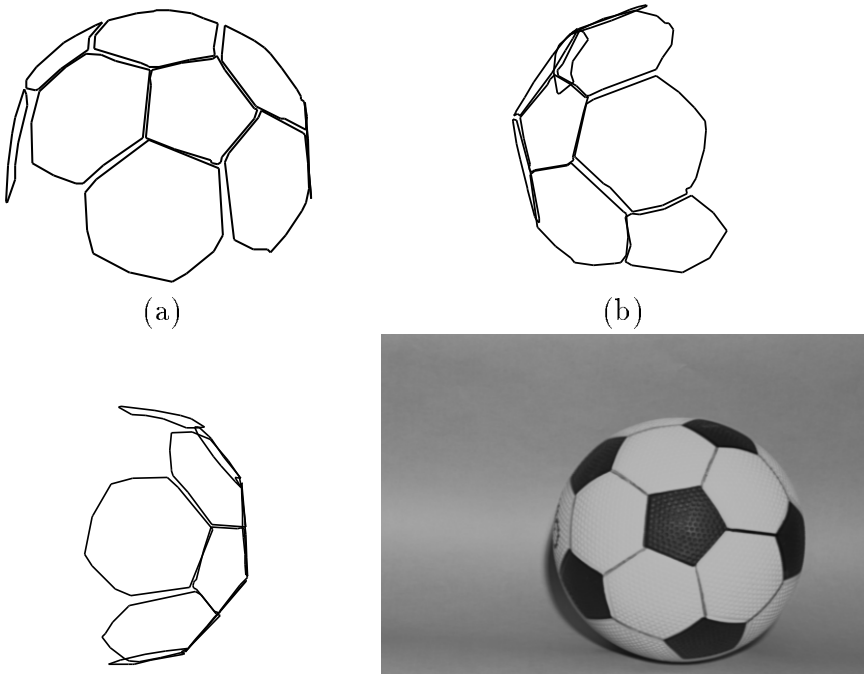


Figure 5: Front (a), side (b) and top (c) views of a calibration setup reconstructed from the original image (d) (right image of the pair not shown).



(a) (b) (c) (d)

Figure 6: Front (a), side (b) and top (c) views of the soccer ball reconstructed from the original image (d) (right image of the pair not shown).

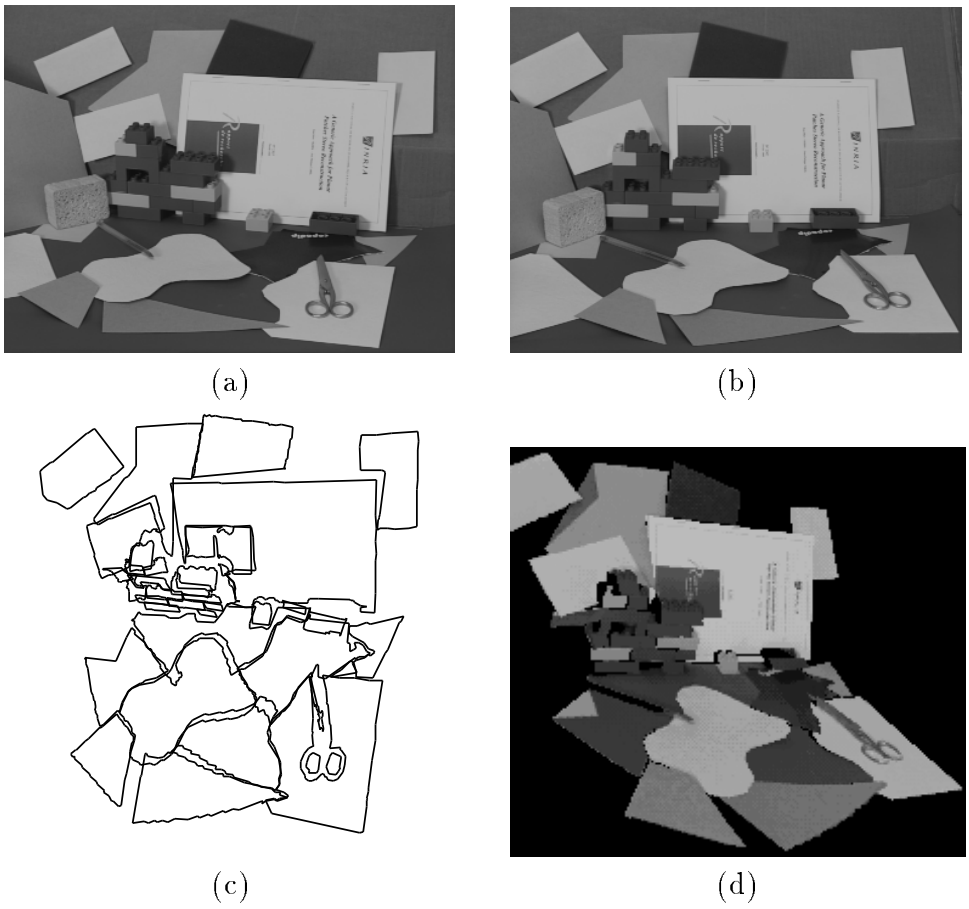


Figure 7: Left (a) and right (b) images of the original stereogram before rectification. Resulting reconstruction viewed as wireframes (c) and with left image texture in another viewpoint (d).

i.e the $(0,0)$ order moment. We will limit our attention to the three fundamental invariants of order no higher than 3, namely:

$$I_1 = (\overline{u^2}\ \overline{v^2} - \overline{uv}^2)/S^4 \quad (16)$$

$$I_2 = (\overline{u^3}^2\overline{v^3}^2 - 6\overline{u^3}\ \overline{u^2v}\ \overline{uv^2}\ \overline{v^3} + 4\overline{u^3}\ \overline{uv^2}^3 + 4\overline{u^2v}^3\overline{v^3} - 3\overline{u^2v}^2\overline{uv^2}^2)/S^{10} \quad (17)$$

$$I_3 = (\overline{u^2}(\overline{u^2v}\ \overline{v^3} - \overline{uv^2}^2) - \overline{uv}(\overline{u^3}\ \overline{v^3} - \overline{u^2v}\ \overline{uv^2}) + \overline{v^2}(\overline{u^3}\ \overline{uv^2} - \overline{u^2v}^2))/S^7 \quad (18)$$

Consequently, before starting the reconstruction process, the computation of invariants allows the validity of the planar hypothesis to be tested, i.e. to check if the shape of a given match of regions is consistent with the planar patch projection hypothesis. Any violation of the hypothesis (due to segmentation errors, occlusions, or other unidentified causes) can produce a variation between the estimates of I_1, I_2 and I_3 computed from the images, and can be used as a discrimination tool. To determine the efficiency of this consistency check, it is necessary to investigate the stability of moment-based invariants versus noise or segmentation error due to viewpoint changes. For example, it is clear that the higher the order, the more sensitive the invariant is. A systematic study using known synthetic data is presented in the next section.

5.2 Robustness to Occlusion

We have chosen to test the robustness of the proposed algorithms on known synthetic data, consisting of a succession of 8 stereo images of two planar patches, with increasing amounts of occlusion due to viewpoint changes. Figure 8(b) shows the maximum occlusion case.

Moreover, Figure 8 shows the error of orientation of the occluded patch as a function of occlusion. We have plotted the occlusion-free experiment with same viewpoints for comparison.

Experimentally, it has been shown that the bias produced by occlusions on the 3D patch orientation and pose is reasonable – about 6° for the data of Figure 8, and about the same magnitude as noise and segmentation errors due to viewpoint changes. For more detail about this test, see the complete study in [5].

5.3 Validating the Reconstruction using Invariants

We have utilized the previous scene to test the use of invariants defined in section 5.1. First, we have removed the occluding plane and plotted the 3 invariants in the different views of the sequence. Invariants of the first image are taken as references.

It is well known that moments become more susceptible to noise as their order increases: in Figure 10, I_2 and I_3 clearly show insufficient stability to viewpoint changes. We have therefore selected I_1 as the most robust invariant. Figure 9 confirms this conclusion: I_2 and I_3 are very sensitive to small occlusion errors, whereas I_1 degrades nicely.

In this example, region couples with an invariant ratio out of the interval $[0.96, 1.04]$ correspond to couples with more than 10% of their area occluded. Other experiments corroborate this result, and region couples displaying more than 5% invariant disparity should be considered unreliable. Figure 11 shows an example of occluded planar patches selection by using second order moment invariant I_1.

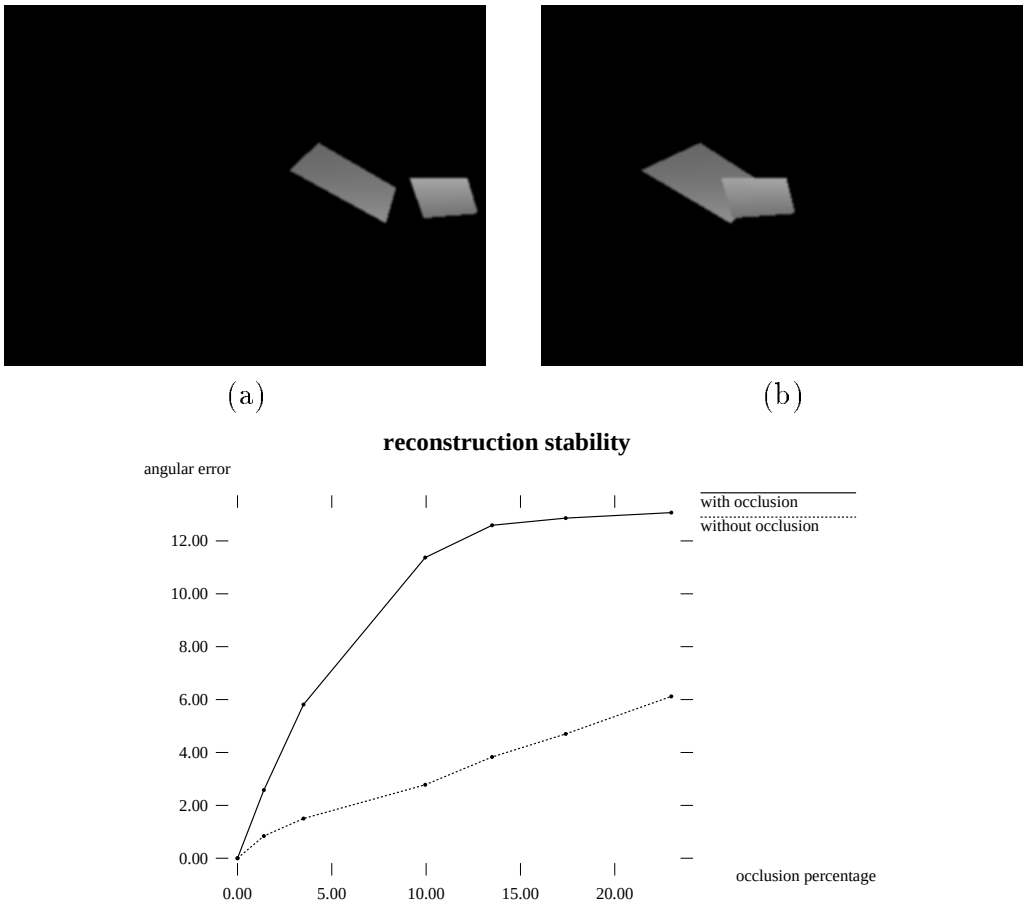


Figure 8: A stereogram (a)(b) of the occlusion test sequence, and the reconstruction stability result. Different occlusion percentage are obtained as a function of successive viewpoints of the camera rotating around the test scene.

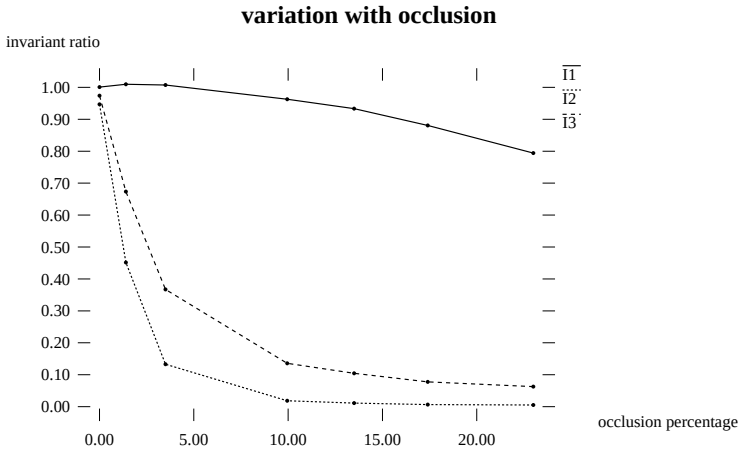


Figure 9: Ratio of the left and right invariants I_1, I_2 and I_3 computed for different values of the occlusion.

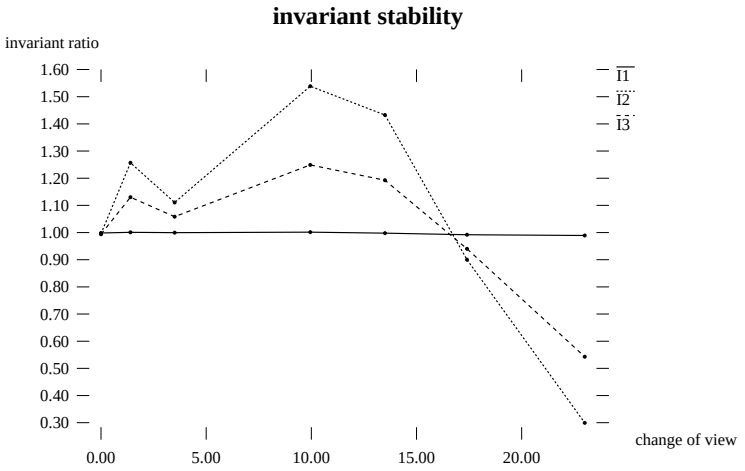


Figure 10: Ratio of the left and right invariants I_1, I_2 and I_3 computed for the successive viewpoints as in Figure 8.

6 View Morphing

In this section, we illustrate advantages of reconstructing planar patches and using disparity space to compute new views of the scene for computer graphic purposes.

6.1 Calibrated Cameras

Since the algorithm reconstructs surface patches and not 1D features, visualizations better than wireframes like those in Figures 5 or 6 can be obtained. In particular, the left image can be retroprojected on the planar patch reconstruction with a retroprojecting texture mapping technique. This technique is usual in computer graphics, and the free software Rayshade [19] provides its fast implementation. Synthetic viewpoints of Figure 12 are generated from a stereo pair (Figure 12(d) is the left original image). Another example is Figure 7(d). The relevance and the accuracy of the reconstruction method is also illustrated by these figures.

6.2 Uncalibrated Cameras and Rectified Stereo Geometry

With calibrated cameras, synthetic views in all possible viewpoints can be generated, but a projective texture mapping is needed. By taking advantage of the *disparity space*, only an orthographic projection is necessary to yield textured reconstructions with uncalibrated cameras.

It was noticed [20] that new views between left and right images can be produced easily, arguing that corresponding pixels are linearly interpolated with parallel viewpoints. Indeed, any parallel viewpoint which is intermediate to the left and right viewpoints is also a rectified viewpoint. we can go further by using the disparity space: parallel intermediate views can be generated as an orthogonal projection on an image plane which is rotated between the left and right image planes. Figure 13 illustrates this property with oriented cameras, with

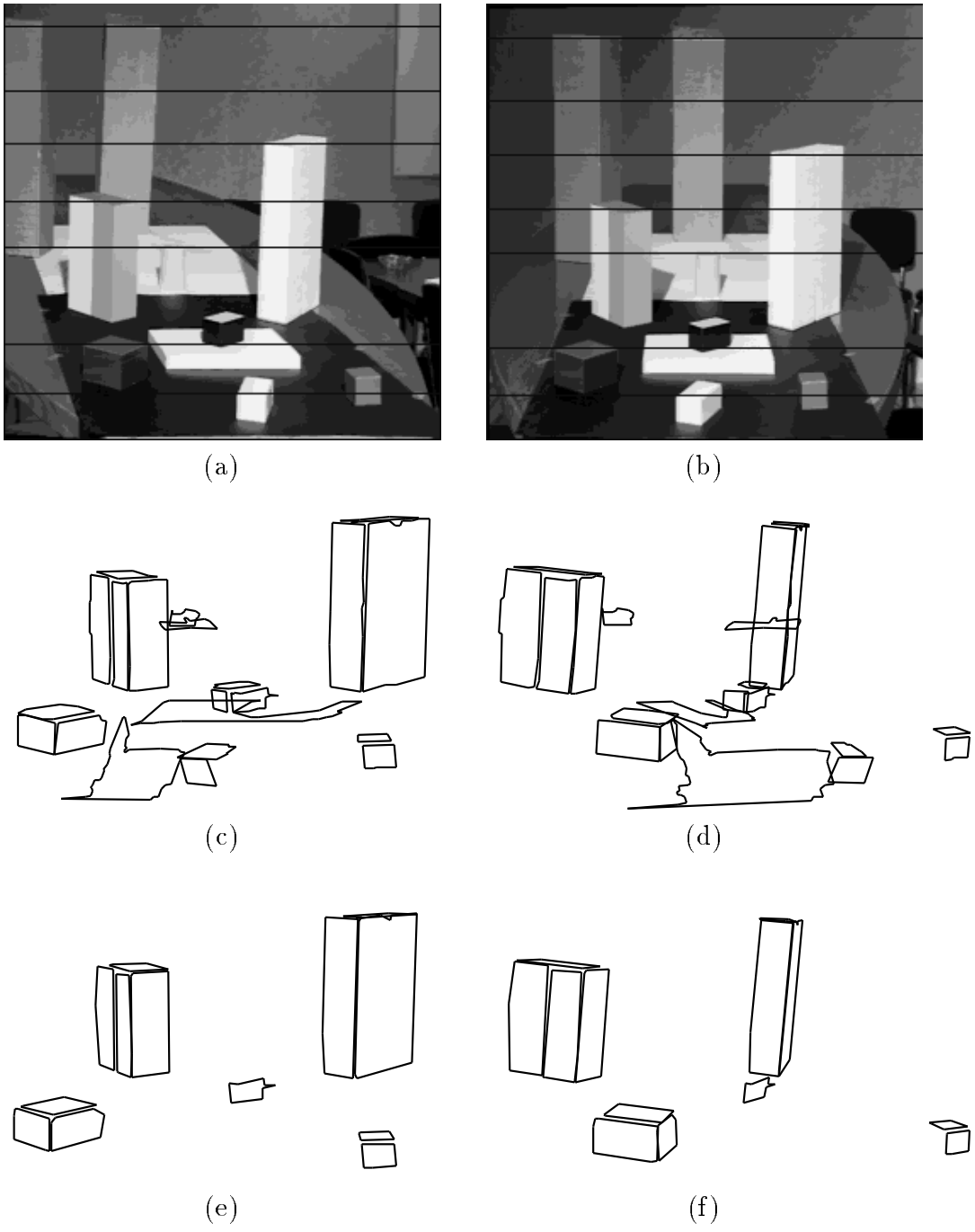


Figure 11: (a) and (b) is the original stereogram. Front (c) and side (d) views of the reconstruction before selection, and (e) and (f) are the same views after selection.

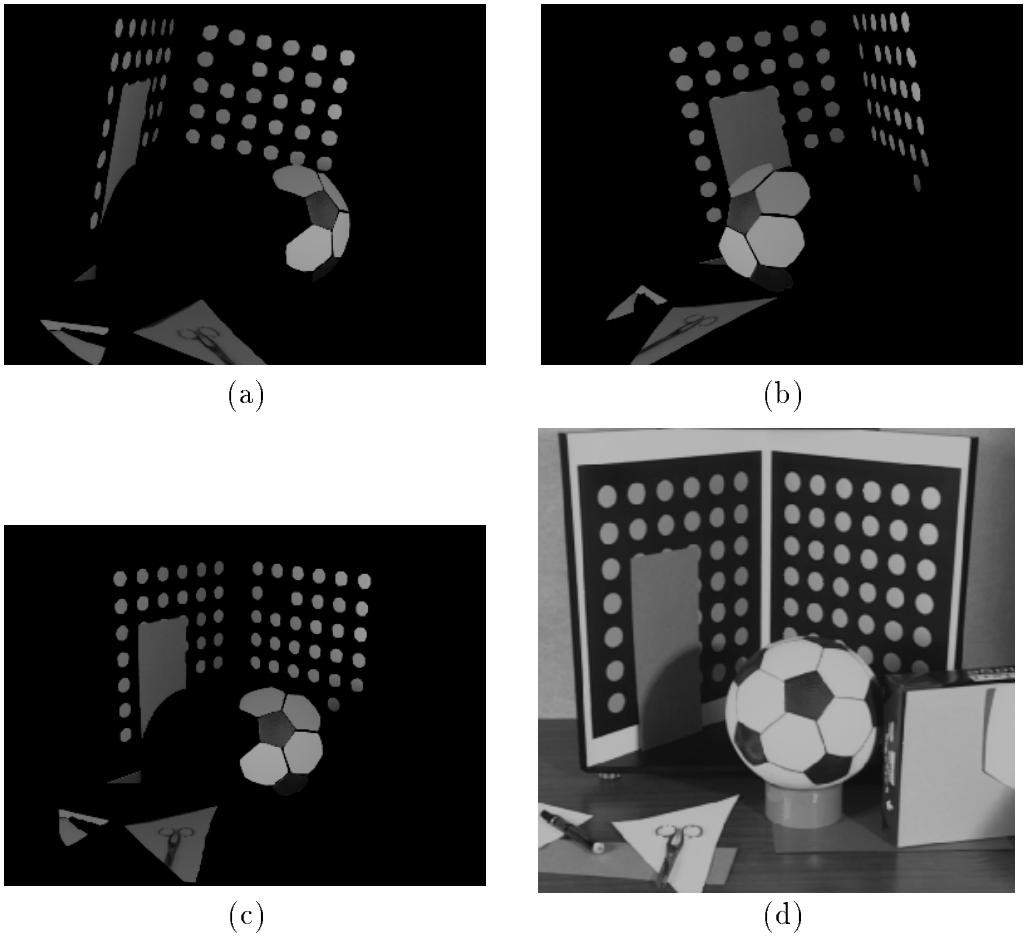


Figure 12: Left side (a), right side (b) and front (c) views of a reconstruction where original image (d) texture is retroprojected on all reconstructed planar patches.

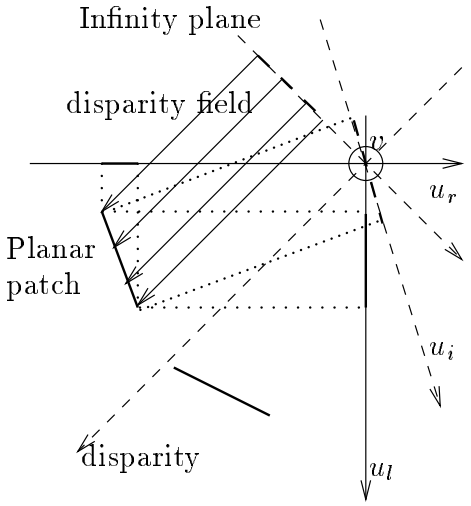


Figure 13: 2D cross section of the disparity space which shows how to obtain the disparity field of a reconstructed planar patch with intermediate views.

a 2D cross section of the disparity space. In this figure, the (u_i, v) plane is an example of an intermediate image plane which can be used to generate easily an intermediate view of the observed scene. Notice that by considering carefully Figure 13, one can perform reconstruction tasks from parallel multiple views in the disparity space.

A particular advantage of using regions based, rather than line or point based, techniques is that it is easy to constrain a part known as planar to be perfectly planar in the reconstruction. The disadvantage of the view interpolation in the disparity space is that the interpolation is exact if and only if the viewpoint is on the line defined by the camera centers C_l and C_r and the image plane of the synthetic camera is parallel to the left and right ones.

In practice, intermediate views are generated with a simple linear interpolation of the affine mapping (10) for each planar patch. This fact allows the synthesis of new views of the reconstructed scene such as in Figure 14(f). Holes between textured regions are fill in by image interpolation between the reconstructed regions. This step achieves realistic view synthesis. An example of interpolation (between left and right images) and extrapolation around the right image of Botticelli's Venus is shown in Figure 14(c) and (d) respectively. These two images are extracted from a short image sequence of the face rotating available at `http://www-syntim.inria.fr/~tarel/`.

As shown in Figure 13, by a simple axis change in the disparity space (alignment with first and second bisectors) it is possible to compute the disparity map of the reconstructed scene. When the surface of the scene is smooth (a human face for example), holes between regions are fill in by interpolation to generate Figure 14(e).

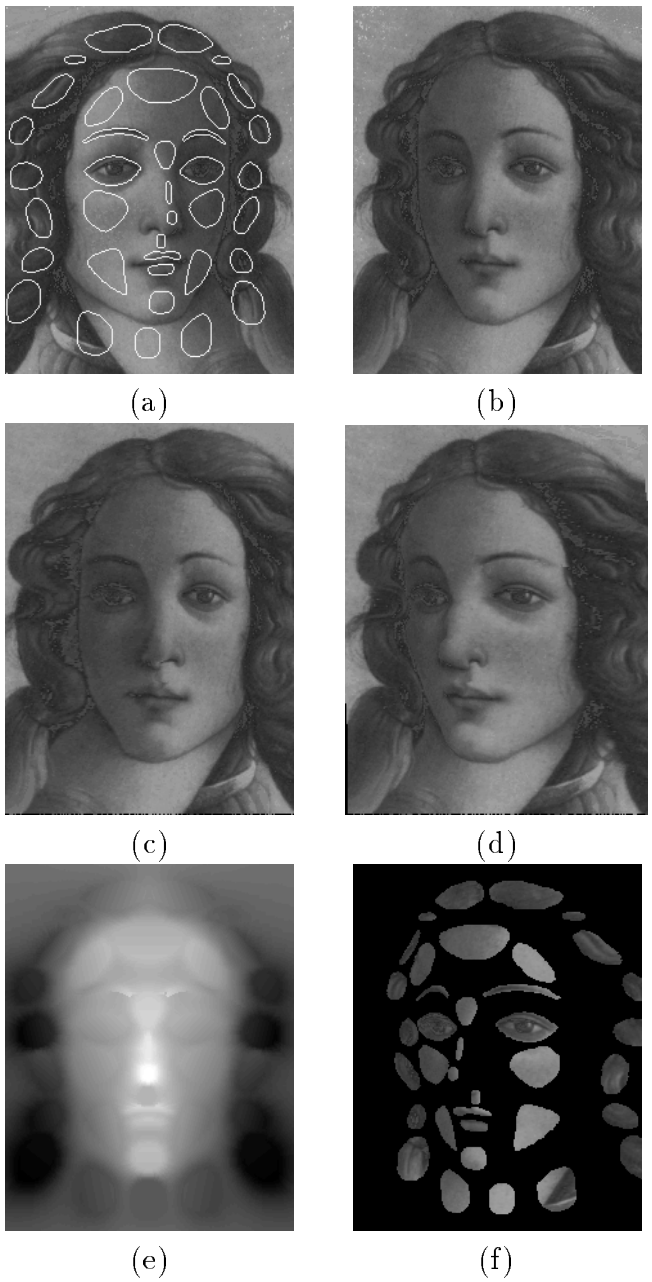


Figure 14: (a)(b) is the uncalibrated stereogram. (b) is the original image extracted from Botticelli's painting "The Birth of Venus". (a) is its mirror reflection with a hand-made segmentation superimposed. (c) is a synthetic face view intermediate to the left and right ones. (d) is an extrapolated view to the right. (f) is the same view as (d) without interpolation between planar patches. (e) is the disparity map associated to (c).

7 Conclusion

The work presented currently relies on epipolar geometry estimation, and region segmentation/matching preprocessing.

We have proposed a generic framework for the 3D reconstruction of planar surfaces from their image projections on a rectified stereogram with or without calibrated cameras. The approach, which does not rely on any local feature matching, makes use of geometric global attributes, and is relatively robust to segmentation errors and occlusions. The use of photometric attributes is investigated in [21]. It is possible to extend this work to handle not only planar patches but also open planar curves, or data sets of 3D coplanar points. In this framework, future work will deal with global reconstruction from several views.

A new method for reconstruction has been proposed, along with a method to check for its self-consistency. Reconstructions have been obtained in the *disparity space*, but the Euclidean structure is also obtained if explicit calibration parameters are available.

Finally, we have demonstrated how the use of the *disparity space* can simplify and clarify the reconstruction process of planar patches by taking advantage of the homographic transform between the 3D real space and the *disparity space*. Moreover, a linear view morphing technique based on planar patches is proposed for computer graphic applications.

Acknowledgments

The author is grateful to Jean-Marc Vézien for helpful discussions and for his contribution in making some of the figures possible through experimentation.

Author

Jean-Philippe Tarel graduated from the École Nationale des Ponts et Chaussées, Paris, France (1991). He received his PhD degree in Applied Mathematics from Paris IX-Dauphine University in 1996 and he was with the Institut National de Recherche en Informatique et Automatique (INRIA) from 1991 to 1996. Since 1997, he has been a research associate at the Brown University, USA. His research interests include computer vision, computer graphics and pattern recognition.

References

[1] D. B. Cooper, Y. P. Hung, and G. Taubin. A new model-based stereo approach for 3D surface reconstruction using contours on the surface pattern. In *Second Interna-*

tional Conference on Computer Vision (Tampa,, FL, December 5–8, 1988), pages 74–83, Washington, DC, 1988. Computer Society Press.

[2] Steen Kristensen, Henrik Nielsen, and Henrik Christensen. Cooperative depth extraction. In *Proceedings, 8th Scandinavian Conference on Image Analysis (SCIA)*, pages 321–328, Tromsø, Norway, May 1993.

[3] Y. P. Hung D. B. Cooper, J. Subrahmonia and B. Cernuschi-Frias. *Markov Random Fields: Theory and Application*, chapter Use of Markov Random Fields in Estimating and Recognizing Objects in 3D Space, pages 335–367. Academic Press, 1992.

[4] John Aloimonos and Jean-Yves Hervé. Correspondenceless stereo and motion: Planar patches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):504–510, May 1990.

[5] J.-M. Vézien. *Techniques de reconstruction globale par analyse de paires d'images stéréoscopiques.* PhD thesis, Université Paris-VII, 1995.

[6] N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 29(9):1277–1294, 1993.

[7] S. Randriamasy and A. Gagalowicz. Region based stereo matching oriented image processing. In *CVPR'91 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, HI, June 3-6, 1991)*, Washington, DC., June 1991. Computer Society Press.

[8] O. D. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2:485–508, 1988.

[9] R. Mohr. *Handbook of Pattern Recognition and Computer Vision*, chapter Projective Geometry and Computer Vision, pages 369–393. World Scientific, 1993.

[10] N. Ayache. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception.* MIT Press, 1991.

[11] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–198, 1985.

[12] S. D. Cochran and G. Medioni. 3-D surface description from binocular stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(10):981–994, October 1992.

[13] A. Fusiello, E. Trucco, and A. Verri. Rectification with unconstrained stereo geometry. In A. F. Clark, editor, *Proceedings of the British Machine Vision Conference*, pages 400–409. BMVA Press, September 1997.

[14] L. Robert, C. Zeller, O. Faugeras, and M. Hébert. *Applications of non-metric vision to some visually-guided robotics tasks*, chapter 5, pages 89–134. Lawrence Erlbaum Associates, 1997.

[15] B. Serra and M. Berthod. 3D model localization using high-resolution reconstruction of monocular image sequences. *IEEE Transactions of Image Processing*, 6(1):175–188, 1997.

[16] Jean-Philippe Tarel. *Estimation géométrique et appariement en modélisation automatique*. PhD thesis, Université Paris-IX, 1996.

[17] A. Sluzek. Using moment invariants to recognize and locate partially occluded 2D objects. *Pattern Recognition Letters*, 7:253–257, 1988.

[18] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167–174, 1993.

[19] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practices (2nd Edition)*. Addison Wesley, 1990.

[20] Steven M. Seitz and Clarles R. Dyer. View morphing. In Addison Wesley, editor, *Computer Graphics (ACM SIGGRAPH '96 Proceedings)*, volume 23(3), pages 21–30, August 1996.

[21] Jean-Philippe Tarel and Jean-Marc Vézien. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandinavian Conference on Image Analysis*, volume 2, pages 1061–1070, Uppsala, Sweden, 1995. Swedish Society for Automated Image Analysis.