A Generic Approach for Planar Patches Stereo Reconstruction

Jean-Marc Vezien, Jean-Philippe Tarel

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Generic Approach for Planar Patches Stereo Reconstruction

Jean-Marc Vezien, Jean-Philippe Tarel

N° 2507

February 1995

PROGRAMME 4

ISSN 0249-6399

A Generic Approach for Planar Patches Stereo Reconstruction

Jean-Marc Vezien [*], Jean-Philippe Tarel [**]

Programme 4 — Robotique, image et vision
Projet Syntim

Rapport de recherche n˚2507 — February 1995 — 25 pages

Abstract: Whereas edge-based stereo has traditionally received a lot of attention, relatively few explicit region-based 3D reconstruction techniques have been developed, despite the significant advantages such global features present for representing indoor or urban environments. We propose a general framework for 3D reconstruction of planar patches from stereoscopic pairs of images, and explicit a generic planar equation recovery scheme, which can then be applied to several types of stereosystem geometry. Another advantage of the formulation is its possible use on various cues: we propose three different applications, using moments of inertia, parametric and non-parametric photometric analysis. In each case, the robustness of the results is assessed, both with analytic data consistency check and tests performed on synthetic stereograms. Experiments on real data are also presented.

Key-words: Planar Assumption, Stereo Reconstruction, Rectification, Inertia Moments, Autocorrelation.

(Résumé : tsvp)

[*] E-mail : Jean-Marc.Vezien@inria.fr
[**] E-mail : Jean-Philippe.Tarel@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Une approche générique pour la reconstruction de facettes 3D

Résumé : Alors qu'il existe de nombreux travaux en reconstruction 3D de contours en stéréovision, peu de techniques utilisent des régions. Ces primitives présentent pourtant d'importants avantages pour représenter des scènes d'intérieur ou d'extérieur en environnement urbain. Nous proposons une approche générique pour réaliser la reconstruction 3D de facettes planes à partir de paires d'images stéréoscopiques et nous explicitons les équations qui permettent de retrouver le plan de la facette pour différents types de configurations du système binoculaire. Un avantage de cette formulation est la possibilité d'utiliser différents type de caractéristiques des régions : nous proposons trois applications utilisant soit les moments d'inertie, soit une représentation paramétrées ou non de la photométrie. Dans chaque cas, la robustesse des résultats est vérifiée, en contrôlant analytiquement leur cohérence et en réalisant des tests sur des paires stéréoscopiques de synthèse. Des résultats sur des images réelles sont aussi présentés.

Mots-clé : Hypothèse Planaire, Reconstruction Stéréoscopique, Rectification, Moments d'inertie, Autocorrelation.

1 Introduction

Since the introduction of early stereo technics and paradigms [Mar82], steady efforts were made to robustly recover depth from stereo. A lot of work was dedicated to the classical goal of robot autonomous navigation, but new applications in scene analysis for virtual environment construction or video-conference are calling a growing interest. Such topics 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 (e.g. corners) or segments.

The planar surfaces assumption, which holds reasonably well in indoor static scenes (as long as humans and plants have been removed) has received most of the attention so far. Aloimonos [Alo90] [AH90] presented moment-based relationships between the two projections of a planar region in a calibrated stereoscopic pair, but used them in the derivation of invariants instead of explicit 3D reconstruction. The planar assumption is often used as a constraint for other techniques, such as stereo contour reconstruction [CHT88] or MRF field-based correlation [SHC90]. Kristensen [KNC93] proposes to use it in collaboration with corner and line detectors, but relies on a depth from focus process to obtain robust reconstructions. It is worth noting that most works rely on the assumption that the stereo geometry is rectified (scanlines correspond to each other in both images), a condition which may not be readily available on a given vision system.

We propose a generic scheme for the 3D reconstruction of planar patches from a single stereo pair. It does *not* require any additional localized information. The method assumes that region segmentation has been performed on both images, as well as the matching between 2D primitives, with methods such as described in [AMG93] and [RG91].

Our procedure fundamentally relies on the linearization of the pixel matching relationship around a reference point of each planar facet, typically the center of inertia. In section 1, we show that such approximations become exact for a rectified system, and allow for simple computations in a transformed coordinate system that we call the *disparity space*. In the case of a general transform between images, we use a paraperspective model to approximate the perspective projection (each 3D surface is shallow with respect to the distance scene-observer). A locally affine transformation between the coordinates of the two projections of a 3D point lying on a planar patch is then derived.

Section 2 is dedicated to using the obtained equations to find the plane parameters which gives the best *global* correspondence between the pair of projected regions. Our emphasis is on avoiding any local feature matching. We propose three different schemes: the first one is solely based on the regions shape and uses centered inertia-moments. The second is based on an uniform albedo assumption only, and thus ignores the shape of the region. The third takes into account both geometry and photometry by using correlation maps computed on the regions.

Many authors have pointed out the main problem with region-based stereo algorithms, which is how to deal with occlusions. Segmentation is often prone to errors as well. Yet very few attempts have been made to detect, take into account or even quantify these effects precisely. In section 3, we therefore propose several ways of checking the coherency of the stereo data which enable to decide whether the reconstruction is valid or not. When using geometrical cues, affine invariants computed on the matched regions prove to be robust enough to detect unreliable data (i.e. if the regions geometry is *inconsistent* with the hypothesis of the projection of a planar patch). The photometric hypothesis of constant albedo can be tested against the image data as well. Finally, a 3D facet can also still be rejected if both tests are positive but the reconstruction methods give inconsistent results.

2 Problem Formulation

Let us consider the set-up of figure 1, where a stereo system (calibrated or not) is looking at a planar patches-made world. One can establish the equation linking the two projections m^r and m^l of a given 3D point M lying on a plane $P : Z^l = p^l X^l + q^l Y^l + c^l$ (or $Z^r = p^r X^r + q^r Y^r + c^r$), as a function of the plane parameters. It can be easily shown that, if one denotes the displacement between the cameras as the composition of rotation R and translation T:

$$\begin{pmatrix} x^l \\ y^l \end{pmatrix} = \frac{1}{(\frac{T_z}{c^r} + R_{33}) + (R_{31} - p^r\frac{T_z}{c^r})x^r + (R_{32} - q^r\frac{T_z}{c^r})y^r} \left[\begin{pmatrix} R_{11} - p^r\frac{T_x}{c^r} & R_{12} - q^r\frac{T_x}{c^r} \\ R_{21} - p^r\frac{T_y}{c^r} & R_{22} - q^r\frac{T_y}{c^r} \end{pmatrix} \begin{pmatrix} x^r \\ y^r \end{pmatrix} + \begin{pmatrix} R_{13} + \frac{T_x}{c^r} \\ R_{23} + \frac{T_y}{c^r} \end{pmatrix} \right] \quad (1)$$

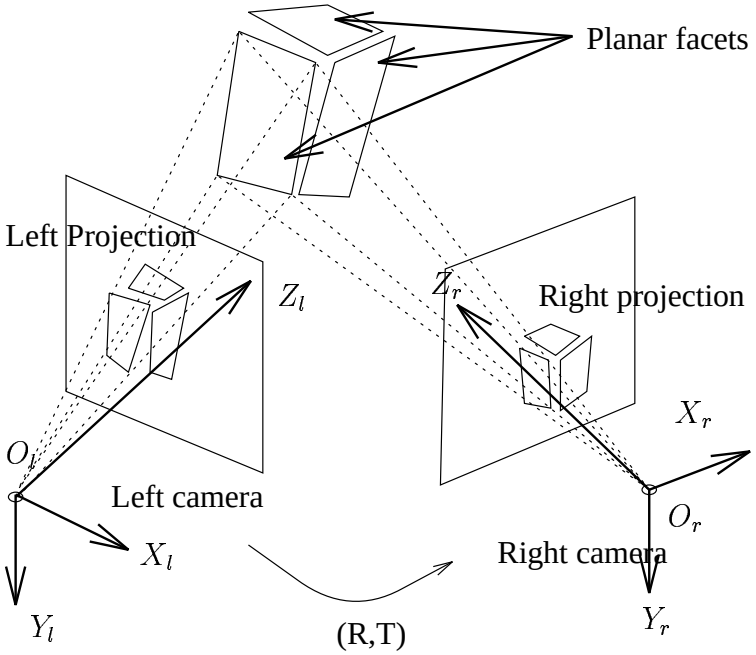

Figure 1: 3D stereo reconstruction of planar patches.

This relationship is nonlinear and as such of little practical use. Nevertheless, it can be drastically simplified in the case of a *rectified* stereo system, where $T = (T_x, 0, 0)^t$ and $R = Id$. In that case, $y^l = y^r = y$, $p^r = p^l = p$, $q^r = q^l = q$ and:

$$\begin{pmatrix} x^l \\ y^l \end{pmatrix} = \begin{pmatrix} x^r + \frac{T_x}{c^r}(1 - px^r - qy) \\ y \end{pmatrix} \tag{2}$$

In other words, epipolar lines are rasterlines, and the relative displacement of the matching points along the epipolar lines is a linear function of x, y, p and q. At this point, it is convenient to map the Cartesian 3D space (X, Y, Z) into the coordinate space (x^r, x^l, y) which we call *disparity space*, and in which the 3D plane $P : Z^r = pX^r + qY^r + c^r$ is transformed into $P' : ex^l + fx^r + gy + h = 0$ (the mapping is homographic), with:

$$\begin{cases} p & = & -\frac{e+f}{h} \\ q & = & -\frac{g}{h} \\ c^r & = & -T_x\frac{e}{h} \end{cases} \tag{3}$$

In the rectified case, it is thus **equivalent to recover planar patches equations in the disparity space or the real 3D space**.

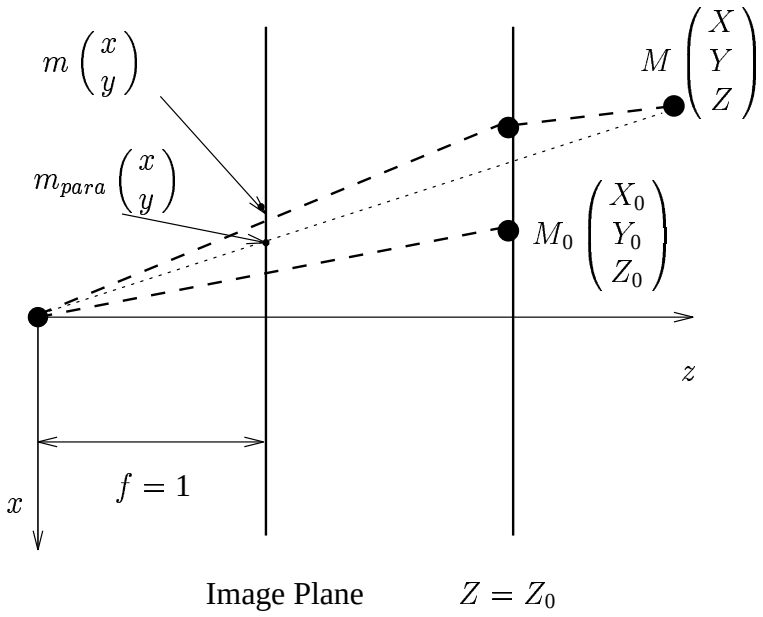

Figure 2: Paraperspective projection. M_0 is the reference point (see text).

In the general case, a local linearization of (1) can still be achieved by approximating the true projection by a paraperspective model, defined by figure 2:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{Z_0}\begin{pmatrix} X \\ Y \end{pmatrix} - \frac{Z - Z_0}{Z_0^2}\begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} \tag{4}$$

where $M_0 = (X_0, Y_0, Z_0)^t$ is a reference point. This model relies on two hypotheses:

- the projective projection $\frac{1}{Z}$ is approximated by the first term of its Taylor series around Z_0. Therefore, one can only apply it to points such that $Z - Z_0$

is small with respect to Z_0. This is sometimes refered as the *shallow object* hypothesis, which holds well for indoor scenes or satellite imagery,

- the object X and Y dimensions are smaller than its depth.

Under these two conditions, the approximation holds, and one can then derive equation 5 as an approximation of (1).

$$\begin{pmatrix} x^l \\ y^l \end{pmatrix} = \frac{Z_0^r}{Z_0^l} \begin{pmatrix} (R_{11} - \frac{T_x}{Z_0^l} R_{31}) - R_{33} \frac{Z_0^r}{Z_0^l} T_x \frac{p^r}{c^r} & (R_{12} - \frac{T_x}{Z_0^l} R_{32}) - R_{33} \frac{Z_0^r}{Z_0^l} T_x \frac{q^r}{c^r} \\ (R_{21} - \frac{T_y}{Z_0^l} R_{31}) - R_{33} \frac{Z_0^r}{Z_0^l} T_y \frac{p^r}{c^r} & (R_{22} - \frac{T_y}{Z_0^l} R_{32}) - R_{33} \frac{Z_0^r}{Z_0^l} T_y \frac{q^r}{c^r} \end{pmatrix} \begin{pmatrix} x^r \\ y^r \end{pmatrix} + \frac{Z_0^r}{Z_0^l} \begin{pmatrix} R_{13} \\ R_{23} \end{pmatrix} + \left(\frac{1}{Z_0^l} \left(2 - R_{33} \frac{Z_0^r}{Z_0^l} \right) + \left(\frac{Z_0^r}{Z_0^l} \right)^2 R_{33} \left(\frac{p^r}{c^r} x_0^r + \frac{q^r}{c^r} y_0^r \right) \right) \begin{pmatrix} T_x \\ T_y \end{pmatrix} \tag{5}$$

Again, we obtain an affine equation, with a linear dependence in the plane coefficients $\frac{p}{c}$ and $\frac{q}{c}$. For the choice of M_0 see next section.

Starting from equations (2) and (5), we will now derive global relationships between the two projections of a 3D planar patch in the images of a stereo pair, and use them to recover the position parameters (p, q, c).

3 3D Plane Computation

For our reconstruction scheme, we assume that two images of a stereo pair have been segmented into regions. A lot of segmentation methods exist [PP93], and we are currently using different algorithms developed by our team, based on region growing [GM86], region splitting [RG91] or energy minimization [AMG93]. Segmented regions are subsequently matched, using epipolar constraints and image intensity based characteristics [VSCG89, RG91]. 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. Therefore we will not try to use matching equations (2) and (5) at a local level, but to integrate them over the whole regions. In this section, we propose three different techniques, directly derived

from the planar-constrained correspondence equations, which rely on geometric *and* photometric cues, along with additional assumptions about the content of the scene. Results are shown in section 5.

3.1 Geometry-Based Reconstruction

Given a region R (a set of N_R connected pixels in an image), (i, j) two integers of $\mathbb{N}$, such as $i + j = n$, the inertia moment of order n L_{ij} is defined as:

$$L_{ij} = \sum_{m(x,y)\in R} x^i y^j \tag{6}$$

In particular $L_{00} = N_R$, and the center of inertia of R is $\vec{\mu_R} = (x_\mu, y_\mu)^t = \frac{1}{N_R}(L_{10}, L_{01})^t$. We will use a more intrinsic representation, the *centered* moments:

$$C_{ij} = \sum_{m(x,y)\in R} (x - L_{10})^i (y - L_{01})^j \tag{7}$$

In this section, we use only moments of order 2 or lower.

- **Rectified geometry:**

Let us define the *cross-moment* between the two matching regions R^l and R^r as:

$$\sigma^{rl} = \sum_{m(x^l,y^l)\in R^l} (x^l - x^l_\mu)(x^r - x^r_\mu) \tag{8}$$

Starting from the disparity plane equation $ex^l + fx^r + gy + h = 0$, using equation (7) leads to the set:

$$\begin{cases} eC^l_{11} + fC^r_{11} + gC_{02} = 0 \\ eC^l_{20} + f\sigma^{rl} + gC^l_{11} = 0 \\ e\sigma^{rl} + fC^r_{20} + gC^r_{11} = 0 \end{cases} \tag{9}$$

Eliminating σ^{rl} (we don't compute any pixel matching), one gets:

$$\frac{e}{f} = \pm\sqrt{\frac{{C^r_{11}}^2 + C^r_{20}C_{02}}{{C^l_{11}}^2 + C^l_{20}C_{02}}}$$

$$g = -\frac{eC_{11}^l + fC_{11}^r}{C_{02}}$$

The normal vector $(e, f, g)^t$ being normalized, there are therefore two solutions, but a negative $\frac{e}{f}$ ratio will be rejected as it corresponds to the situation where the two cameras would look at different sides of the 3D plane (which is unlikely for real scenes).

• **General case:**

If one cannot work in a simplified geometry, a local linearization is performed through equation (5). M_0 is triangulated from matching points μ^r and μ^l, thus assuming they are the projections of the real center of gravity of the 3D region (which is a reasonable first estimate if the two cameras are not too far apart). We will denote by F_{para} the linear part of the left-to-right correspondence equations. It depends on $\frac{p}{c}$ and $\frac{q}{c}$, denoted P and Q respectively. $F_{para} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, with:

$$F_{para} = \begin{pmatrix} (R_{11} - \frac{T_x}{Z_0^l}R_{31}) - R_{33}\frac{Z_0^r}{Z_0^l}T_x P^r & (R_{12} - \frac{T_x}{Z_0^l}R_{32}) - R_{33}\frac{Z_0^r}{Z_0^l}T_x Q^r \\ (R_{21} - \frac{T_y}{Z_0^l}R_{31}) - R_{33}\frac{Z_0^r}{Z_0^l}T_y P^r & (R_{22} - \frac{T_y}{Z_0^l}R_{32}) - R_{33}\frac{Z_0^r}{Z_0^l}T_y Q^r \end{pmatrix} \quad (10)$$

This time, the inertia matrix correspondence equation is:

$$\mathcal{C}^l = (\vec{m^l} - \vec{\mu^l})(\vec{m^l} - \vec{\mu^l})^t = F_{para}\mathcal{C}^r F_{para}^t \quad (11)$$

Matrices C are positive definite, and can be decomposed as $C = (U)^T\Psi^T\Psi U$, where U is a rotation and Ψ a diagonal matrix (this is obtained by standard eigenvalue decomposition). Replacing C^l and C^r in (11), one finally deduces that matrix $\mathcal{K}$, defined as $\mathcal{K} = (\Psi^l)^{-1}U^l F_{para} U^{rT}\Psi^r$ is necessarily a rotation, and as such verifies $\mathcal{K}_{11} = \mathcal{K}_{22}$ and $\mathcal{K}_{12} = -\mathcal{K}_{21}$. The elements $\mathcal{K}_{ij}$ are analytically obtained as first order polynomials in P^r and Q^r, so that these last two equalities form a system which solution is the best planar fit.

3.2 Photometry-Based Reconstruction

Let us now consider planar patches with a constant albedo (i.e. made of one single material). It can be proved that, if the lighting consists of a pointwise light source

(or a combination of several such sources), and if the extension of the surface is small with respect to the average distance to the light source, the photometric surface profile captured from a given viewpoint is close to linear (see figure 3), i.e. $I(x, y) = \alpha x + \beta y + \gamma$.

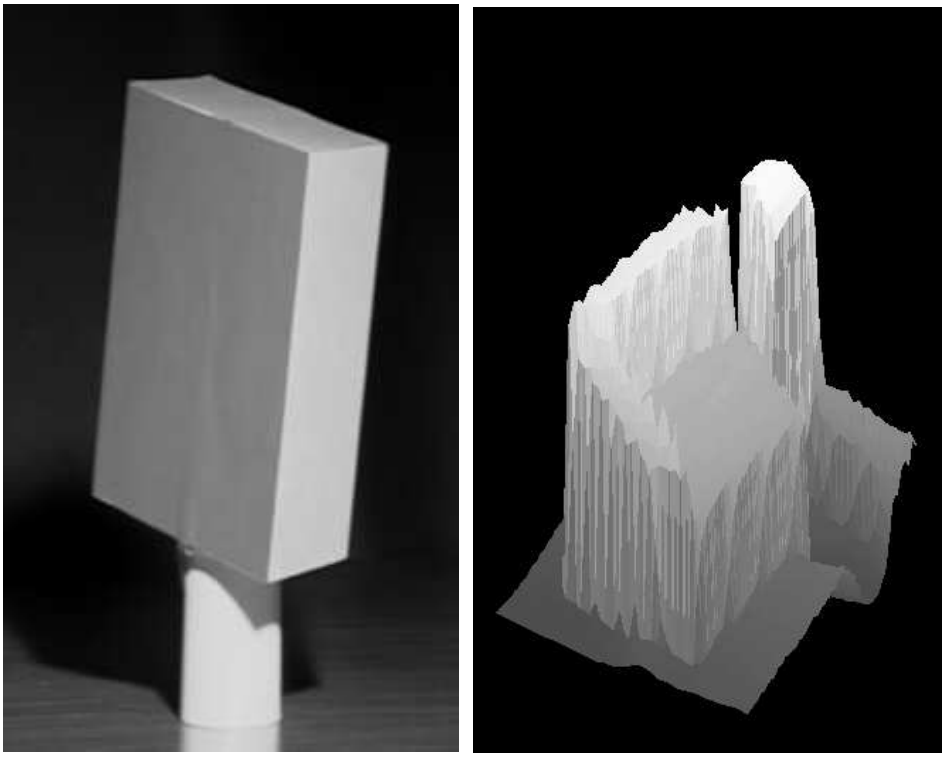

Figure 3: A lambertian cube. Left: Original 8-bit image, Right: intensity map $I = f(x, y)$. The object was about 20 cm in size, with a spotlight 3 m away.

Now we introduce an additional constraint, which we call the *stereo lambertian* assumption, stating that the intensity of the projection of a 3D point in both images of a stereo system is practically the same. Note that this is different from the classical lambertian model: true lambertian surfaces are stereo lambertian, but a wide range of specular surfaces seen over a wide range of orientations display similar intensities

in the two images. This is because the two stereo cameras are looking at the scene from practically the same vantage point.

• **Rectified geometry:**
Using the same notations as in section 2, a point (x^r, x^l, y) in the disparity space displays the same intensity on both images:

$$\begin{aligned} \alpha^r x^r + \beta^r y + \gamma^r &= \alpha^l x^l + \beta^l y + \gamma^l \Leftrightarrow \\ e = \alpha^l \quad , \quad & f = -\alpha^r, \\ g = \beta^l - \beta^r \quad \text{and} \quad & h = \gamma^l - \gamma^r \end{aligned} \tag{12}$$

The corresponding euclidian space plane equation is computed via formula (3).

• **General case:**
Let M_0 be a reference point 3D on the surface (computed as in 3.1), and its projection $\vec{m_0} = (x_0, y_0)^t$ in the image. We thus define the *centered intensity* at point $\vec{m} = (x, y)^t$ as

$$\begin{aligned} \hat{I}(x,y) &= I(x,y) - I(x_0,y_0) = \vec{k}^t(\vec{m} - \vec{m_0}) \\ &\text{with } \vec{k} = (\ \alpha\ ,\ \beta\ )^t \end{aligned}$$

The stereo-lambertian assumption can be expressed as:

$$\begin{aligned} \hat{I}^l(\vec{m^l}) &= \vec{k^l}^t(\vec{m^l} - \vec{m_0^l}) \\ &= \vec{k^l}^t F_{para}(\vec{m^r} - \vec{m_0^r}) \\ &= \hat{I}^r(\vec{m^r}) = \vec{k^r}^t(\vec{m^r} - \vec{m_0^r}) \end{aligned} \tag{13}$$

so that:

$$\vec{k^r} = F_{para}^t \vec{k^l} \tag{14}$$

Replacing the elements of F_{para} by their parametric value of equation (10), and solving for P^r and Q^r, one obtains:

$$P^r = \frac{\alpha^r - a_0\alpha^l - c_0\beta^l}{a_1\alpha^l + c_1\beta^l} \tag{15}$$

$$Q^r = \frac{\beta^r - b_0\alpha^l - d_0\beta^l}{b_1\alpha^l + d_1\beta^l} \tag{16}$$

To write the above equations, we assume that the picture is monochromatic. But since these equations are linear, it is possible to use colored images, for example using the luminance signal $Y = 0.299\ R + 0.587\ G + 0.114\ B$. Numerical stability and results are discussed in section 5.

3.3 Correlation-Based Reconstruction

As in the previous section, we assume that the surface considered is *stereo-lambertian*. Provided that the region gray levels are not uniform, a correlation-based method can be applied. As opposed to standard correlation approaches [SCM89], the proposed algorithm does not operate locally but globally using all the region photometry.

Let us define the autocorrelation function of the bounded region as [Hor86]:

$$T(\vec{s}) = \sum_{(x,y)\in R} I(x + s_x, y + s_y)I(x, y) \tag{17}$$

in its discrete version. The mapping between matching points in the right and left images is affine (equation (2) with a rectified geometry and equation (5) with a paraperspective projection). Hence:

$$\vec{s^l} = F_{para}\vec{s^r} \tag{18}$$

where F_{para} is a function of the facet specification (p, q, c) (equation (10)). Consequently, we deduce from (18) that the dependency between the left and right autocorrelation is linear:

$$\begin{aligned} T^l(\vec{s^l}) &= \sum_{R^l} I^l(\vec{u^l})I^l(\vec{u^l} + \vec{s^l}) \\ &= \sum_{R^r} I^r(\vec{u^r})I^r(\vec{u^r} + F_{para}\vec{s^l}) \\ &= T^r(F_{para}\vec{s^l}) \end{aligned} \tag{19}$$

Autocorrelation images are numerically computed with a Fast Fourier Transform before normalization. As previously, centers of gravity provide the c patch position (see 3.1). The (p, q) orientation of the face is given by fitting the left autocorrelation map on the right one, by minimizing the least-square criterion:

$$\epsilon(p, q, c) = \sum_{\vec{s^l} \in C^l} \left[T^r(M\vec{s^l}) - T^l(\vec{s^l}) \right]^2 \tag{20}$$

The minimization problem is achieved using Powell's iterative algorithm. The initial estimate is given by one of the previous reconstruction methods to ensure a fast and reliable convergence.

4 Checking data consistency

For each of the three proposed method of planar patch reconstruction, it is possible to test the relevance of its application on a given region pair.

4.1 Geometric Moments Invariants

In section 2, we produced affine mapping formulas between matching points in the right and left images. It follows that affine moment-based invariants exist between corresponding regions of a stereo pair [FS93, Rei93]. Like in the previous sections, C_{pq} is the centered moment of order (p, q) in x and y respectively (thus C_{00} is the area of the considered region). We will limit ourselves to the three fundamental invariants of order no higher than 3, namely:

$$I_1 = (C_{20}C_{02} - C_{11}^2)/C_{00}^4 \tag{21}$$

$$\begin{aligned} I_2 = {} & (C_{30}^2C_{03}^2 - 6C_{30}C_{21}C_{03} + 4C_{30}C_{12}^3 \\ & + 4C_{21}^3C_{03} - 3C_{21}^2C_{12}^2)/C_{00}^{10} \end{aligned} \tag{22}$$

$$\begin{aligned} I_3 = {} & (C_{20}(C_{21}C_{03} - C_{12}^2) \\ & - C_{11}(C_{30}C_{03} - C_{21}C_{12}) \\ & + C_{02}(C_{30}C_{12} - C_{21}^2))/C_{00}^7 \end{aligned} \tag{23}$$

Consequently, before starting the reconstruction process, the computation of invariants allows to test the validity of our hypothesis, i.e. 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 on the images, and 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 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 section 5.3.

4.2 Checking Planar Photometry

The reconstruction method of section 3.2 assumes a constant albedo on the planar patch and a linear luminance profile $I(x, y) = \alpha x + \beta y + \gamma$. To test the validity of the photometric assumption, coefficients α, β and γ can be obtained by minimizing the least square error:

$$\epsilon = \sum_{(x,y)\in R} (I(x, y) - \alpha x - \beta y - \gamma)^2 \tag{24}$$

Let us denote ϵ_{min} the residual error ϵ when the minimum is reached. The lower ϵ_{min}, the better the validity of the constant albedo assumption. The reconstruction method will be used if and only if ϵ_{min} is below a given threshold.

Moreover, the photometric reconstruction method can be only applied in presence of intensity gradients, which amplitudes can be tested by observing the slope (α, β) of the region profile I.

4.3 Checking Correlation Information

We can also take into account both geometric and photometric informations at the same time, using the generalized centered moments, which are simply the inertia moments weighted by the intensity $I(x, y)$ of each pixel $m(x, y)$:

$$C'_{ij} = \sum_{m(x,y)\in R} I(x, y)(x - L_{10})^i (y - L_{01})^j \tag{25}$$

As in section 4.1, invariants are built out of the generalized moments and can be used in a similar way.

5 Experimental results

5.1 Synthetic scenes

The three suggested reconstruction methods are first tested with synthetic data, where all the environment (geometry and photometry) is controlled. Synthetic scenes also allow to isolate errors produced by the reconstruction process, without interferences induced by the previous analysis processes (calibration, segmentation and matching).

5.1.1 Paraperspective and rectified geometry

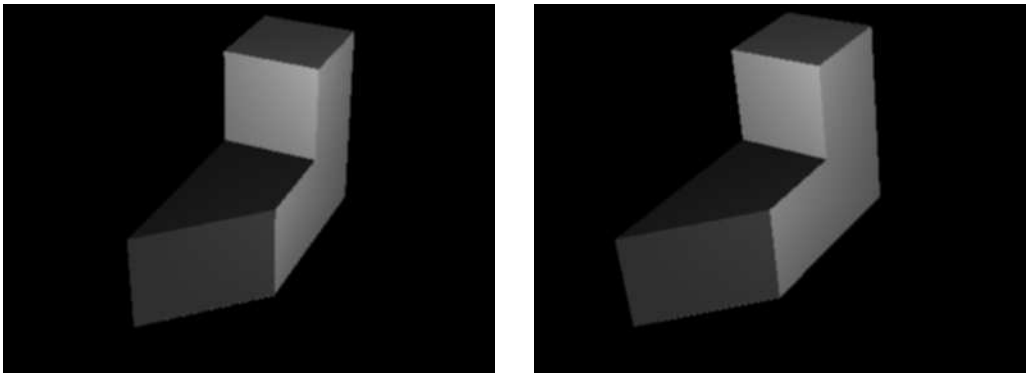

Figure 4: *Stereo pair of a synthetic object.*

First, we verify that the paraperspective model (equation (5)) gives accurate results despite approximations. For example, in the original stereo pair of figure 4, the system vergence is 10°, and the geometric reconstruction obtained in the paraperspective case is shown in figure 5b.

The orientations of the facets relatively to one another are shown in table 1, where they are compared to the real angles and to the values obtained with the rectified geometry as well.

This experiment, along with others not presented here, show that the method is more accurate in the rectified case, but the two algorithms tend to produce similar estimates if the vergence angle becomes small.

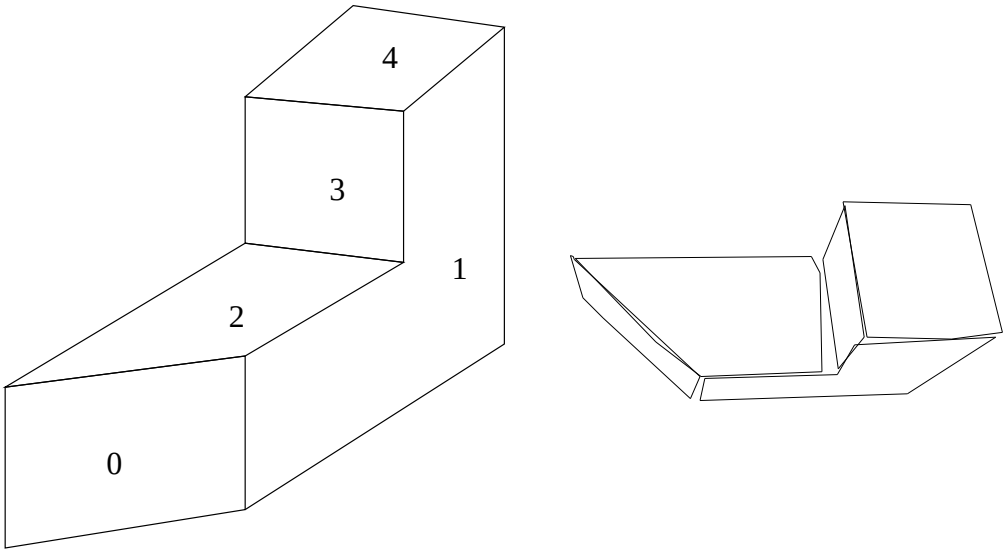

Figure 5: *a- Facets number used in table 1&2. b- top view of the reconstructed object with the geometry based method.*

5.1.2 Photometry-Based Reconstruction

On the same scene (same as figure 4, but with rectified images), we compare the result of geometry and photometry based reconstructions with the exact values. We observe that the geometry based method is more accurate than the photometric one.

As expected, one can see the photometric method is less stable, mainly because it depends on comparatively smaller gradients. Also, the smaller the region is, the less accurate the photometric estimates are. In general, this method should be used only if the geometry is poorly estimated or unavailable.

Facets pair	real	parapersp.	rectified
0–1	135.0°	132.6°	135.45°
0–2	90.0°	89.18°	89.61°
1–2	90.0°	85.55°	91.24°
1–3	90.0°	83.22°	87.79°
1–4	90.0°	85.09°	88.98°
2–3	90.0°	88.60°	94.33°
2–4	0.0°	2.96°	2.45°
3–4	90.0°	91.44°	95.20°

Table 1: *Real normal angle of pair of facets, and estimations with geometry based reconstruction in paraperspective and rectified geometry.*

Facet	geometry	photometry
0	0.407°	1.395°
1	0.183°	3.231°
2	0.154°	14.933°
3	0.916°	46.028°
4	0.183°	6.624°

Table 2: *Angular error between real and estimated normals of the facets (see figure 5) with geometry-based and photometry-based reconstructions.*

5.1.3 Correlation-Based reconstruction

By definition, the correlation-based method assumes the presence of a texture (for example, see figure 6). But as it also uses geometry implicitely, in the actual correlation computation (equation (17)), this method turns out to be the most accurate (but also the most computationally intensive). See table 3.

Figure 6: Stereo synthetic pair with a real texture.

Plane	Equation (p,q,c)	Error Angle
Real	(-0.004,-0.171,18.29)	
Geometric	(-0.014,-0.198,18.337)	1.6°
Correlation	(-0.023,-0.188,18.4)	1.45°

Table 3: *Results on the scene of figure 6*

5.2 Real Scene

Figure 7 shows the original stereo images of a football, and figure 8 the results with the geometry based reconstruction algorithm. Please note that the method is not disturbed by the fact the patches (hexagons and pentagons) are only approximately planar.

The experiments described above show the general framework presented here is suited for planar patch reconstruction, and that furthermore the three proposed methods allow to process a large spectrum of data behaviour.

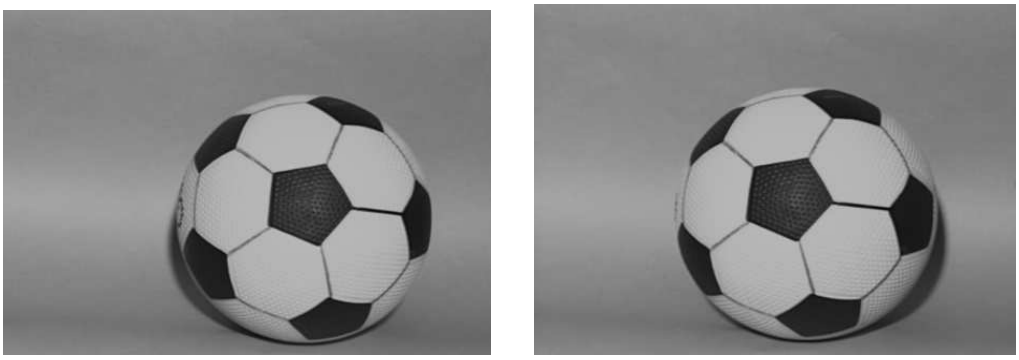

Figure 7: *Stereo pair of the real '98 soccer world championship football.*

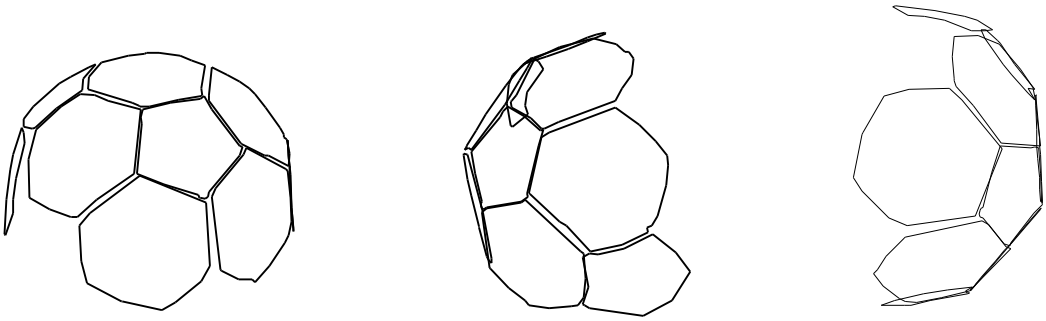

Figure 8: *Three views of the reconstructed "ball".*

5.3 Stability of the methods

We have chosen to test the robustness of our algorithms on known synthetic data, consisting of a succession of 8 stereo images of two planar patches, with increasing

amounts of occlusion (figure 9 shows the maximum occlusion case, i.e. the stereo pair consisting of images 1–8).

5.3.1 Robustness to occlusion

The table below shows the error of orientation of the occluded facet as a function of occlusion (we have plotted the corresponding occlusion-free experiment for comparison).

Experimentally, the bias produced by occlusions on the 3D patch orientation and pose is reasonable (about $12°$ for the data of figure 9), and about the same magnitude as noise and segmentation errors.

5.3.2 Invariant Checking

We use the same scene as previously to test the use of invariants defined in section 4.1. First, we remove the occluding plane and plot 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 therefore select I_1 as the more robust invariant. Figure 11 confirms this conclusion: I_2 and I_3 are very sensitive to big occlusion errors, whereas I_1 degrades nicely.

In this example, region couples with an invariant ratio out of the interval $[0.96, 1.04]$ corresponds to couples with more than 10% of their area occluded. Other experiments corroborate this result, and regions couples displaying more than 5 % invariant disparity should be considered unreliable.

Weighted invariant I'_1 holds well on textured surfaces: the table below shows the relative error in the estimation of I_1 and I'_1 in both image of figure 6.

	I_1 (Geom.)	I'_1 (Text.)
Invariant	0.7 %	1.9 %

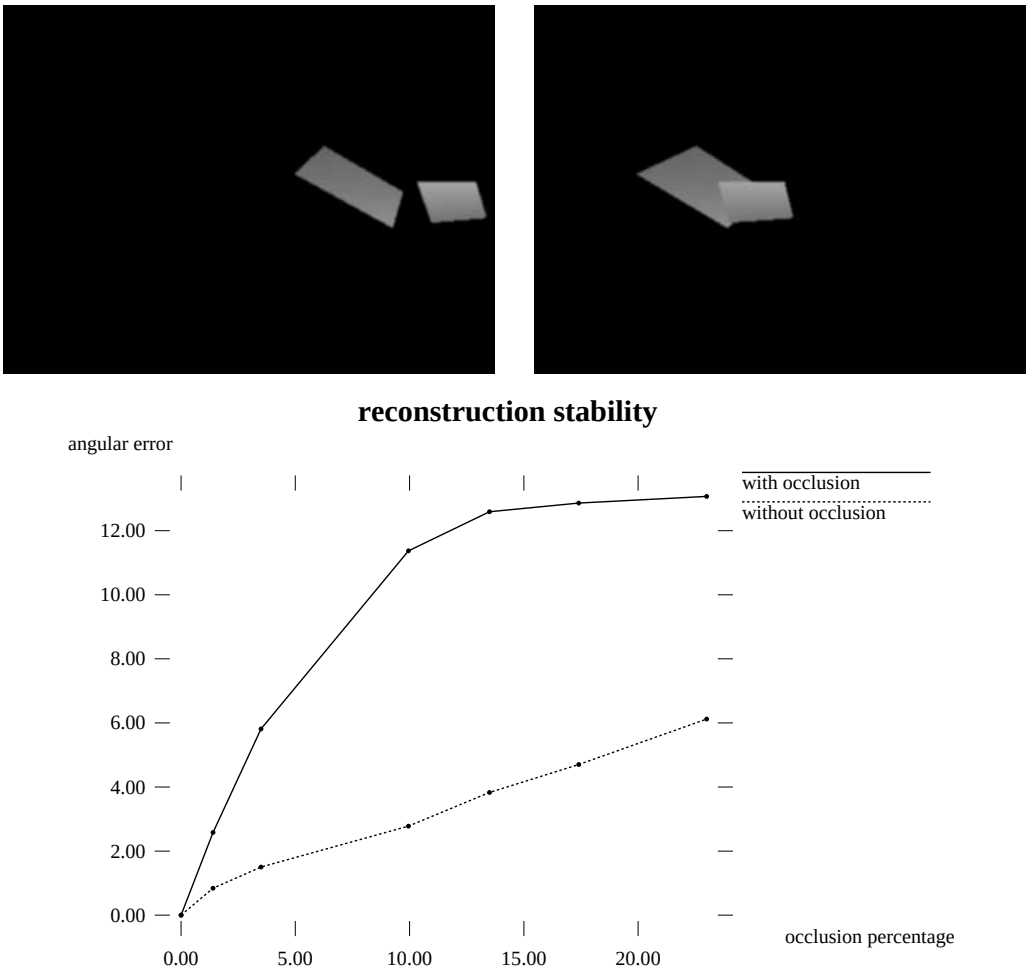

Figure 9: *Two images of the occlusion test sequence, and the reconstruction stability results (see text).*

6 Conclusion

We proposed a generic framework for the 3D reconstruction of planar surfaces from their image projections on a stereo couple. The approach, which does not rely on

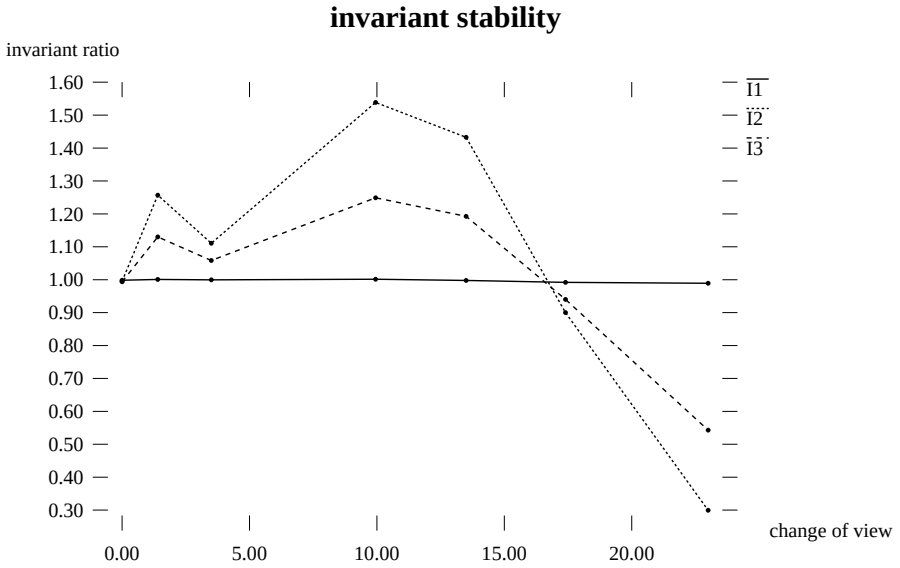

Figure 10: *Ratio of the left and right invariants I_1, I_2 and I_3 computed on the successive viewpoints.*

any local feature matching, can make use of both geometric and photometric cues, and is relatively robust to segmentation errors or occlusions.

Three examples of planar equation recovery schemes were proposed, along with methods to check for their self-consistency. Reconstructions were obtained in the 3D space, but a similar method could be applied in the projective space [Sha94] if no explicit calibration parameters are available.

The work presented here currently relies on a classical region segmentation/matching process. Future extensions include the development of a segmentation-less stereo reconstruction scheme making use of similar equations as geometric and photometric constraints, to get richer 3D information.

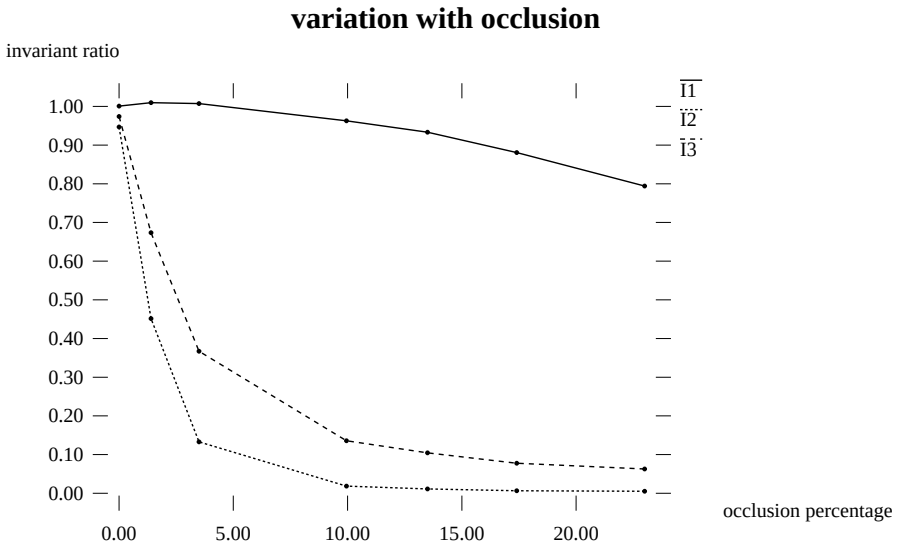

Figure 11: *Ratio of the left and right invariants I_1, I_2 and I_3 computed for different value of the occlusion.*

References

[AH90] John Aloimonos and Jean-Yves Hervé. Correspondenceless stereo and motion: Planar patches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):504–510, 1990.

[Alo90] John Y. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, August 1990.

[AMG93] A. Ackah-Miezan and A. Gagalowicz. Discrete models for energy minimizing segmentation. In *Proceedings of the 3rd International Conference on Computer Vision*, pages 200–207, 1993. 11-13 May, Berlin.

[CHT88] David B. Cooper, Yi-Ping Hung, and Gabriel Taubin. A new model-based stereo approach for 3d surface reconstruction using contours on

the surface pattern. In *International Conference on Computer Vision*, pages 74–82, 1988.

[FS93] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26:167–174, 1993.

[GM86] André Gagalowicz and Olivier Monga. A new approach to image segmentation. In *Proceedings of the Eighth International Conference on Pattern Recognition*, pages 265–267, Paris, October 1986.

[Hor86] Berthold Klaus Paul Horn. *Robot Vision*. MIT Press, Cambridge, 1986.

[KNC93] Steen Kristensen, Henrik Nielsen, and Henrik Christensen. Cooperative depth extraction. In *8th Scandinavian Conference on Image Analysis*, pages 321–328, 1993.

[Mar82] David Marr. *Vision*. Freeman, San Francisco, 1982.

[PP93] N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 29(9):1277–1294, 1993.

[Rei93] Thomas H. Reiss. Recognizing planar objects using invariant image features. In *Lecture Notes in Computer Science*. Springer-Verlag, 1993.

[RG91] S. Randriamasy and A. Gagalowicz. Region based stereo matching oriented image processing. In *Proceedings of Computer Vision and Pattern Recognition*, Maui, Hawaii, June 1991.

[SCM89] Y.C. Shah, R. Chapman, and R.B. Mahani. A new technique to extract range information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):768–773, 1989.

[Sha94] A. Shashua. Projective structure from uncalibrated images - structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, 1994.

[SHC90] J. Subrahmonia, Y. P. Hung, and D. B. Cooper. Model-based segmentation and estimation of 3d surfaces from two or more intensity images using markov random field. In *Proceedings of the 10th International Conference on Pattern Recognition*, pages 390–397, June 1990.

[VSCG89] Laurent Vinet, Peter T. Sander, Laurent Cohen, and André Gagalowicz. Hierarchical regions based stereo matching. In *Proceedings of the Sixth Scandinavian Conference on Image Analysis*, pages 71–78, Oulu, Finland, June 1989.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399