

Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models

F. Devernay and O. D. Faugeras

INRIA. 2004, route des Lucioles. B.P. 93. 06902 Sophia-Antipolis. FRANCE.

Abstract

We are considering the problem of recovering the three-dimensional geometry of a scene from binocular stereo disparity. Once a dense disparity map has been computed from a stereo pair of images, one often needs to calculate some local differential properties of the corresponding 3-D surface such as orientation or curvatures. The usual approach is to build a 3-D reconstruction of the surface(s) from which all shape properties will then be derived without ever going back to the original images. In this paper, we depart from this paradigm and propose to use the images directly to compute the shape properties. We thus propose a new method extending the classical correlation method to estimate accurately both the disparity and its derivatives directly from the image data. We then relate those derivatives to differential properties of the surface such as orientation and curvatures.

1 Introduction

1.1 Motivation

Three-dimensional shape analysis in computer vision has often been considered as a two step process in which a) the structure of the scene is first recovered as a set of coordinates of 3-D points and b) models are fitted to this data in order to recover higher-order shape properties such as normals or curvatures which are first and second order differential properties of the shape surface. The original images are not used anymore whereas this is clearly where the information lies.

In some applications one may want to use even higher order properties such as affine or projective differential invariants, that would be especially useful in the situation of an uncalibrated stereo rig [5]. All these quantities can be expressed, using the perspective projection matrices of the two cameras (or the fundamental matrix in the case of an uncalibrated system), in terms of the derivatives of the disparity field. As a

consequence, we are confronted to the task of estimating the spatial derivatives of the disparity map and we explore the possibility of estimating these derivatives directly from the images rather than applying to the disparity map the same paradigm as to the 3-D shape.

1.2 Related Work

We very briefly review the work that has been done in interpreting stereo disparity. Up till now the major part of the existing studies is on the interpolation or approximation of the possibly sparse disparity map by a surface. This was done using either minimization of spline functions [1, 12], or interpolation by polynomial surface patches [8]. For both methods, surface orientation and the presence of surface discontinuities can be detected and taken into account. To calculate the orientation of the observed surface, another approach was to simply differentiate numerically the point-by-point distance reconstruction [2].

Theoretical results were also obtained for the calculation of surface orientation by studying the local projections of a surface and the displacement vector field generated by movement (i.e. the optical flow) [10], or the disparity field in the case of stereoscopy [9]. A result that is closer to the approach presented here is the calculation of the three-dimensional surface orientation from the disparity field and its first derivatives [13], but, as we show in this article, it can be done much more simply, thus allowing to reach the second order derivatives, i.e. curvatures. A method similar to ours was also applied to estimation of traversability for robot motion planning [11].

1.3 Contributions

If we want to calculate the local differential properties of a 3-D surface, we can go at least two ways; first, we can reconstruct the scene points in three dimensions, fit some surface to the reconstructed points and compute the differential properties from the fitted surfaces. A second possibility is to avoid the explicit

reconstruction and work directly from the disparity map. Because of the computational effort involved in the first approach, we choose here the second one.

We thus first have to compute the derivatives of the disparity. Since the precision of the dense disparity map calculated by a standard correlation technique is only about one pixel, we must either regularize the disparity map or compute its derivatives differently. The first solution implies the use of *local* regularization techniques, because we must keep in mind that the disparity map may contain holes due for example to object discontinuities or occluding contours. We chose to explore a second solution and present a new method to compute precise values of the disparity and its derivatives, up to second derivatives, directly from the image intensity data, i.e. without using explicitly any derivation operator.

We then present a method to compute the three-dimensional surface orientation from the first-order derivatives of the disparity. The analytic expressions are very simple when working in standard coordinates (i.e. when the images are rectified so that epipolar lines are horizontal). We also extend this to the computation of surface curvature from the second-order derivatives of the disparity, but the resulting expressions are less simple.

We tested our algorithms on real images successfully, and the results are presented at the end of this paper.

Our method can be easily extended to the case of an uncalibrated stereo rig [5], but in that case we will need to use projective differential invariants instead of Euclidean invariants. The use of weak calibration seems to be a very promising and still mostly unexplored field of research.

2 Computing derivatives of disparity

If we want to know the local surface orientation and curvatures from a stereoscopic pair of images, we have to calculate somehow the derivatives with respect to the image coordinates of the disparity map of a 3-D surface. Two problems arise if we try to differentiate the disparity map: first, because the classical correlation algorithms give the disparity at a precision of about one pixel, there is noise in the disparity map, and second, the disparity map may contain holes because of occluded regions or points where the correlation failed (Figure 5). One solution is to use a deriving operator with a finite support that would regularize the data, like the one we present below.

We also present another method that uses directly

the image intensity data to compute the derivatives without any kind of derivation. Using this method gives more accurate results, but at a much higher computational cost.

2.1 From the disparity map

As we pointed out, if we want to use the disparity map itself to calculate its derivatives we have to perform some local regularization that can handle holes in the disparity map. We chose to do a local least square approximation of the disparity data by a model, and then recover the derivatives from the coefficients of the fitted model.

To calculate the first derivatives of disparity, we fit a plane model on the data located in a rectangular window centered at the considered point [3]. Then we consider we can trust the result if both the χ^2 of the approximation is under a fixed threshold and the derivatives of disparity verify the ordering constraint [6]. This last condition can also be replaced by the disparity gradient limit.

A perhaps better method to compute these derivatives from the disparity map can be found in [8], but the quality of the results will still depend on the precision of the disparity map, which is the crucial problem.

2.2 An enhanced correlation method

We thought that instead of trying to look for the derivatives in the disparity map, where they may be definitively lost because of noise or correlation errors, why not look for these directly in the image intensity data?¹ The idea comes from the following observation: a small surface element that is viewed as a square of pixels in the left image can be seen in the right image as a distorted square, and an approximation of this distorted square can be computed from the derivatives of the disparity.

Let us call $d(u, v)$ the disparity at point (u, v) , that is to say that the point in the right image corresponding to (u, v) is $(u + d, v)$. Let α and β be the derivatives of d with respect to u and v , respectively. Then the point corresponding to $(u + du, v + dv)$ is $(u + d + (1 + \alpha)du + \beta dv + o(du + dv), v + dv)$. This means that the region corresponding to a small square centered at (u, v) in the left image is a sheared and slanted rectangle centered at $(u + d, v)$ in the

¹In this subsection we work in standard coordinates, i.e. the original images are rectified so that the epipolar lines are horizontal and consequently the disparity between the left and right images is only horizontal (Figure 3). The reference image used for the computation of disparity is the left image.

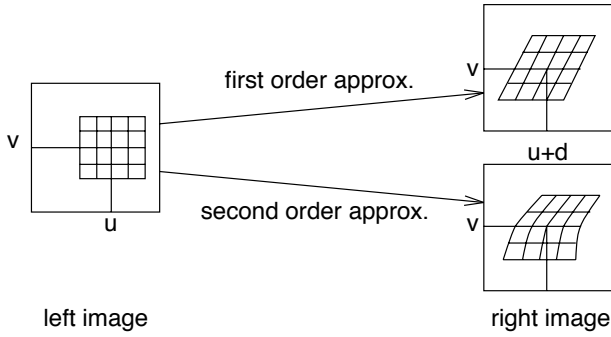


Fig. 1.: How a small square region in the left image is transformed in the right image: first (top) and second (bottom) order approximations of the deformation.

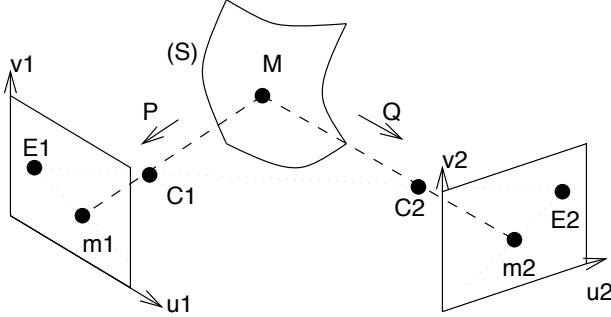


Fig. 2.: The stereoscopic system and the epipolar geometry

right image (Figure 1). We can use the same scheme to compute a higher order approximation of the deformation that operates on a square region in the left image. We obtain the point corresponding to $(u + du, v + dv)$ by calculating the Taylor series expansion of $(u, v) \rightarrow (u + d(u, v), v)$ up to order n . An example of such a deformation is shown in Figure 1 for $n = 2$.

Now that we know how an element of surface of the left image is deformed in the right image given the derivatives of disparity, we can inversely try to guess the derivatives of the disparity as the parameters of the deformed element which maximize the correlation between both regions. For example to calculate the first derivatives of the disparity we simply have to calculate the values of d , α , and β that maximize the correlation between the square region in the left image and the slanted and sheared region in the right image. To calculate the correlation between both regions, that are of different shape, we just do an interpolation between the intensity values of the right image, at the positions corresponding to the centers of the pixels of the deformed left image, and then we calculate the correlation criterion as a finite sum, as we do in classical correlation techniques.

Then, to find the values of the disparity and its derivatives that maximize the correlation, we calculate

a dense disparity map by a standard correlation technique, that is used as the first component of the initialization vector for a classical minimization method. The other components, which are the derivatives of disparity, are initialized at 0, and a classical minimization method finds the best values for the disparity and its first and second derivatives. We finally get six maps: one for the disparity itself, two for its first derivatives, and three for its second derivatives.

3 From derivatives of disparity to 3-D differential properties

Let us consider a pair of calibrated cameras, i.e. we know the projection matrices P and Q associated respectively to the first and the second camera. From P and Q we can compute the optical centers C_1 and C_2 (Figure 2) and their epipoles E_1 and E_2 (the epipoles correspond to the projection in each camera of the optical center of the other camera).

A 3-D surface (S) is projected on both cameras and we want to calculate the orientation and the curvature of the surface in each point. We first study the generic case where the cameras are in a general position, and then we will restrict ourselves to the case of standard geometry which simplifies the calculations.

3.1 Cameras in generic position

Let \tilde{e}_1 and \tilde{e}_2 be the vectors in homogeneous coordinates representing the two epipoles, a_1 and b_1 be two points of the first camera not aligned with E_1 , and a_2 and b_2 be two points of the second camera not aligned with E_2 (the points a_1 and b_1 do not have to match a_2 and b_2). Let M be a physical point belonging to the observed surface (S) , whose projections on the two cameras are m_1 and m_2 . The corresponding vectors in homogeneous coordinates \tilde{m}_1 and \tilde{m}_2 can be written as linear combinations of the other points in each camera, $\tilde{m}_1 = \lambda_1 \tilde{e}_1 + \tilde{a}_1 + \mu_1 \tilde{b}_1$ and $\tilde{m}_2 = \lambda_2 \tilde{e}_2 + \tilde{a}_2 + \mu_2 \tilde{b}_2$.

The epipolar constraint is $\tilde{m}_2^T \mathbf{F} \tilde{m}_1 = 0$, where \mathbf{F} is the fundamental matrix [7] corresponding to the stereoscopic system. Since $\mathbf{F} \tilde{e}_1 = \mathbf{F}^T \tilde{e}_2 = 0$, it can be rewritten $(\tilde{a}_2 + \mu_2 \tilde{b}_2)^T \mathbf{F} (\tilde{a}_1 + \mu_1 \tilde{b}_1) = 0$, so there exists a homographic relationship between μ_1 and μ_2 :

$$\mu_2 = -\frac{\tilde{a}_2^T \mathbf{F} (\tilde{a}_1 + \mu_1 \tilde{b}_1)}{\tilde{b}_2^T \mathbf{F} (\tilde{a}_1 + \mu_1 \tilde{b}_1)} = \frac{a\mu_1 + b}{c\mu_1 + d} \quad (1)$$

Moreover, by matching the points of both images we obtain the dense disparity map f :

$$\lambda_2 = f(\lambda_1, \mu_1) \quad (2)$$

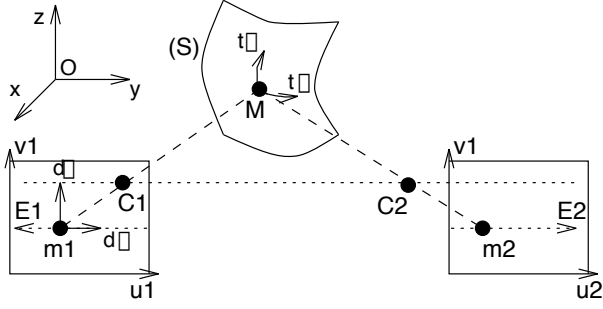


Fig. 3.: A configuration corresponding to standard geometry. The vectors \mathbf{t}_λ and \mathbf{t}_μ , tangents to (S) at M , correspond to small displacements of the image point.

Since matrices P and Q are known, we also have the reconstruction formula (which is developed later in the case of standard geometry), $\mathbf{M} = \mathbf{r}(\lambda_1, \mu_1, \lambda_2, \mu_2)$.

Surface orientation: To calculate the surface orientation we just have to differentiate once the reconstruction formula and substitute the values of $d\lambda_2$ and $d\mu_2$ found by differentiation of equations 1 and 2:

$$d\mathbf{M} = \mathbf{t}_\lambda d\lambda_1 + \mathbf{t}_\mu d\mu_1 \quad (3)$$

$$\mathbf{t}_\lambda = \frac{\partial \mathbf{r}}{\partial \lambda_1} + \frac{\partial \mathbf{r}}{\partial \lambda_2} \frac{\partial f}{\partial \lambda_1}, \mathbf{t}_\mu = \frac{\partial \mathbf{r}}{\partial \mu_1} + \frac{\partial \mathbf{r}}{\partial \lambda_2} \frac{\partial f}{\partial \mu_1} + \frac{ad - bc}{(c\mu_1 + d)^2} \frac{\partial \mathbf{r}}{\partial \mu_2}$$

Consequently, the tangent plane to the surface (S) at point \mathbf{M} is the vector space generated by the two vectors \mathbf{t}_λ and \mathbf{t}_μ .

Surface curvature: Knowing the local surface curvature may be more useful than the surface orientation, so we continued our computation to get the surface curvature from the first and second derivatives of disparity. The method consists of differentiating twice the reconstruction formula, and then computing the second order properties of the surface such as principal directions and curvatures from this equation. More details can be found in [3].

3.2 In standard geometry

Standard geometry [6] consists of rectifying the images so that the epipolar lines are horizontal (Figure 3). The preceding calculations can be simplified a lot because the epipoles are at infinity, $\tilde{\mathbf{e}}_1 = \tilde{\mathbf{e}}_2 = (1, 0, 0)^T$, and the epipolar constraint is simply $\mu_1 = \mu_2$. Besides we can choose $\tilde{\mathbf{a}}_1 = \tilde{\mathbf{a}}_2 = (0, 0, 1)^T$ and $\tilde{\mathbf{b}}_1 = \tilde{\mathbf{b}}_2 = (0, 1, 0)^T$, i.e. $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ correspond to the horizontal and the vertical direction in each image.

3-D reconstruction: In standard geometry the projection matrices associated to each camera have

only a few differences:

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{p}_1^T & p_{14} \\ \mathbf{p}_2^T & p_{24} \\ \mathbf{p}_3^T & p_{34} \end{bmatrix}, \tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{q}_1^T & q_{14} \\ \mathbf{p}_2^T & p_{24} \\ \mathbf{p}_3^T & p_{34} \end{bmatrix}$$

By writing that points $\mathbf{m}_1 = (u_1, v_1)$ and $\mathbf{m}_2 = (u_2, v_1)$ (Figure 3) are the projections of the same 3-D point, we obtain the reconstruction formula, $\mathbf{r}(u_1, v_1, u_2) = \mathbf{A}^{-1} \mathbf{B}$, with:

$$\mathbf{A} = \begin{bmatrix} \mathbf{p}_1^T - u_1 \mathbf{p}_3^T \\ \mathbf{p}_2^T - v_1 \mathbf{p}_3^T \\ \mathbf{q}_1^T - u_2 \mathbf{p}_3^T \end{bmatrix}, \mathbf{B} = \begin{pmatrix} u_1 p_{34} - p_{14} \\ v_1 p_{34} - p_{24} \\ u_2 p_{34} - q_{14} \end{pmatrix} \quad (4)$$

We also need the Jacobian matrix \mathbf{J}_r , whose columns are the partial derivatives of \mathbf{r} with respect to u_1, v_1 and u_2 , to calculate surface orientation, and the differential of the Jacobian to calculate the surface curvatures:

$$\mathbf{J}_r = ((\mathbf{p}_3^T p_{34}) \tilde{\mathbf{r}}) \mathbf{A}^{-1} \quad (5)$$

$$d\mathbf{J}_r = (\mathbf{p}_3^T d\mathbf{r}) \mathbf{A}^{-1} - \mathbf{A}^{-1} d\mathbf{A} \mathbf{J}_r \quad (6)$$

Surface orientation: The equation 3 becomes in standard geometry:

$$d\mathbf{M} = \left(\frac{\partial \mathbf{r}}{\partial \lambda_1} + \frac{\partial \mathbf{r}}{\partial \lambda_2} \frac{\partial f}{\partial \lambda_1} \right) d\lambda_1 + \left(2 \frac{\partial \mathbf{r}}{\partial \mu_1} + \frac{\partial \mathbf{r}}{\partial \lambda_2} \frac{\partial f}{\partial \mu_1} \right) d\mu_1$$

so that the tangent vectors can be written $\mathbf{t}_\lambda = \mathbf{J}_r \left(1, 0, \frac{\partial f}{\partial u_1} \right)^T$ and $\mathbf{t}_\mu = \mathbf{J}_r \left(0, 2, \frac{\partial f}{\partial v_1} \right)^T$, and the orientation of the surface is given either by $\mathbf{t}_\lambda \times \mathbf{t}_\mu$ if the image frame is direct, or by $\mathbf{t}_\mu \times \mathbf{t}_\lambda$ if the image frame is indirect (Figure 3).

Surface curvature: The expression of $d^2 \mathbf{M}$ we obtain in the generic case can be simplified in the case of standard geometry. More details can be found in [3].

4 Results

We present here some results that we obtained using the different techniques described in this paper. The stereoscopic system we used consists of a pair of CCD cameras with 16mm lenses (we use a big focal length because we want minimum distortion). The image resolution is 512×512 , and the subject of our stereograms were several textured objects representing moulds of human faces, human torsos, vases, and real faces (Figure 4). The system was calibrated using weak calibration [7] for stereo and a calibration grid for reconstruction.



Fig. 4.: Two sample cross-eyed stereograms.

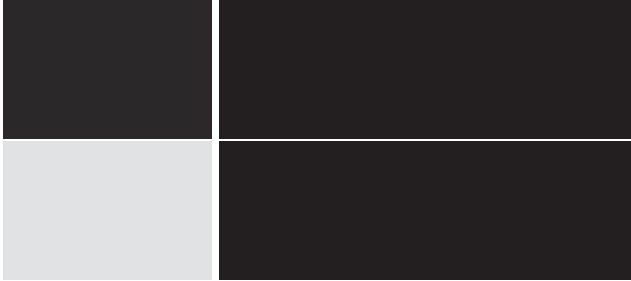


Fig. 5.: The disparity field obtained by standard correlation (left) and the first derivatives computed by plane fitting (right).

4.1 The disparity and its derivatives

The first step for the estimation of differential properties of 3-D surfaces is to calculate a dense disparity map from the considered pair of images and its derivatives, up to the order we need. We present here some results for the two methods presented in this paper and compare them.

By plane fitting: Some results of the application a standard correlation method followed by plane fitting are shown in figure 5. The contrast was augmented so that we can see the main defect of this method: some bumps appear all over the surface due to the noise that was present in the original disparity data. They can disappear if we increase the size of the region used for fitting, but we will lose precision and localization of surface features. A solution would be to fit a higher degree model or do some regularization before processing the data.

By correlation: This method gives better results, at the price of a higher computational cost. The most remarkable one is the new disparity map, which is a lot more precise because we take into account the local image deformations. The first derivatives of disparity seem to be also accurate, especially when compared to those obtained by plane fitting (Figure 6), but some

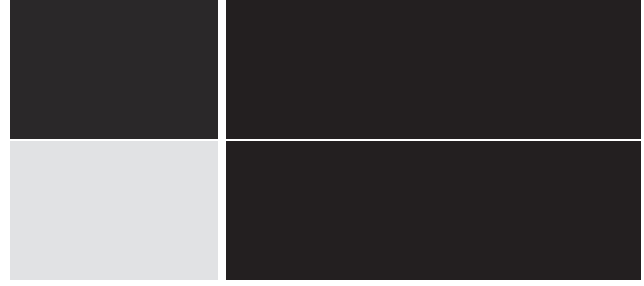


Fig. 6.: The disparity field (left) and the first derivatives of disparity obtained by correlation (right).



Fig. 7.: The second derivatives of the disparity: $\frac{\partial^2 f}{\partial x^2}$ (left), $\frac{\partial^2 f}{\partial x \partial y}$ (center), and $\frac{\partial^2 f}{\partial y^2}$ (right).

strange phenomenon occur in the images of the second derivatives of disparity.

In fact, there seem to be horizontal stripes all over the image of the second derivative with respect to y , and their amplitude and frequency decrease when the size of the correlation window increases, so that it must come from some kind of noise (Figure 7). Since this appears only in this image, this must be the consequence of the synchronization error at the beginning of each video line. After verification, the amplitude of the waves correspond to the pixel jitter value given in the technical data of the acquisition system, which can be up to 0.5 pixel.

4.2 3-D reconstruction and orientation

We compared the 3D reconstruction obtained from two different dense disparity maps: one obtained by a standard correlation algorithm, and one refined by our enhanced correlation method. The result of our method is by far better than the other one, as can be seen in Figure 8 which represents a close view of a 100×100 region of the face stereogram, where the amplitude of the variation of disparity is less than 10

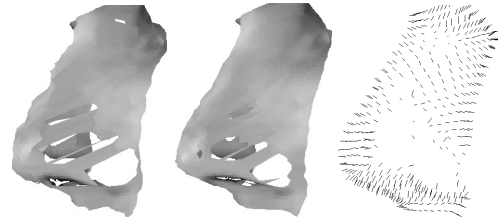


Fig. 8.: Reconstruction of the nose from the pair of Figure 4: using standard correlation (left), using enhanced correlation (center), and the field of normals (right)

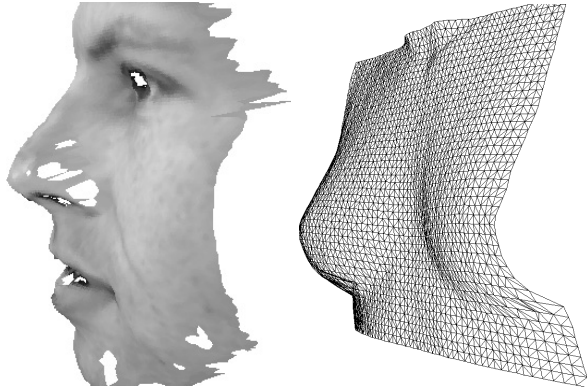


Fig. 9.: The face reconstruction with intensity mapping and the bust reconstruction, both using enhanced correlation.

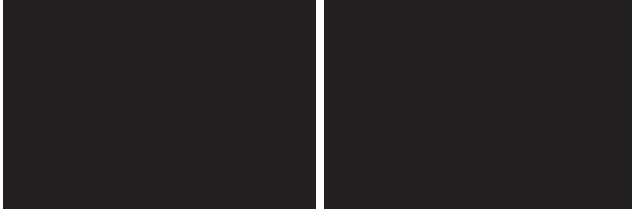


Fig. 10.: Images of the Gaussian (left) and mean (right) curvatures

pixels. We also show a subsampling of the field of normals that was obtained together with the disparity map. The whole face is represented in Figure 9.

This is not an easy case (the surface is only slightly textured), so we can hope that our correlation technique will work on many kinds of 3D surfaces. Using it with first order approximation of the local image distortion is enough to get both a precise reconstruction and the field of normals, and it is much faster than with the second order approximation.

We also calculated the Gaussian and the mean curvatures of the torso stereogram (Figure 10). The problem is that the stripes that appeared in the second derivative of the disparity over y are still present, so that there is some error in the curvature maps.

5 Conclusion

We have described a method to compute the differential properties of 3-D shapes, such as surface orientation or curvature, from stereo pairs. The advances are both theoretical and practical: we first have shown how the 3-D differential properties are related to the derivatives of the disparity map, and second we have described a new method to compute these derivatives directly from the image intensity data by correlation.

This enhanced correlation method is more accurate than classical methods but slower, so that we may

want to use it only locally (e.g. for a better interpretation of regions of interest). The second order derivatives of the disparity computed by this method are nonetheless not as stable as we may wish, and a good solution may be a hybrid method using both the image intensity and the 3-D reconstruction.

The next step of our work is to use the enhanced correlation method on non-rectified images, using the epipolar constraint, in order to get rid of the noise caused by the rectification. We also have to eliminate the holes that may be present in the original disparity map, because this one is used as an initialization of our process. We also plan to make the calculations in the case of weak calibration, and to compute higher-order properties such as affine or projective curvature. This can be applied to feature detection and recognition, or it can be used to find regions of interest in an active vision approach.

References

- [1] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [2] M. Brady, J. Ponce, A. Yuille and Asada H. Describing Surfaces. *Computer Vision, Graphics, and Image Processing*, 32:1–28, 1985.
- [3] F. Devernay and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. Technical report, INRIA, 1994. To appear.
- [4] O.D. Faugeras. Euclidean, Affine and Projective Planar Differential Geometry for Scale-Space Analysis. Technical report, INRIA, 1994. To appear.
- [5] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In Giulio Sandini, editor, *Proc. 2nd Eur. Conf. Computer Vision*. Springer-Verlag, 1992.
- [6] O.D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [7] O.D. Faugeras, T. Luong and S. Maybank. Camera self-calibration: theory and experiments. In Giulio Sandini, editor, *Proc. 2nd Eur. Conf. Computer Vision*. Springer-Verlag, 1992.
- [8] W. Hoff and N. Ahuja. Extracting surfaces from stereo images. In *Proc. Int. Conf. Computer Vision*, 1987.
- [9] J.J. Koenderink and A.J. van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29–35, 1976.
- [10] H. C. Longuet-Higgins and K. Prazdny. The interpretation of moving retinal images. *Proceedings of the Royal Society of London*, B 208:385–387, 1980.
- [11] L. Robert and M. Hebert. Deriving Orientation Cues from Stereo Images. In Jan-Olof Eklundh, editor, *Proc. 3rd Eur. Conf. Computer Vision*, 1994. Springer Verlag.
- [12] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:413–424, 1986.

- [13] R.P. Wildes. Direct Recovery of Three-Dimensional Scene Geometry from Binocular Stereo Disparity. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(8):761–774, August 1991.