Shape from Stereo Using Fine Correlation: Method and Error Analysis

Frédéric Devernay

I3S / Université de Nice Sophia-Antipolis

BP.121, 06903 Sophia Antipolis. FRANCE.

`Frederic.Devernay@unice.fr`

Olivier Faugeras

INRIA Sophia-Antipolis

B.P. 93. 06902 Sophia-Antipolis. FRANCE.

`Olivier.Faugeras@sophia.inria.fr`

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, called *fine correlation*, 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.

We present the results of the reconstruction and of the estimation of the surface orientation and curvatures on some stereo pairs of real images.

The performance and accuracy of both classical and fine correlation are analysed using synthetic images. By modeling the correlation error distribution as a mixture of Gaussian, we can compute the intrinsic accuracy of each method and the proportion of false matches, depending on the local disparity slope. The results show that the accuracy is almost constant with respect to slope using fine correlation, whereas it increases dramatically using classical correlation.

Keywords: Stereoscopy, Differential properties of surfaces, Disparity interpretation, Correlation, Three-dimensional vision.

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 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 [Fau95] that would be especially useful in the situation of an uncalibrated stereo rig [Fau92]. 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 deriving them from the disparity map, as is usually done with 3-D urfaces.

We also want to better understand the error model underlying the estimation of these surface properties, and how to chose the parameters controling the stereo algorithm (such as the correlation window size) in order to get optimal results for a given scene.

1.2 Related Work

Just as the disparity map of a stereo pair gives information about the position of each point in the scene, the derivatives of disparity with respect to image coordinates are directly linked to differential properties of the observed surface (such as normal or curvatures) [KvD76, Wil91].

Up till now the major part of the existing studies is on the interpolation or approximation of the possibly sparse reconstruction or disparity map by a surface. This was done using either minimization of spline functions [BZ87, Ter86, Gri81], or interpolation by polynomial surface patches [HA87, EW87]. The major backdraw of these methods is that they smooth out both the surface and its first or higher order discontinuities. These discontinuities can be identified by looking at the residuals of the function minimization [HA87, EW87]. Another solution, which is used in snakes-based methods, is to use a non-convex minimization technique, in which the discontinuities are allowed within a smooth surface [BZ87, Ter86]. One last method to calculate the orientation of the observed surface is simply to differentiate numerically the point-by-point distance reconstruction [BPYH85, MN84] (see sec. 2.2).

Another way to compute the derivatives of disparity is to take into account the orientation of the observed surface in the stereoscopy algorithm, such as in some Bayesian approaches [Bel93, SFB96]. These methods use the *a priori* probability of each surface orientation, which can be computed from the kind of scene [SFB96]. Surface orientation can also be used and computed with correlation-based stereo methods: these *stretch-and-shear* correlation methods forget about the classical fronto-parallel constraint, and use correlation windows which can be stretched and sheared [DF94, LTS94, RH94, JM92]. The *fine correlation* method presented here belongs to this class.

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 3-D, fit some surfaces 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 and the extra noise added by the reconstruction process, we choose here the second approach.

Thus, the derivatives of the disparity have to be computed first. Since the accuracy 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.

For these reasons 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. For this we use the local deformations of the image of a surface from the left view to the right view, which give information on the derivatives of the disparity. We implemented an extension of a correlation technique based on this principle, called fine correlation, which gives good results up to the second order derivatives of disparity. The counterpart is that the computation is much slower than calculating directly derivatives from the disparity map itself. However, the quality of the results is worth the effort.

We then show how 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). Similarly, the computation of surface curvatures can be obtained from the second-order derivatives of the disparity, but the resulting expressions are less simple.

Our algorithms have been tested on real images successfully, and the results are presented at the end of this paper. We also analysed the improvement of fine correlation over classical correlation using synthetic data, which provide ground-truth disparity and its derivatives. Each error distribution was modeled as a mixture of Gaussian, which enabled the computation of both the intrinsic accuracy of the method, and the proportion of false matches. The computation of these measures as a function of disparity slope showed that the proportion of false matches increases with slope for both methods, but the intrinsic accuracy is constant for fine correlation, whereas it increases with slope for classical correlation.

Our method can be easily extended to the case of an uncalibrated stereo rig [Fau92], but in that case we will need to use either affine or projective invariants of the surface instead of Euclidean invariants. The notion of surface orientation is the same in an affine or projective world, but higher order affine or projective differential invariant need high-order derivatives of the disparity map, which may be hard to extract from images, even using fine correlation.

2 Computing derivatives of disparity

2.1 Difficulties of the computation

If we want to know the local surface orientation and curvatures from a stereoscopic pair of images, we have to compute somehow the derivatives with respect to the image coordinates of the disparity map of a three-dimensional object. As always we encounter the most common problem in computer vision: the data contains too much noise to be simply differentiated because the accuracy of a disparity map obtained by a standard correlation technique is not much better than one pixel.

Another obstacle to the computation of the derivatives from the disparity map is that the map is *dense* but nevertheless contains holes due to occluded regions, discontinuities of the disparity

or positions where no maximum was found for the correlation criterion. For these reasons, if we want to regularize the disparity map, we have to use *local* operators like the one presented below.

Because the first order derivatives that we computed by regularization of the dense disparity map were somewhat unsatisfactory, we thought of another method to compute these directly from the image intensity data, without using any differentiation operator. An enhanced correlation method, called *fine correlation*, that takes into account the local deformations from one image to the other, is described below. This method gives more accurate results, but at a much higher computational cost than classical correlation and differentiation.

2.2   Computing derivatives from the disparity map

As we pointed out, if we want to use the disparity map itself to compute its derivatives, some local regularization method that can handle holes in the disparity map has to be performed. We chose to fit a model by a least square approximation of the data located in the neighborhood of a point where we want to compute the derivatives, and then deduce the derivatives from the equations of this model.

To compute the first derivatives of the disparity from the disparity map, the model we use is simply a plane of equation $d(x, y) = \alpha x + \beta y + \gamma$ and the approximation is made on the disparity values present in a $h \times v$ window centered at the current point. Consequently, the fitting algorithm is only a bilinear regression of a set of 3-D points (x_i, y_i, d_i) (d_i is the disparity at (x_i, y_i)). Once we have the equation of the fitted plane, the coefficients α and β correspond to the first derivatives of the disparity in the x and y directions (the value of γ can be considered as a regularized value of the disparity). We can also compute the standard deviations σ_α and σ_β on α and β. We decide to consider the values α and β as valid if there are enough data points in the $h \times v$ image window, $\sigma_\alpha < s$, $\sigma_\beta < s$, and $\alpha > -1$, where the threshold s must be fixed (we chose $s = 0.05$ for our experiments).

The condition $\alpha > -1$ corresponds to the differentiation of the classical ordering constraint, which says that if two points A_1 and B_1 lie close enough on the same epipolar line in a certain order (e.g. A_1 is closer to the epipole E_1 than B_1) in the first retina, then the images of the corresponding 3-D points in the second retina lie in the same order. This condition can also be replaced by the disparity gradient limit [PMF85] which may be closer to the human visual system limit. Since the disparity gradient is related to local surface orientation, this is equivalent to setting a maximum angle between the local orientation of the surface and the optic ray.

A better method to compute the derivatives of disparity from the disparity map can be found in [HA87], but the quality of the results will still depend on the precision of the disparity map, which is the crucial problem.

2.3   Fine correlation

The disparity map computed by correlation is known to be accurate to about one pixel (as shown section 5), so it cannot be differentiated to compute accurate local orientation and curvatures. Besides, we will see that the disparity accuracy is worse for high disparity slopes. Instead of trying to look for the derivatives in the disparity map, where they may be definitively lost, why not look for these directly in the image intensity data? This is the goal of the method presented here, called fine correlation, which extracts both the disparity and its derivatives from the image intensity data. In 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 2). The reference image used for the computation of disparity is the left image.
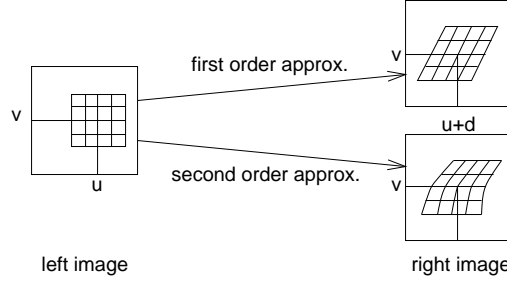
4

Figure 1: To a small square region in the left image corresponds in the right image a sheared and slanted rectangle if we do a first order approximation of the deformation, or a distorted rectangle of the same height at second order.

2.3.1 Principle

The idea comes from the following observation: a small surface element that is viewed as a square window of pixels in the reference image can be seen in the other image as a distorted square. To better understand this, imagine that the left camera is a projector which casts the image of a small square on the 3-D surface, and the righ camera re-projects the image of this square in the right image. Actually, it can easily be shown that the distortion of the square in the other image is closely related to the derivatives of disparity, and we will explain later how these derivatives are related to the differential properties of the surface.

Let us call $d(u,v)$ the disparity at point (u,v), which means that the point in the right image corresponding to (u,v) is $(u + d(u,v), v)$. Let $\alpha(u,v)$ and $\beta(u,v)$ be the derivatives of $d(u,v)$ with respect to u and v, respectively. Then, if we do a first order approximation of the function which maps (u,v) to $(u + d(u,v), v)$, the point corresponding to $(u + \delta_u, v + \delta_v)$ is $(u + d(u,v) + (1 + \alpha)\delta_u + \beta\delta_v, v + \delta_v)$. This means that the region corresponding to a small square centered at (u,v) in the left image is a sheared and slanted parallelogram centered at $(u + d(u,v), v)$ in the right image (Figure 1).

The same scheme can be used to compute a second order approximation of the deformation that operates on a square region in the left image. We obtain that the point corresponding to $(u + \delta_u, v + \delta_v)$ by calculating the Taylor series expansion of $(u,v) \rightarrow (u + d(u,v), v)$ up to second order : $(u + \delta_u, v + \delta_v) \rightarrow (u + d(u,v) + \delta_u(1 + \alpha) + \delta_v\beta + \frac{\delta_u^2}{2}\frac{\partial^2 d}{\partial u^2} + \frac{\delta_u\delta_v}{2}\frac{\partial^2 d}{\partial u\partial v} + \frac{\delta_v^2}{2}\frac{\partial^2 d}{\partial v^2}, v + \delta_v)$

An example of such a deformation is shown in Figure 1. The same calculations can be made for even higher degree deformations.

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 do guess the derivatives of the disparity as the parameters of the deformed element which maximize the correlation between both regions. For example to compute the first derivatives of the disparity (Figure 1) we simply have to compute the values of $d(u,v)$, $\alpha(u,v)$, and $\beta(u,v)$ that maximize the correlation between the square region in the left image and the slanted rectangle region in the right image.

2.3.2 Implementation

The implementation of the proposed method requires the answers to two essential questions: First, how do we compute the correlation between two regions of pixels of different shape? Second, how do we find the parameters that maximize the criterion of correlation?

The first problem was solved by computing the transformed coordinates of each point in the correlation window in the reference image, and doing a linear interpolation of the image intensity in the other image at each of these points. Then, a standard correlation criterion is used (such as zero-mean normalized cross-correlation). This solution is simple to implement, fast, and easily adaptable to any class of deformation, so that we can compute higher degree derivatives with only a few modifications in the code.

For the optimization step, which consists of finding the values of the disparity and its derivatives that maximize the correlation (i.e. minimize the criterion of correlation in our case), we first compute a dense disparity map by a standard correlation technique. Let $p_0 = (d_0, \alpha_0, \beta_0)$ be a vector whose components are respectively the disparity at point (u, v) computed by classical correlation, and its derivatives computed by the local plane fitting method described before. This vector is then used as the initialization for a classical minimization method (we used a Levenberg-Marquard method), which finds the values for the disparity and its first and second derivatives that maximize the correlation score. We repeat this process for each point in the left image and we finally get three maps: one for the disparity itself and two for its first derivatives. Optionally, if we use a second order model, we get three more maps for the second derivatives of disparity. It is to be noted that the derivatives of the disparity are measured locally at point (u, v), and are not guaranteed to be consistent with the derivatives that could be computed from the disparity itself, e.g. by finite differences. In fact, the 3 components (6 for the second order model) should be considered as a separate measurement on the image pair, and a continuous disparity map could be obtained by interpolation between the points at integer image coordinated, using the derivatives computed by fine correlation.

Another possibility would have been to sample the space of disparity and disparity values and compute the disparity for each set of values (as in [LTS94, RH94]). This way, we would not rely on a proper initialization from the classical correlation method, but the dimension of the search space would increase dramatically with the disparity derivation order, and the solution would be less accurate than with a continuous optimization method. Sampling the search space is fully justified for classical correlation, where the continuous correlation score function can be deduced from its values at integer disparities, but not for fine correlation.

3 From derivatives of disparity to 3-D differential properties

3.1 The problem

Let us consider a pair of calibrated cameras in standard geometry. We know the projection matrices P and Q associated respectively to the first and the second camera, and that the epipolar lines are identical in both images: both image planes are parallel to the straight line joining the two optical centers C_1 and C_2 (Figure 2) and that the epipoles e_1 and e_2 are at infinity in the horizontal direction in each image plane. The image coordinates are called u (horizontal coordinate, parallax direction) and v (vertical coordinate, no parallax). We do not lose generality by considering only the standard (or rectified) geometry case, since any stereo pair can be rectified by computing and applying on both images a pair of rectifying homographies.

A 3-D surface (S) is projected on both cameras and what we want to do is to calculate the position, orientation and the curvature of the surface at each point, respectively from the disparity and its first and second derivatives
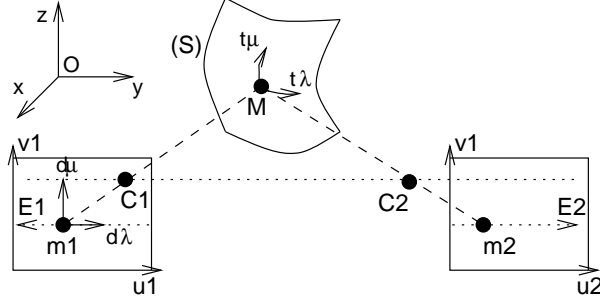
Figure 2: A configuration corresponding to standard geometry

3.2 3-D reconstruction

In standard geometry, the 3×4 projection matrices associated to each camera only differ by their first line:

$$\tilde{\mathbf{P}} = \left[\begin{array}{c} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{array} \right] \text{ and } \tilde{\mathbf{Q}} = \left[\begin{array}{c} \mathbf{q}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{array} \right]$$

Let $\mathbf{m}$ be a point in the projective space $\mathcal{P}^3$, and $\mathbf{m}_1 = \tilde{\mathbf{P}}\mathbf{m} = \alpha(u_1, v_1, 1) \cong (u_1, v_1, 1)$ its projection in the reference image (Figure 2). α is an unknown non-zero scale factor. The projection of $\mathbf{m}$ in the other image can be written as:

$$\mathbf{m}_2 = \tilde{\mathbf{Q}}\mathbf{m} = \alpha(u_2, v_2, 1) \cong (u_2, v_2, 1) = (u_1 + d(u_1, v_1), v_1, 1)$$

so that, for any point $\mathbf{m}$,

$$\left[\begin{array}{c} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{q}_1 - \mathbf{p}_1 \\ \mathbf{p}_3 \end{array} \right] \mathbf{m} = \alpha \left(\begin{array}{c} u_1 \\ v_1 \\ d(u_1, v_1) \\ 1 \end{array} \right)$$

Let $\mathbf{R}$ be the matrix, called *reconstruction matrix*, defined as:

$$\mathbf{R}^{-1} = \left[\begin{array}{c} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{q}_1 - \mathbf{p}_1 \\ \mathbf{p}_3 \end{array} \right]$$

The kernel of the $\mathcal{L}(\mathcal{P}_3, \mathcal{P}_2)$ application represented by $\tilde{\mathbf{P}}$ is the vector space of $\mathcal{P}_3$ generated by $\mathbf{c}$, the optical center of the first camera, and the kernel of $\tilde{\mathbf{Q}}$ is generated by $\mathbf{c}'$, the optical center of the second camera. Since $\mathbf{c}' \neq \mathbf{c}$, the four vectors $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ and $\mathbf{q}_1$ are linearly independant, and the matrix $\mathbf{R}$ is of full rank.

If we associate to each point in the disparity map the point $\mathbf{m}_d$ in $\mathcal{P}^3$, defined as $\mathbf{m}_d = (u_1, v_1, d(u_1, v_1), 1)$, the Euclidean reconstruction of this point is simply:

$$\mathbf{m} = \mathbf{R}\mathbf{m}_d$$

Among other things, this means that the disparity map *is a projective reconstruction* [Fau92], since it is related to a Euclidean reconstruction by a collineation of $\mathcal{P}^3$, represented by matrix $\mathbf{R}$.

7

3.3 Orientation of the observed surface

The equation of the tangent plane to the disparity map at point $\mathbf{m}_d^0 = (u_0, v_0, d_0, 1)$ can be written as:

$$\mathbf{p}_d^T \cdot \mathbf{m}_d = (\alpha, \beta, \gamma, \delta)^T \cdot (u, v, d, 1) = 0 \tag{1}$$

with

$$\mathbf{p}_d = (\alpha, \beta, \gamma, \delta) = \left(\frac{\partial d}{\partial u}(u_0, v_0), \frac{\partial d}{\partial v}(u_0, v_0), -1, d_0 - u_0\alpha - v_0\beta \right)$$

The matrix $\mathbf{R}$ being of full rank, we can rewrite eq. 1 as:

$$\mathbf{p}_d^T \mathbf{R}^{-1} \mathbf{R} \mathbf{m}_d = 0$$

or, since $\mathbf{m} = \mathbf{R}\mathbf{m}_d$,

$$\left(\mathbf{R}^{-T} \mathbf{p}_d \right)^T \mathbf{R}\mathbf{m} = 0. \tag{2}$$

Equation 2 is the equation of the tangent plane to the reconstructed Euclidean surface, which is represented by the element of the dual projective space $\mathbf{p}$, with:

$$\mathbf{p} = \mathbf{R}^{-T} \mathbf{p}_d$$

This shows that the equation of the tangent plane to the reconstruction can be easily calculated from the disparity and its first derivatives. Next section shows that this also holds for the curvatures of the Euclidean reconstruction.

3.4 Curvatures of the surface

The same way we used the equation of the tangent plane to the disparity surface to find the tangent plane to the reconstruction, we now use the equation of the osculating quadric to deduce the curvatures of the reconstruction.

The second order Taylor expansion of the disparity function at point (u_0, v_0) can be written:

$$\begin{aligned} d(u, v) = {} & d + (u - u_0)p + (v - v_0)q \\ & + \frac{(u - u_0)^2}{2}r + (u - u_0)(v - v_0)s + \frac{(v - v_0)^2}{2}t, \\ & + o((u - u_0)^2 + (v - v_0)^2) \end{aligned}$$

where we use the classical Monge notations for the partial derivatives:

$$d = d(u_0, v_0) \qquad\qquad p = \frac{\partial d}{\partial u}(u_0, v_0) \qquad\qquad q = \frac{\partial d}{\partial v}(u_0, v_0)$$

$$r = \frac{\partial^2 d}{\partial u^2}(u_0, v_0) \qquad\qquad s = \frac{\partial^2 d}{\partial u \partial v}(u_0, v_0) \qquad\qquad t = \frac{\partial^2 d}{\partial v^2}(u_0, v_0)$$

This leads to an equation for the osculating quadric of the disparity map at $\mathbf{m}_0 = (u_0, v_0, d(u_0, v_0), 1)$:

$$\mathbf{m}_d^T \mathbf{Q}_d \mathbf{m}_d = 0, \tag{3}$$

with $\mathbf{m}_d = (u, v, d, 1)$ and

$$\mathbf{Q}_d = \begin{bmatrix} r & s & 0 & p - ru_0 - sv_0 \\ s & t & 0 & q - su_0 - tv_0 \\ 0 & 0 & 0 & -1 \\ p - ru_0 - sv_0 & q - su_0 - tv_0 & -1 & \begin{matrix} 2(d - pu_0 - qv_0) \\ +ru_0^2 + 2su_0v_0 + tv_0^2 \end{matrix} \end{bmatrix}$$

The osculating quadric (S_d) to the disparity map in $\mathcal{P}^3$ is defined as the kernel of the bilinear symmetric form $\mathbf{Q}_d \in Bilsym(\mathcal{P}^3)$.

The reconstruction matrix $\mathbf{R}$ being of full rank, we can rewrite eq. 3 as:

$$\mathbf{m}_d^T \mathbf{R}^T \mathbf{R}^{-T} \mathbf{Q}_d \mathbf{R}^{-1} \mathbf{R} \mathbf{m}_d = 0$$

or, since $\mathbf{m} = \mathbf{R} \mathbf{m}_d$,

$$\mathbf{m}^T \mathbf{Q} \mathbf{m} = 0 \qquad (4)$$

with

$$\mathbf{Q} = \mathbf{R}^{-T} \mathbf{Q}_d \mathbf{R}^{-1}$$

The collineation associated with $\mathbf{R}$ being at least C^2 (in fact, C^∞) on its definition domain, and the osculating quadric defined by equation 3 being tangent at second order to the disparity map, the quadric defined by equation 4 is osculating to the Euclidean reconstruction (S).

Let $\mathbf{U}$, $\mathbf{v}$, w, $\mathbf{T}$ and $\mathbf{m}'$ be defined by the identities:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{U} & \mathbf{v} \\ \mathbf{v}^T & w \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & t \end{bmatrix}, \text{ and } \mathbf{m}' = \begin{bmatrix} \mathbf{U} & \mathbf{v} \end{bmatrix} \mathbf{m}$$

Then it can be shown [Dev97] that the Gaussian curvature K and the mean curvature H at a point $\mathbf{m}$ on the quadric are:

$$K = -\frac{\det\left(\mathbf{T}^T \mathbf{Q} \mathbf{T}\right)}{\|\mathbf{m}'\|^4}$$

$$H = -\frac{1}{2}\left(\frac{\mathbf{m}'^T \mathbf{U} \mathbf{m}'}{\|\mathbf{m}'\|^3} + \frac{\operatorname{trace}\left(\mathbf{U}\right)}{\|\mathbf{m}'\|}\right)$$

4 Results

4.1 The experimental system

The results presented here were obtained using the different techniques described in this paper. The stereoscopic system we used consists of a pair of CCD cameras. The image resolution of our system is 512×512, and the subject of our stereograms were several textured objects representing moulds of human faces, human torsos, or vases, and real faces (Figure 3). The system was calibrated using a calibration grid.

4.2 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 desired order. Some results for the two methods presented in this paper are shown and compared.

Figure 3: Two cross-eyed stereograms: a mold of torso and a real face. Images are presented so that parallax is horizontal.

4.2.1 By plane fitting

Some results of the application a standard correlation method followed by plane fitting are shown in figure 4. 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 accuracy and localization of surface features is lost. A solution would be to fit a higher degree model or do some regularization before processing the data.

4.2.2 By fine correlation

The method based on correlation 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 accurate because the local image deformations are taken into account. The first derivatives of disparity seem also to be accurate, especially when compared to those obtained by plane fitting (Figure 5), but some strange phenomenon occurs 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 v (Figure 6). After verification, the amplitude of these variations corresponds to the pixel jitter value given in the technical data of the acquisition system, which causes a random horizontal shift of each video line of up to 0.5 pixel, averaged over the height of the correlation window. Consequently, these are only due to the limits of the acquisition system.

4.3 Reconstruction and surface orientation

The 3D reconstructions obtained from two different dense disparity maps, one obtained by a standard correlation algorithm [Fua91], and one refined by fine correlation, are shown Figure 7,
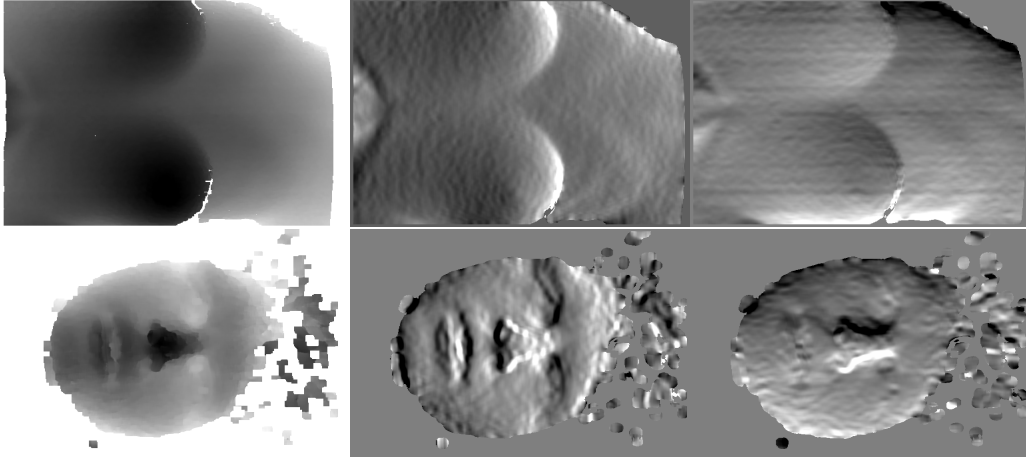
Figure 4: The disparity field obtained by standard correlation (left) and the computation of first derivatives by plane fitting (center and right).
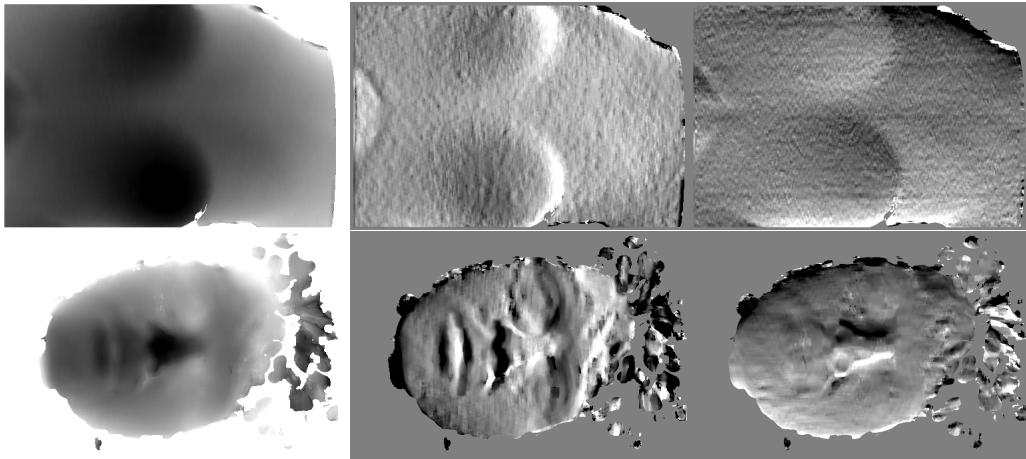

Figure 5: The disparity field (left) and the first derivatives of disparity obtained by fine correlation (center and right).
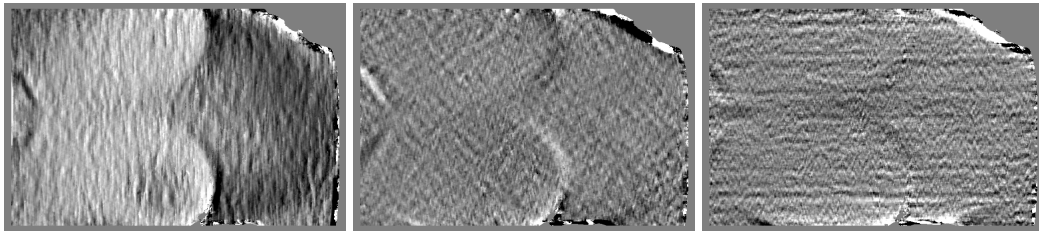

Figure 6: The second derivatives of the disparity: $\frac{\partial d}{\partial u^2}$ (left), $\frac{\partial d}{\partial u \partial v}$ (center), and $\frac{\partial d}{\partial v^2}$ (right).
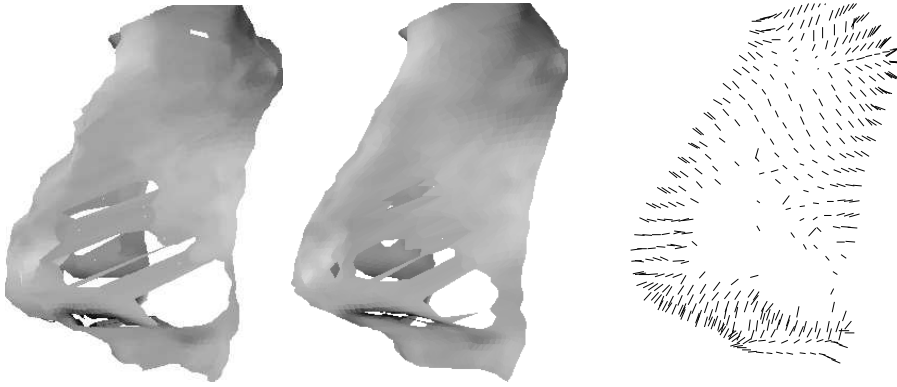
11

Figure 7: The reconstruction of the nose from the pair of Figure 3: using standard correlation (left), using enhanced correlation (center), and the field of normals (right). This represents a 100×100 pixels region, where the amplitude of the variation of disparity is less than 10 pixels.
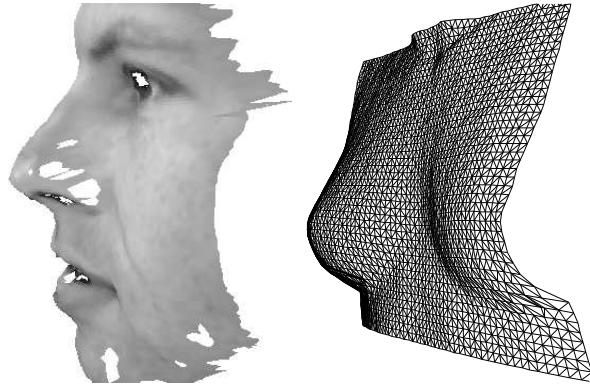

Figure 8: The face reconstruction with intensity mapping and the bust reconstruction, using enhanced correlation.

together with a subsampling of the field of normals that was obtained together with the disparity map by fine correlation. The whole face is represented in Figure 8.

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. Another example of the field of normals is shown in Figure 9.

We also calculated the Gaussian curvature and the mean curvature of the torso stereogram. The problem is that the stripes that appeared in the second derivative of the disparity over v are still present, so that there is too much error in the curvature maps. Hopefully, further experiments with digital cameras (which do not have the pixel jitter problem) should give better results.
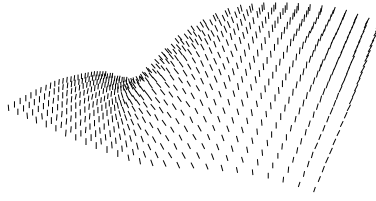
Figure 9: The field of normals of a 100×100 region of the bust stereogram. There are 4 pixels between two samples.

5 Error analysis

In this section, we analyse the error on disparity and its derivatives for the classical and fine correlation methods, using ground-truth synthetic data representing an ideal case for surface-based stereoscopy.
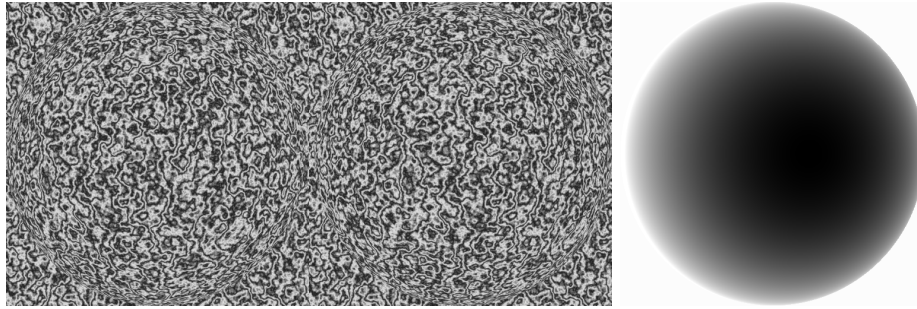

Figure 10: The synthetic image pair (cross-eyed stereogram) and the corresponding disparity map.

In fact, the synthetic pair (Figure 10) is a raytracer-generated scene (a half sphere put on a plane) for which we can compute exactly the disparity and its derivatives. Since it is a half sphere, all possible orientations are present in the scene, which makes it a good benchmark for computing first derivatives of disparity. This scene also has several properties that makes it an very nice candidate for stereoscopy by correlation :

- There are no depth discontinuities, except at the edge of the sphere,

- there is almost no intensity noise (only intensity sampling noise), since it is generated by ray-tracing,

- the texture we use is isotropic (it is a parametric texture), has high dynamic, and the surface is fully Lambertian.

For these reasons, the results presented here should be considered as a case study made in the best possible conditions: the numerical results are better than they would be on a real scene, but the method and qualitative results will probably hold for any stereo pair for which we have ground truth.

First, we should define what error measurement we use. What we call the disparity error in the following is not the difference between the analytical disparity and the computed disparity at one point in the reference image: the reference point is the point for which the analytical 3-D position is closest to the 3-D reconstruction of a given point in the disparity map. Once again, since we are studying a sphere, this point can be found easily by projecting the reconstruction on the sphere. This choice was guided by the fact that algorithm performance and accuracy should be considered with respect to 3-D reconstruction error, though it introduced 3-D geometry knowledge in a purely image-based error analysis.

A naive approach to error analysis in the case of stereo by correlation would be to consider the error on disparity or its derivatives as Gaussian, and just compute the mean and standard deviation of these errors. Unfortunately, the error is not Gaussian at all, simply because there are two separate random processes underlying correlation error: The first one is whether the stereo method hits or misses, i.e. if we got a false match or if we found the right correspondence. The second is the intrinsic accuracy of the method, i.e. how accurate is this correspondence between the point at the center of the correlation window in the reference image and a point in the other image.

If we suppose, as a first approximation, that each of these two processes are Gaussian, which they certainly aren't, then the error distribution can be modeled as a mixture of two Gaussian, which is a weighted sum of Gaussian. For a given error distribution, the maximum likehood mixture of Gaussian was computed using the EM algorithm [MK96], which proved to be quite simple and effective in our case. In fact, Figure 11 show how bad the disparity error distribution is modeled by a Gaussian distribution, and how well it is modeled by a mixture of two Gaussian.
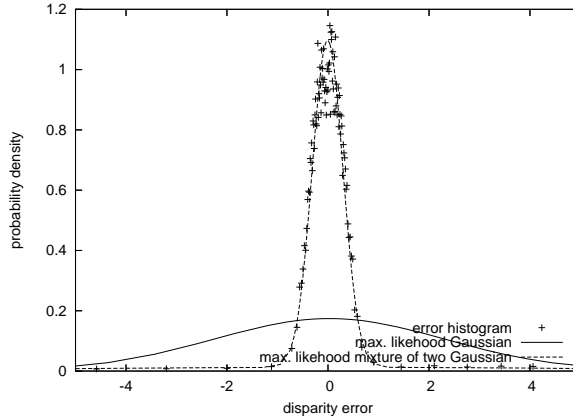

Figure 11: Modeling the error distribution by a mixture of Gaussian: histogram (crosses) for classical correlation (window width=7, disparity slope=0.8), maximum likehood Gaussian model (solid line), and maximum likehood mixture of two Gaussian model (dotted line).

It is very easy to tell which Gaussian represents which process: the narrow Gaussian, which has the highest weight, represents the intrinsinc accuracy, and the wide Gaussian represents false matches. The size and position of the narrow Gaussian represent the accuracy and bias of the correlation method. The parameters of the wide Gaussian aren't really meaningful, since they mainly depend on the choice of the disparity interval, but its weight in the mixture gives us another valuable information: the proportion of false matches.

5.1 On classical correlation

Because of the fronto-parallel approximation made by classical correlation, the error distribution depends highly on surface orientation, which is related to first derivatives of disparity. In fact, we found that a dimensioning parameter was disparity slope, $\sqrt{(\partial d/\partial u)^2 + (\partial d/\partial v)^2}$, as shown Figure 12. All the error measurements were classified by slope, and for each slope a maximum likehood mixture of Gaussian was computed for the error distribution.
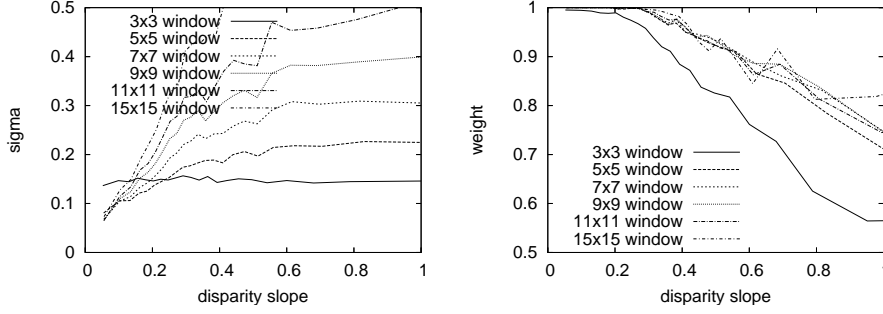

Figure 12: σ of the main Gaussian (left) and its weight (right) for the classical correlation method. Intrinsic correlation error increase with disparity slope. Proportion of false matches is the same for window widths greater than 7.

The same statistical analysis performed on various correlation window sizes shows that the correlation error increases with the window size. This is due to the fact that no particular point of the correlation window is privileged, so that the measured disparity can be this of a point at the edge of the correlation window. This is confirmed by the asymptotic values of the disparity error curves, which seem to be proportional to the correlation window size. But using a larger correlation window also means getting less false matches, so we must find a compromise between accuracy and proportion of false matches. Fortunately enough, this compromise is easy to find, since using correlation windows larger than 7 pixels doesn't seem to decrease significantly the proportion of false matches. This analysis confirms that we should use a window size of 7 pixels, which is empirically known by many as the "good" window size for correlation.

A prerequesite for fine correlation is computing initial values for the derivatives of disparity, e.g. by bilinear regression (section 2.2). Before this process, we filter the disparity map using morphological operations, in order to reduce again the proportion of false matches (Figure 13) : fine correlation will only succeed if the initial disparity is good enough.

5.2 On fine correlation

The disparity and its derivatives are initialized from the results of bilinear regression, and are optimized by fine correlation. We first must check that this initialization method is satisfactory and that the following results on fine correlation do not depend on the initialization scheme: a bad initialization should generate more false matches, and thus modify the weight of the main Gaussian, but the intrinsinc accuracy of the method –measured using the width and position of the main Gaussian– should not change. To check this, we ran the fine correlation initialized by the exact disparity and derivatives values, and we compared the parameters of the main Gaussian with those obtained after initialization by bilinear regression. The results (Figure 14) show that
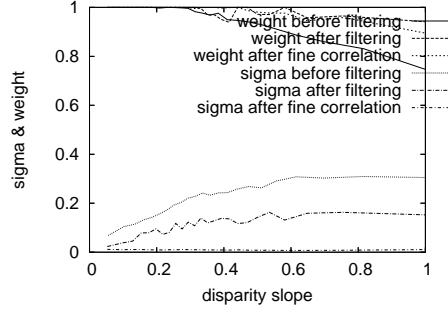
15

Figure 13: σ of the main Gaussian and its weight after filtering and bilinear regression.

initialization does not change the width and position of the main Gaussian: our methodology for error analysis does not depend on the initialization method.
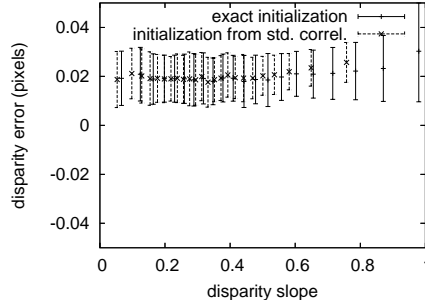

Figure 14: Error bars for the main Gaussian, using fine correlation with different initialization schemes: from exact disparity and derivatives value, and from standard correlation using bilinear regression. Initialization does not influence the main Gaussian, only the proportion of false matches.

As we did with classical correlation, we can try to guess the best correlation window size from the proportion of false matches and the width of the first Gaussian mode (Figure 15). Using the same arguments as for classical correlation, we find a best correlation window width of 11 pixels for fine correlation with first order derivatives, and 15 pixels for fine correlation with second order derivatives. From these figures, a good rule of thumb to compute the best window size seems to be that the number of pixels in the window (or its surface) should be approximately proportional to the number of parameters of the deformation model: 1 for classical correlation, 3 for first order fine correlation, 6 for second order fine correlation, which give window sizes of respectly 7×7, 11×11 and 15×15 pixels.

One noticeable fact is that not only the standard deviation of this Gaussian is much lower than for classical correlation, but it is also almost constant with disparity slope: only the proportion of false matches increases with slope. The very small bias on disparity is due to the fact that the surface is non-planar (the first order fine correlation makes a locally planar approximation of the surface) and the curvature of the sphere is not negligible. This bias disappears when we use second order fine correlation (Figure 16), since the locally planar approximation is replaced

16

by locally quadratic. Figures 15 and 16 show that the accuracy obtained by fine correlation *on this scene* is better than 1/50 pixel of disparity.
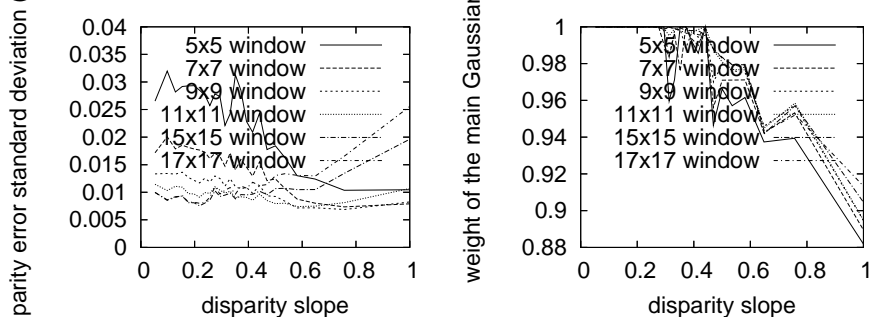

Figure 15: σ of the main Gaussian (left) and its weight for fine correlation, to be compared with Figure 12.
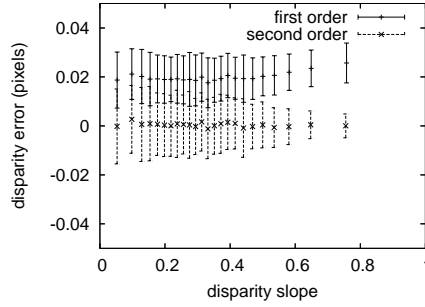

Figure 16: Error bars for the main Gaussian, using fine correlation and a first order (with a 11×11 window) or second order model (with a 15×15 window). The standard deviation is almost identical: both have an accuracy better than 1/50 pixel. First order model causes a small bias because of surface curvature, whereas second order model gives no bias.

5.3 First derivatives of disparity

Since fine correlation computes both disparity and its derivatives, and we have ground truth data for the derivatives, we can also measure the accuracy of these derivatives. A first comparison can be made between the values used for the initialization, which come from the bilinear regression method described before, and the output of fine correlation. Figure 17 shows the accuracy of the horizontal derivative of disparity before and after fine correlation, and proves that fine correlation computed derivatives of disparity properly, even for high disparity slopes.

Figure 18 shows how the standard deviation of the error on horizontal and vertical derivatives of disparity vary with the correlation window size, and confirms that using a window as small as 11×11 pixels already gives a very good accuracy. The bias on both derivatives is almost zero, since the surface we use is approximately locally quadratic. The second derivatives of disparity, as computed by second order fine correlation, can not be properly analysed on this synthetic scene.
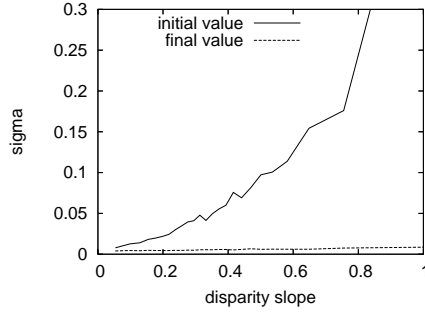
17

Figure 17: Standard deviation of initial value for first derivative computed by standard correlation with a 7×7 correlation window followed by bilinear regression on a 7×7 area (which makes a total 13×13 neighborhood in the initial images), and final value after fine correlation using a 11×11 window.

In fact, the synthetic scene does not cover a wide range of curvature (or second derivatives) values, and a collection of scenes should be used instead, each having different properties (e.g. given values for the second derivatives of disparity). Then comes the difficult problem of analysing the coupling between the effects of first derivatives of disparity (i.e. slope) and its second derivatives (curvatures): the second derivatives will certainly be more difficult to extract on a steep surface than on an almost fronto-parallel surface. This problem may be adressed in the future.
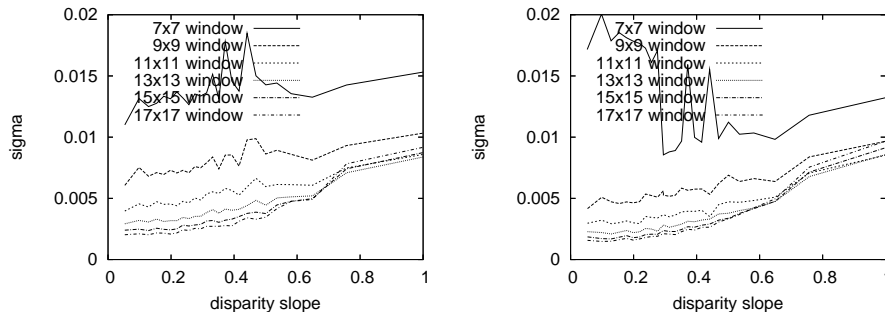

Figure 18: Standard deviation of main Gaussian for disparity derivatives computed by fine correlation: derivative over u (left) and derivative over v (right).

6   Conclusion

We described in this paper 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 method, called fine correlation, is more accurate than classical methods but of course computationally more expensive. Besides, it depends on proper initialization by another method,

18

such as classical correlation or stretch-and-shear correlation. The difference between fine correlation and stretch-and-shear correlation methods – which are close in mind [LTS94, RH94] – is that the latter only searches a finite set of disparity and orientation values, whereas our method does an optimization in the continuous space of disparity and its derivatives, thus giving more accurate results.

The error on disparity and its derivatives was studied on ideal synthetic data, and this lead to several observations. The disparity error distribution was modeled by a mixture of Gaussian, which enabled us to separate the intrinsic disparity error from the errors generated by false matches. We were able to point out, for each method, what is the best correlation window size for that particular synthetic scene. Finally, we showed that the intrinsic disparity error increases with the disparity slope for classical correlation, whereas it keeps almost constant for fine correlation, which is the most important and most encouraging result.

Future work will be mainly on better modeling and understanding the disparity error for both the classical and fine correlation methods. The synthetic case studied for that purpose in this paper is an ideal stereoscopic pair, and our goal is to establish how disparity error varries with degraded image or scene properties: intensity noise, local or global texture properties (dynamic, orientation, repeating patterns), non-lambertian surfaces, local differential properties of surfaces, occlusions and discontinuities. Moreover, we would like to be able to predict *a priori* a disparity error probability at each point, given local and global image measurements. Hopefully, this will lead to a better understanding of the correlation-based stereo algorithms, which could be crucial for highly sensitive domains where stereo could be used (such as military or medical applications).

References

[Bel93]   Peter N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *Proceedings of the 4th International Conference on Computer Vision*, Berlin, Germany, May 1993. IEEE Computer Society Press. 2

[BPYH85] Michael Brady, Jean Ponce, Alan Yuille, and Asada H. Describing surfaces. *Computer Vision, Graphics, and Image Processing*, 32:1–28, 1985. 2

[BZ87]   Andrew Blake and Andrew Zisserman. *Visual Reconstruction*. MIT Press, 1987. 2

[Dev97]   Frédéric Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, École Polytechnique, Palaiseau, France, February 97. 9

[DF94]   Frédéric Devernay and Olivier Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 208–213, Seattle, WA, June 1994. IEEE. 2

[EW87]   R.D. Eastman and A.M. Waxman. Using disparity functionals for stereo correspondence and surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 39:73–101, 1987. 2

[Fau92]   Olivier Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 563–578, Santa Margherita, Italy, May 1992. Springer-Verlag. 2, 3, 7

[Fau95]   Olivier Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, 12(3):465–484, March 1995. 2

[Fua91]   P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. In *International Joint Conference on Artificial Intelligence*, Sydney, Australia, August 1991. 10

[Gri81]   W.E.L. Grimson. *From Images to Surfaces*. MIT Press : Cambridge, 1981. 2

[HA87]   W. Hoff and N. Ahuja. Extracting surfaces from stereo images. In *Proceedings of the 1st International Conference on Computer Vision*, London, England, June 1987. IEEE Computer Society Press. 2, 4

[JM92]   D.G. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proceedings of the 2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag. 2

[KvD76]   J.J. Koenderink and A.J. van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29–35, 1976. 2

[LTS94]   R. A. Lane, N. A. Thacker, and N. L. Seed. Stretch-correlation as a real-time alternative to feature-based stereo matching algorithms. *Image and Vision Computing*, 12(4):203–212, May 1994. 2, 6, 19

[MK96]   Geoffrey J. McLachlan and Thiriyambakam Krishnan. *The EM algorithm & extensions*. John Wiley & Sons, 1996. 14

[MN84]   G. Medioni and R. Nevatia. Description of three-dimensional surfaces using curvature properties. In L.S. Baumann, editor, *Proceedings of the ARPA Image Understanding Workshop*, pages 219–229, New Orleans, LA, October 1984. Defense Advanced Research Projects Agency, Science Applications International Corporation. 2

[PMF85]   S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. PMF : a stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985. 4

[RH94]   L. Robert and M. Hebert. Deriving orientation cues from stereo images. In J-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800-801 of *Lecture Notes in Computer Science*, pages 377–388, Stockholm, Sweden, May 1994. Springer-Verlag. 2, 6, 19

[SFB96]   Charles Stewart, Robin Flatland, and Kishore Bubna. Geometric constraints and stereo disparity computation. *The International Journal of Computer Vision*, 20(3):143–168, 1996. 2

[Ter86]   Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:413–424, 1986. 2

[Wil91]   Richard P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):761–774, August 1991. 2