Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach

Luis Alvarez [*]
Depart. de Informática y Sistemas
Universidad de Las Palmas

Rachid Deriche [†]
Comput. Vision Laboratory, Robotvis Group
INRIA Sophia-Antipolis

Javier Sánchez [‡]
Depart. de Informática y Sistemas
Universidad de Las Palmas

Joachim Weickert [§]
Depart. of Mathemat. and Comput. Science
University of Mannheim

Abstract

We present an energy based approach to estimate a dense disparity map from a set of two weakly calibrated stereoscopic images while preserving its discontinuities resulting from image boundaries. We first derive a simplified expression for the disparity that allows us to estimate it from a stereo pair of images using an energy minimization approach. We assume that the epipolar geometry is known, and we include this information in the energy model. Discontinuities are preserved by means of a regularization term based on the Nagel–Enkelmann operator. We investigate the associated Euler–Lagrange equation of the energy functional, and we approach the solution of the underlying partial differential equation (PDE) using a gradient descent method. The resulting parabolic problem has a unique solution. In order to reduce the risk to be trapped within some irrelevant local minima during the iterations, we use a focusing strategy based on a linear scale-space. Experimental results on both synthetic and real images are presented to illustrate the capabilities of this PDE and scale-space based method.

1 Introduction

Energy based methods have been extensively used in the last years [4, 5, 6, 8, 9, 11, 12] for estimating the disparity map between images. The goal of this paper is to present a variational approach to recover a dense disparity map from a set of two *weakly calibrated* stereoscopic images. To solve this problem, we first make full use of the knowledge of the so-called *fundamental matrix* [7] to derive the equations that relate corresponding pixels in the two views. The intrinsic and extrinsic parameters of the camera are not known. We directly compute the disparity map from the grey-level image intensities without dealing with any intermediate process such as rectification and we address the problem of accurately determining the dense disparity map while smoothing and regularizing this disparity map along the contours of the grey level image and inhibiting smoothing across the image discontinuities.

The preservation of discontinuities in the disparity map is obtained using an anisotropic *linear* operator [3, 10] which allows to develop discontinuities in the disparity map across the edges of one of the 2 images. This important step is achieved by considering a well adapted regularization term that has already proven to be very useful in optical flow estimation.

In order to avoid converging to irrelevant minima, a focusing strategy embedding our method in a linear scale-space is used, as it has already been successfully applied in optical flow estimation [3]. The coarse-scale solution serves then as initial data for solving the problem at a finer scale. We have shown that our method leads to a mathematically correct concept and we have proven the existence and uniqueness of the solution of the parabolic equation which governs our method.

Finally, our approach has been validated on a large set of synthetic and real stereo data. All these results are presented in a previous work [1].

[*]Address: Edificio de Informática y Sistemas, Campus Universitario de Tafira. SP-35017 Las Palmas, Spain. E-mail: lalvarez@dis.ulpgc.es

[†]Address: 2004 Route des Lucioles. F-06902 Sophia-Antipolis, France. E-mail: Rachid.Deriche@sophia.inria.fr

[‡]Address: Edificio de Informática y Sistemas, Campus Universitario de Tafira. SP-35017 Las Palmas, Spain. E-mail: jsanchez@dis.ulpgc.es

[§]Address: D-68131 Mannheim, Germany. E-mail: Joachim.Weickert@ti.uni-mannheim.de

2 Formalism of the matching process

2.1 Notation and Background

In this paper we use a projective camera model. This model maps a 3D point $\mathbf{M} = [X, Y, Z]^t$ to a 2D image point $\mathbf{m} = [x, y]^t$ through a 3×4 projection matrix $\mathbf{P}$ via $s\hat{\mathbf{m}} = \mathbf{P}\hat{\mathbf{M}}$, where s is a nonzero scale factor and the notation $\hat{\mathbf{p}}$ is such that if $\mathbf{p} = [x, y, \ldots]^t$ then $\hat{\mathbf{p}} = [x, y, \ldots, 1]^t$. In the case of two images acquired by a binocular stereo system, every physical point $\mathbf{M}$ in space yields a pair of 2D projections $\mathbf{m_1}$ and $\mathbf{m_2}$ on the two images. The 3×4 projection matrices are defined by the following relations:

$$s_1 \hat{\mathbf{m}}_1 = \mathbf{P_1}\hat{\mathbf{M}} \qquad s_2 \hat{\mathbf{m}}_2 = \mathbf{P_2}\hat{\mathbf{M}} \qquad (1)$$

Assuming that the world coordinate system is associated with the first camera, the two projection matrices are given by

$$\mathbf{P_1} = [\mathbf{A}|\mathbf{0}] \qquad \mathbf{P_2} = [\mathbf{A}'\mathbf{R}|\mathbf{A}'\mathbf{t}] \qquad (2)$$

where $\mathbf{R}$ and $\mathbf{t}$ represent the 3×3 rotation matrix and the 3×1 translation vector defining the rigid displacement between the two cameras, and $\mathbf{0}$ denotes the 3×1 null vector. The matrices $\mathbf{A}$ and $\mathbf{A}'$ are the 3×3 intrinsic parameters matrices of the two views, each depending on five parameters and having the following well-known form [7]:

$$\mathbf{A} = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 \\ 0 & \alpha_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

All these matrices and parameters can be computed with good accuracy by means of a classical calibration method [7]. In such a case, the system is said to be *calibrated*.

By eliminating the scalars s_1 and s_2 associated with the projection equations (1) as well as the point $\mathbf{M}$, an equation relating the pair of projections of the same 3D point is obtained:

$$\hat{\mathbf{m}}_2{}^t \mathbf{F} \hat{\mathbf{m}}_1 = 0. \qquad (3)$$

For a point $\mathbf{m_1} = [x, y]^t$ in the first image I_l, the fundamental matrix $F = (f_{i,j})$ provides the epipolar line Δ of equation $\hat{\mathbf{m}}_2{}^t \mathbf{F}\hat{\mathbf{m}}_1 = 0$ in the image I_r. Let us introduce the notation

$$a(x, y) \quad := \quad f_{11}x + f_{12}y + f_{13}, \qquad (4)$$
$$b(x, y) \quad := \quad f_{21}x + f_{22}y + f_{23}, \qquad (5)$$
$$c(x, y) \quad := \quad f_{31}x + f_{32}y + f_{33}. \qquad (6)$$

Using this notation, the epipolar line Δ can be written as

$$a(x, y)x' + b(x, y)y' + c(x, y) = 0. \qquad (7)$$

We will use this equation in order to introduce our specific parameterization of the disparity, developed to yield a simple linear second order differential operator in the minimization part.

2.2 The Disparity Term

Under the *Lambertian* assumption that corresponding pixels have equal grey values, the determination of the disparity from the stereo pair comes down to finding a function $\mathbf{h}(x, y) := (u(x, y), v(x, y))^t$ such that $I_l(x, y) = I_r(x + u(x, y), y + v(x, y))$, where the point $(x', y') = (x + u(x, y), y + v(x, y))$ belongs to the epipolar line associated to (x, y).

Let us denote by $\mathbf{N}$ (resp. $\mathbf{T}$) the unitary normal (resp. tangential) vector of the epipolar line Δ given by the equation $\hat{\mathbf{m}}_2{}^t \mathbf{F}\hat{\mathbf{m}}_1 = 0$, and by $\mathbf{D}$ the unitary disparity vector associated to the point $\mathbf{m_1}$ and $\mathbf{m_2}$

$$\mathbf{N} = \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} \\ \frac{b}{\sqrt{a^2+b^2}} \end{bmatrix} \mathbf{T} = \begin{bmatrix} \frac{-b}{\sqrt{a^2+b^2}} \\ \frac{a}{\sqrt{a^2+b^2}} \end{bmatrix}$$
$$\mathbf{D} = \begin{bmatrix} \frac{u}{\sqrt{u^2+v^2}} \\ \frac{v}{\sqrt{u^2+v^2}} \end{bmatrix} \qquad (8)$$

Then we have

$$\mathbf{m_2} = \mathbf{m_1} + \delta\mathbf{D} = \mathbf{m_1} - \gamma\mathbf{N} - \lambda\mathbf{T} \qquad (9)$$

where $\delta = \sqrt{u^2 + v^2}$ represents the disparity. γ represents the distance (modulus a sign) of the point $\mathbf{m_1}$ to its epipolar line Δ in the second image and λ represents the distance (modulus a sign) of $\mathbf{m_0}$, the projection of the point $\mathbf{m_1}$ on the epipolar line Δ, to the point $\mathbf{m_2}$ that lies along the epipolar line Δ.

2.3 The Energy Functional

Let us now develop an approach to accurately estimate the $\lambda(x, y)$ function associated to a pair of stereoscopic images. The easiest possibility would be to proceed in a classical way and try to recover this important information using a simple correlation scheme. Unfortunately, this naive solution will not provide a correct and accurate result, in particular in the regions where the disparity map has discontinuities, as is often the case at image edges. It is well known that the disparity map of this classical method tends to be very smooth across the image boundaries. The idea we would like to formalize and develop here is to estimate a $\lambda(x, y)$ function which is smooth only along the image boundaries and not across them. This leads us to consider the minimization of the following energy functional:

$$E(\lambda) = \int_\Omega (I_l(x, y) - I_r(\mathbf{x} + \mathbf{h}(\lambda(x, yen))))^2 \, dx \, dy$$
$$+ \quad C \int_\Omega \nabla\lambda^t \, D(\nabla I_l) \, \nabla\lambda \, dx \, dy \qquad (10)$$

Here, Ω denotes the image domain, C is a positive constant, and $D(\nabla I_l)$ is a regularized projection matrix perpendicular to ∇I_l:

$$D\left(\nabla I_l\right) = \frac{\left\{\begin{bmatrix} \frac{\partial I_l}{\partial y} \\ -\frac{\partial I_l}{\partial x} \end{bmatrix}\begin{bmatrix} \frac{\partial I_l}{\partial y} \\ -\frac{\partial I_l}{\partial x} \end{bmatrix}^t + \nu^2 Id\right\}}{|\nabla I_l|^2 + 2\nu^2} \quad (11)$$

where Id denotes the identity matrix. This projection has been introduced by Nagel and Enkelmann in the context of optic flow estimation. We use it here because of its simplicity (the underlying second order differential operator is linear) and because this method has demonstrated its performance numerous times in the context of optical flow estimations [3].

2.4 Minimizing the Energy

In order to minimize our energy functional, we solve its associated Euler–Lagrange equation

$$C \operatorname{div}\left(D\left(\nabla I_l\right) \nabla\lambda\right)$$
$$+ \left(I_l - I_r^\lambda\right) \frac{a\left(\frac{\partial I_r}{\partial y}\right)^\lambda - b\left(\frac{\partial I_r}{\partial x}\right)^\lambda}{\sqrt{a^2 + b^2}} = 0$$

We obtain a solution of the above equation by calculating the asymptotic state $(t \to \infty)$ of the parabolic equation

$$\begin{aligned} \frac{\partial\lambda}{\partial t} &= C\operatorname{div}\left(D\left(\nabla I_l\right)\nabla\lambda\right) \\ &+ \left(I_l - I_r^\lambda\right)\frac{a\left(\frac{\partial I_r}{\partial y}\right)^\lambda - b\left(\frac{\partial I_r}{\partial x}\right)^\lambda}{\sqrt{a^2+b^2}} \quad (12) \end{aligned}$$

We observe that in this diffusion–reaction method the matrix $D(\nabla I_l)$ plays the role of a diffusion tensor. Its eigenvectors are $v_1 := \nabla I_l$ and $v_2 := \nabla I_l^\perp$, and the corresponding eigenvalues are given by

$$\begin{aligned} \lambda_1(|\nabla I_l|) &= \frac{\nu^2}{|\nabla I_l|^2 + 2\nu^2} \\ \lambda_2(|\nabla I_l|) &= \frac{|\nabla I_l|^2 + \nu^2}{|\nabla I_l|^2 + 2\nu^2} \end{aligned}$$

In the interior of objects we have $|\nabla I_l| \to 0$, and therefore $\lambda_1 \to 1/2$ and $\lambda_2 \to 1/2$. At ideal edges where $|\nabla I_l| \to \infty$, we obtain $\lambda_1 \to 0$ and $\lambda_2 \to 1$. Thus, we have isotropic behavior within regions, and at image boundaries the process smoothes anisotropically along the edge.

2.5 A Linear Scale-Space Approach to Recover Large Disparities

In general, the Euler–Lagrange equation (2.4) will have multiple solutions. As a consequence, the asymptotic state of the parabolic equation, which we use for approximating the disparity, will depend on the initial data. Typically, we may expect that the algorithm converges to a local minimum of the energy functional (10) that is located in the vicinity of the initial data. To avoid convergence to irrelevant local minima, we embed our method into a linear scale-space framework [13]. Considering the problem at a coarse scale avoids that the algorithm gets trapped in physically irrelevant local minima.

The basic idea of embedding our method in linear scale-space is as follows : we replace the images I_l and I_r by $I_l^\sigma := G_\sigma * I_l$ and $I_r^\sigma := G_\sigma * I_r$, where $*$ is the convolution operator, and G_σ denotes a Gaussian with standard deviation σ. We start with a large initial scale σ_0. Then we compute the disparity λ_{σ_0} at scale σ_0 as the asymptotic state of the solution using some initial approximation (see below). Next, we choose a number of scales $\sigma_n < \sigma_{n-1} < \dots < \sigma_0$, and for each scale σ_i we compute the disparity λ_{σ_i} as the asymptotic state of the above PDE with initial data $\lambda_{\sigma_{i-1}}$. The final disparity corresponds to the smallest scale σ_n. In accordance with the logarithmic sampling strategy in linear scale-space theory, we choose $\sigma_i := \eta^i \sigma_0$ with some decay rate $\eta \in (0,1)$. A detailed analysis of the usefulness of such a focusing strategy in the context of a related optic flow problem can be found in [3].

Our method is governed by the evolution equation

$$\begin{aligned} \frac{\partial\lambda_\sigma}{\partial t} &= C\operatorname{div}\left(D\left(\nabla I_l^\sigma\right)\nabla\lambda_\sigma\right) \\ &+ \left(I_l^\sigma - I_r^{\sigma,\lambda_\sigma}\right)\frac{a\left(\frac{\partial I_r^\sigma}{\partial y}\right)^{\lambda_\sigma} - b\left(\frac{\partial I_r^\sigma}{\partial x}\right)^{\lambda_\sigma}}{\sqrt{a^2+b^2}} \quad (13) \end{aligned}$$

For the initial value $\lambda_0^{\sigma_0}(x,y)$ we consider two possibilities. The first one is to take a constant value which depends, in general, on a rough *a priori* estimation of the expected disparity, following an estimation of the depth where the interesting objects are located.

3 Experimental Results

In Figure 1 we present the calculated disparity $(\sqrt{u^2 + v^2})$ using the classical correlation method (bottom left) and using our method with the correlation technique as initialization (bottom right). Two epipolar lines are depicted in the right image. They correspond to the points represented by a cross around the left eye and the nose in the left image. Figure 2 shows several views of the 3-D reconstruction of the face in the stereo pair, using the disparity map obtained by our method. The reconstruction looks very realistic.
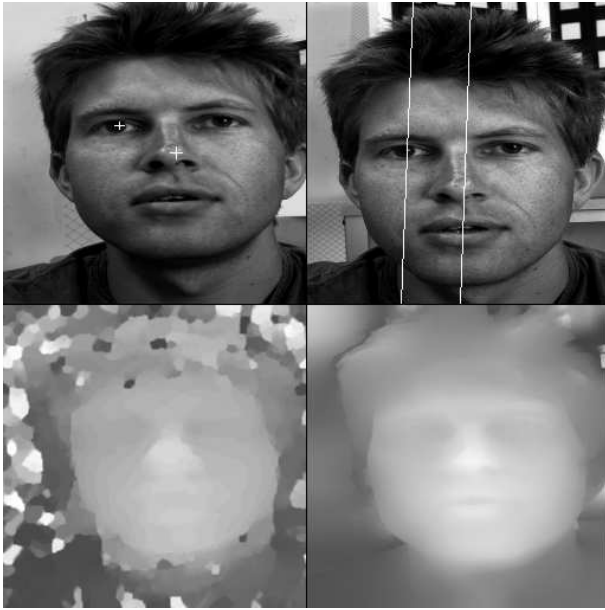
Figure 1: Top: the original stereo pair. Bottom left: the computed disparity map using a correlation window of size 13×13. Bottom right: our method ($\sigma_0 = 7$, $\sigma_n = 0.8$, $\alpha = 1$, $s = 0.5$, $\eta = 0.95$), with the correlation result as initialization.
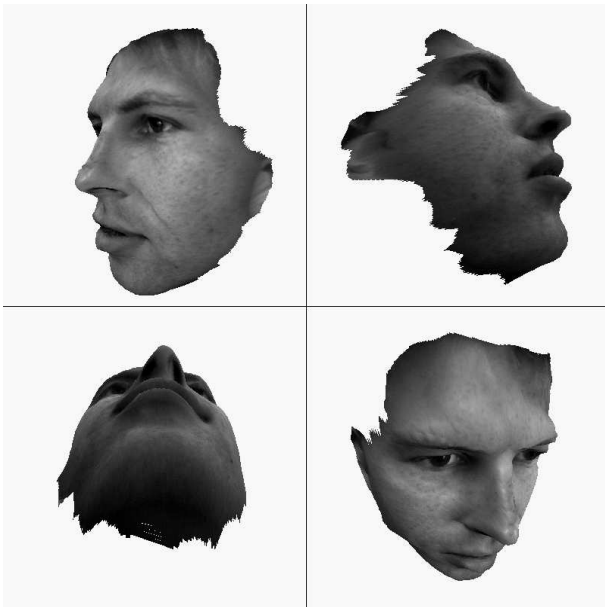

Figure 2: Four views of the 3-D reconstruction of the stereo pair in Figure 1, using the disparity map from our method.

References

[1] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, *Dense Disparity Map Estimation Respecting Image Discontinuities: a PDE and Scale-Space Based Approach*, Technical Report INRIA 3874, Sophia–Antipolis, France, 2000.

[2] L. Alvarez, P.-L. Lions, and J.-M. Morel, Image selective smoothing and edge detection by nonlinear diffusion. II, *SIAM J. Numer. Anal.* **29**, 1992, 845–866.

[3] L. Alvarez, J. Weickert, and J. Sánchez, Reliable estimation of optical flow for large displacements, *Int. J. Comput. Vision*, in press.

[4] S. T. Barnard, Stochastic stereo matching over scale. *Int. J. Comput. Vision* **3**, 1989, 17–32.

[5] R. Deriche, C. Bouvin, and O. Faugeras, A level-set approach for stereo, in *Investigative Image Processing*, (L. I. Rudin and S. K. Bramble, Eds.), Boston, MA, Nov. 1996, SPIE Proceedings, Vol. 2942, pp. 150–161, 1997.

[6] R. Deriche, S. Bouvin, and O. Faugeras, Front propagation and level-set approach for geodesic active stereovision, in *Third Asian Conference On Computer Vision*, Hong Kong, January 1998.

[7] O. Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.

[8] O. Faugeras and R. Keriven, Variational principles, surface evolution, PDE's, level set methods and the stereo problem, *IEEE Trans. on Image Processing* **7**, 1998, 336–344.

[9] F. Heitz, P. Pérez, and P. Bouthemy, Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP : Image Understanding* **59**, 1994, 125–134.

[10] H.-H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from images sequences, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 1986, 565–593.

[11] L. Robert and R. Deriche, Dense depth map reconstruction: a minimization and regularization approach which preserves discontinuities, in *Computer Vision – ECCV '96* (B. Buxton, R. Cipolla, Eds.), Vol. I, pp. 439–451, Lecture Notes in Computer Science, Vol. 1064, Springer, Berlin, 1996.

[12] J. Shah, A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo, in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, New York, June 15–17, 1993, pp. 34–40, IEEE Computer Society Press.

[13] J. Weickert, S. Ishikawa, and A. Imiya, Linear scale-space has first been proposed in Japan, *J. Math. Imag. Vision* **10**, 1999, 237–252.