Differential Geometry Techniques In Stereo Vision

Adee Ran[1] and Nir Sochen[2]
[1]Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, 32000, Israel
[2]Department of Applied Mathematics, University of Tel-Aviv, Tel-Aviv, 69978, Israel

Abstract

This work presents an algorithm for reconstruction of dense depth data from a parallel-axis stereoscopic image pair, using a diffusion process.

In order to get good depth reconstruction we propose an expansion of initial local matching to area matching. The expansion is performed by using non-homogeneous, edge-preserving diffusion. The diffusion PDE creates an evolution of a three-dimensional manifold (describing the matching cost function for different offsets) embedded in four-dimensional space. The evolution of the manifold in time is described by the Beltrami operator, which is a generalization of the Laplacian to non-Euclidean manifolds.

The proposed algorithm was tested on several image pairs, and compared with other algorithms found in literature. Results are shown.

I. Introduction

The problem we are dealing with is an estimation of the 2-D disparity function $d(x, y)$, which can be used to reconstruct the 3-d structure of the scene. it can be shown (for example, [2]) that in parallel-axis stereo, the 3-d coordinates of each point in the scene which appears in two images can be obtained from the coordinates of its projection on one image (x_L, y_L), its disparity $d = x_L - x_R$, the location of the cameras and their focal length.

Extracting a dense disparity map from a stereoscopic image pair is a non-trivial task. It is, in fact, an ill-posed problem [1]. Different approaches to this problem were taken since the mid-70's. Some of them are described in [2], [5]. Two major difficulties are reconstruction of surfaces with non-constant depth, and reconstruction of object boundaries without blurring.

We propose an algorithm which handles these difficulties, using an edge-preserving diffusion equation known as the Beltrami flow, which was previously used successfully for removal of noise from gray-level and color images [8, 11] .

This paper is organized as follows: In section II we outline the task of stereo matching, and describe briefly some previous work. Section III describes the Beltrami framework for low-level vision [8, 11]. We describe a prior algorithm [12] and present the stereo matching algorithm in section IV, and then show the results of our method and compare them to the results of [12] in section V.

II. Stereo Matching

Matching specific feature points which are found in both images is a relatively easy task. It is possible to find the 3-d location of the features and then use surface interpolation to reconstruct the rest of the scene. This idea works fine when the surfaces are smooth (e.g. in DTM construction), but fails when objects of different depths appear in the scene, since the depth changes abruptly across object boundaries and the real depth function is not smooth.

A better approach for the last case is to find a match for each pixel in one image, i.e. a dense disparity image. Consider the 3-d point P which corresponds to a pixel (i, j) in the left image. The projection of P in the right image is in coordinates $(i + d(i, j), j)$. We then call $d(x, y)$ the *disparity* of pixel (i, j). In general the disparity is a positive real number; We shall assume that the image resolution is high and treat d as an integer, quantizing to whole pixels. We shall also limit the set of values that d can take so that $d(i, j) \in \{d_{min}, d_{min} + 1, ...d_{max}\}$.

If we try to find a match for a specific pixel without considering its neighborhood, we might find more than one match in the other image, since the intensity is quantized. Other sources of errors may also exist (e.g. blurring, electronic noise etc.) and might cause a "wrong" pixel to be a better match than the "right" pixel. Consequently, this local matching scheme does not give good results except for specific synthetic images.

In order to prevent local erroneous matches, area-matching is customarily used. As a simple example of an area-matching criterion, consider the sum of squared difference of intensities over a fixed window around the points which are tested for a match. The inclusion of neighbor pixels makes the matching less ambiguous and reduces the effect of noise. However, a large window may cause unwanted blurring of abrupt depth changes. Moreover, a good match over the whole window is possible only when the point is on a surface of constant depth ("fronto-parallel"), whereas slanted or more complex surfaces will not match well with any window.

Several approaches were taken to overcome these problems. Grimson [4] used a 3-dimensional window by summing over a certain range of disparities. Pollard, Mayhew and Frisby [6] used a shaped 3D window in the sum by placing a limit on the disparity gradient. Both methods used linear combination of the squared difference of intensities. Kanade and Okutomi [5] proposed an adaptive-window matching scheme, in which for every pixel the disparity is estimated iteratively, and the window size is adjusted so as to maximize the *certainty* of the match, according to probabilistic models of the disparity and brightness functions. Szeliski and Scharstein [12], [13] used non-linear diffusion to aggregate the support each "candidate" receives from its neighborhood. Their method also uses a probabilistic model of disparity.

One should note that the window size and shape adjustment task is typical in other problems in computer vision such as edge detection and noise removal. In all these cases, different input images require different windows sizes, or equivalently different resolutions in which the processing can be performed. Moreover, different objects in the same scene may require treatment in different scales. Recent work in image processing uses the scale-space approach to address this task. In this approach, a family of images is created from the input image, with a single scale (or time) parameter. As we progress in scale further from the input image, less detail is retained in the image, making it less noisy. The scale-space is built by a partial differential equation (PDE) which describes the evolution of the image in scale. The choice of a PDE makes it possible to create different scale-spaces with various properties, and is a subject of many works (see [10] and references therein).

In the stereo matching context, we can view some of the previously mentioned approaches as choices of a 3D scale-space, and the selection of window size or level of diffusion as the specific scale in which the stereo matching should be treated.

An important property for a scale-space is preserving edge location and sharpness for as long as possible along scale. Sochen et al. [8, 11] have proposed the Beltrami flow – a scale-space PDE using the Beltrami differential operator, which is based on differential geometry concepts, and showed that it preserves edges and flattens homogeneous regions in images. The Beltrami flow is presented in the next section.

III. THE BELTRAMI FRAMEWORK

Let us briefly review our general framework for non-linear diffusion in computer vision.

A. *Riemannian Structure*

We represent an image and other local features as embedding maps of a Riemannian manifold in a higher dimensional space. The simplest example is the image itself which is represented as a 2D surface embedded in $\mathbb{R}^3$. We denote the map by $Y : \Sigma \to \mathbb{R}^3$, Where Σ is a two-dimensional surface, and we denote the local coordinates on it by (x^1, x^2). The map Y is given in general by $(Y^1(x^1,x^2), Y^2(x^1,x^2), Y^3(x^1,x^2))$. In our example we represent it as follows $(Y^1 = x^1, Y^2 = x^2, Y^3 = I(x^1,x^2))$. We choose on this surface a Riemannian structure, namely, a metric. The metric is a positive definite symmetric 2-tensor that may be defined through the local distance measurements:

$$ds^2 = g_{11}(dx^1)^2 + 2g_{12}dx^1dx^2 + g_{22}(dx^2)^2. \quad (1)$$

We use below the Einstein summation convention in which the above equation reads $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, where identical indices that appear one up and one down are summed over. We denote the inverse of the metric by $g^{\mu\nu}$.

The metric g on Σ can be chosen arbitrarily, but we can use the knowledge of the metric on M and the map Y to construct the *induced* metric. The induced metric defines an infinitesimal distance on Σ to be equal to the distance measured in M (See [11] for details). The induced metric in our example is:

$$(g_{\mu\nu}) = \begin{pmatrix} 1+I_x^2 & I_xI_y \\ I_xI_y & 1+I_y^2 \end{pmatrix}. \quad (2)$$

B. Polyakov Action: A measure on the space of embedding maps

Denote by (Σ, g) the embedded manifold and its metric, by m the dimension of Σ, and by (M, h) the space-feature manifold and its metric, then the functional $S[Y]$ attaches a real number to a map $Y : \Sigma \to M$:

$$S[Y^i, g_{\mu\nu}, h_{ij}] = \int dV \langle \nabla Y^i, \nabla Y^j \rangle_g h_{ij} \quad (3)$$

where $dV = d^m x\sqrt{g}$ is an m-dimensional volume element, and the scalar product $<,>_g$ is defined with respect to the image metric i.e. $\langle \nabla R, \nabla G \rangle_g = g^{\mu\nu}\partial_\mu R\,\partial_\nu G$, where $\partial_\mu R \equiv \frac{\partial R}{\partial \mu}$. This functional, for $m = 2$ and $h_{ij} = \delta_{ij}$, was first proposed by Polyakov [7] in the context of high energy physics, and the theory known as *string theory*.

Using standard methods in the calculus of variations (see [11]), the gradient descent equation with respect to the embedding (assuming Euclidean embedding space) is:

$$Y_t^i = \frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}g^{\mu\nu}\partial_\nu Y^i). \quad (4)$$

The operator that is acting on Y^i is the natural generalization of the Laplacian from flat spaces to manifolds and is called *the second order differential parameter of Beltrami* [9], or for short *Beltrami operator*. We will denote it by Δ_g.

The resulting flow preserves edges well while smoothing more homogeneous regions rapidly (see [11]). Also, it is easily defined for higher dimensions. These properties make it suitable for the stereo matching problem as we show in the next section.

IV. The Proposed Method

We propose to perform stereo matching using a Beltrami flow scale-space to remove false matches.

Following [12] we divide the solution to three steps:

1. Computation of local matching cost function for each pixel and each disparity under consideration – in other words, initialization of the 3D disparity space,
2. Aggregation of support for each disparity candidate using Beltrami flow, and
3. For each pixel, choosing as estimated disparity the candidate that has minimum matching cost.

As a local matching cost function we chose the same function used in [12], which corresponds to a contaminated Gaussian model:

$$\rho_M(x) = -\ln\left((1-\varepsilon_M)\exp(\frac{-x^2}{2\sigma_M^2}) + \varepsilon_M\right)$$

By setting ε_M and σ_M, various error measures can be achieved: ε_M controls the maximum cost and can be adjusted according to the nature of the images, e.g. object and background contrast. σ_M can be adjusted according to the expected image noise.

The disparity space $E(x, y, d)$ is initialized as

$$E_0(x, y, d) = \rho_M\big(I_L(x+d, y) - I_R(x, y)\big) \quad (5)$$

In the second step the initial local matching cost should be modified to allow support from some neighborhood. In [12] this was performed by a non-linear diffusion. The non-linearity was introduced by a conversion of the matching cost function $E(x, y, d)$ to a probability distribution at each pixel (x, y), while the aggregation of support across d was by simple linear smoothing.

We propose a different approach. The 3D disparity space is treated as a 3D manifold embedded in 4D space (x, y, d, E) which we assume to be Euclidian. We can then proceed in the Beltrami framework described in section III: We compute the induced metric and apply the Beltrami flow to the disparity space.

The induced metric for the map $Y : \Sigma \rightarrow \mathbb{R}^4$ is

$$(g_{\mu\nu})(x, y, d) = \begin{pmatrix} 1 + E_x^2 & E_x E_y & E_x E_d \\ E_x E_y & 1 + E_y^2 & E_y E_d \\ E_x E_d & E_y E_d & 1 + E_d^2 \end{pmatrix}, \tag{6}$$

for which $g = |g_{\mu\nu}| = 1 + E_x^2 + E_y^2 + E_d^2$. Substituting into 4 we get the flow

$$\begin{aligned} E_t =& \frac{1}{g}\big(E_{xx} + E_{yy} + E_{dd}\big) - \\ & \frac{1}{g^2}\big(E_{xx}E_x^2 + E_{yy}E_y^2 + E_{dd}E_d^2\big) + \\ & \frac{2}{g^2}\big(E_x E_y E_{xy} + E_x E_d E_{xd} + E_y E_d E_{yd}\big). \end{aligned} \tag{7}$$

This flow preserves "edges" in the disparity space which correspond to matches with local support (in some direction on the (x, y, d) manifold). At these points the matching cost will change slowly and should stay low. For false matches with no support the flow will drive the matching cost faster toward higher values.

We note that the geometry of the disparity space depends on the aspect ratio between the spatial dimensions (x, y) and the disparity axis d. This ratio is a parameter of the algorithm which should be set according to the nature of input images, e.g. cameras' focal length.

The final step is the estimation of d for each (x, y). We take as the estimate simply the value of d which minimizes the (diffused) matching cost function (for a different approach in this stage, see [3].)

V. Results

We present the results of the proposed algorithm (labeled "BF") on two image pairs. The result of Szeliski and Scharstein's non-linear diffusion algorithm [12] (labeled "SS") for these image pairs is also shown.

The first pair is a synthetic view of a corridor with objects, taken from `http://www-dbv.cs.uni-bonn.de/~ft/stereo.html` (courtesy of the university of Bonn, Germany). The original images are shown in Figures V and V. We used the same parameters for the local matching cost for both algorithms: $\sigma_M = 2$ and $\varepsilon_M = 0.001$. The other parameters for the SS algorithm were taken from [12]. For our algorithm we set the aspect ratio of $(x, y) : d$ as 10:1. Both algorithms use 10 iterations with $\Delta t = 0.25$ (labeled μ in [12]).

The result of our algorithm is shown in Figure 3 and the result of the SS algorithm is shown in Figure 4.

Figure 1: Corridor - left view.

The second pair are real images of a castle model, taken from `http://www.cs.cmu.edu/afs/cs/project/cil/ftp/cil-0001` (courtesy of Carnegie-Mellon University, Pennsylvania, USA). The original images are shown in Figures V and V. We used 20 iterations with the same parameters noted above except for σ_M which was taken as 5.

The result of our algorithm is shown in Figure V and the result of the SS algorithm is shown in Figure V.

Figure 2: Corridor - right view.

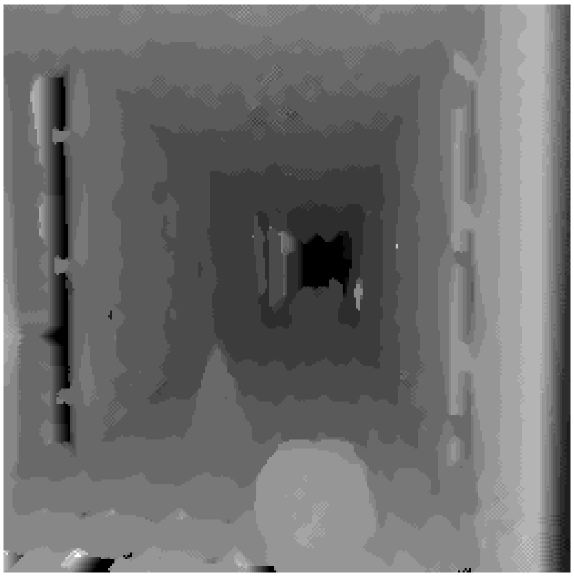

Figure 3: Corridor disparity image - Beltrami flow, 10 iterations.

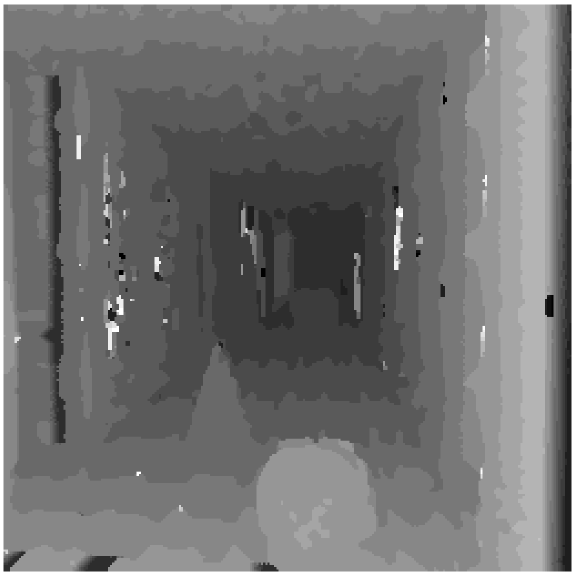

Figure 4: Corridor disparity image - SS Diffusion, 10 iterations.

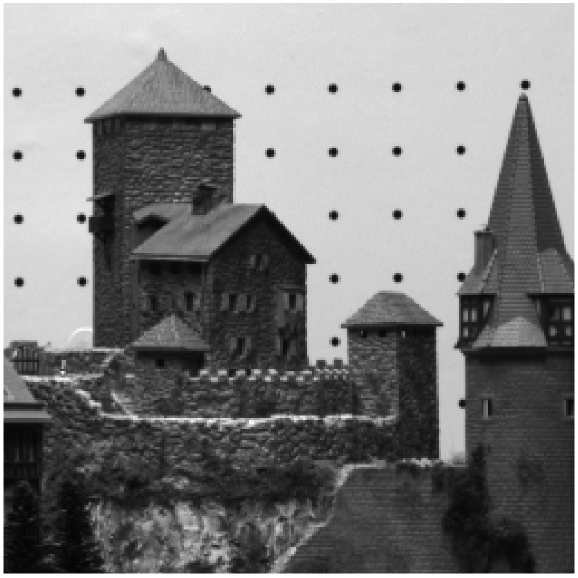

Figure 5: Castle - left view.

VI. CONCLUDING REMARKS

The algorithm proposed in this work achieves better results than SS [12], in most of the cases. Moreover, the proposed method gives good results after fewer iterations than the compared method, with low sensitivity to an increase of the number of iterations after a certain point.

The proposed method can naturally be extended to the case of color images or images of other features, and for more than two images of the same scene. These extensions will be worked out in the future.

VII. REFERENCES

[1] R. Bertero, T. Poggio and V. Torre, Ill-Posed Problems in Early Vision, *Proc. IEEE*, Vol. 76, No. 8, August 1988

[2] U. R. Dhond and J. K. Aggarwal, Structure from stereo - A review, *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489-1509, November/December 1989.

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

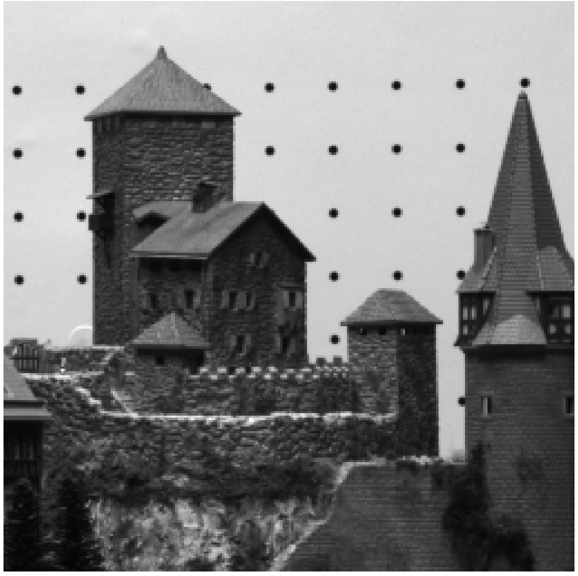

Figure 6: Castle - right view.

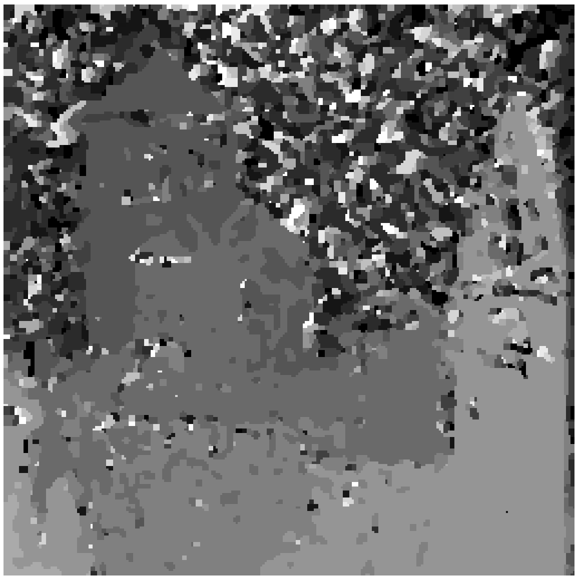

Figure 8: Castle disparity image - SS Diffusion, 20 iterations.

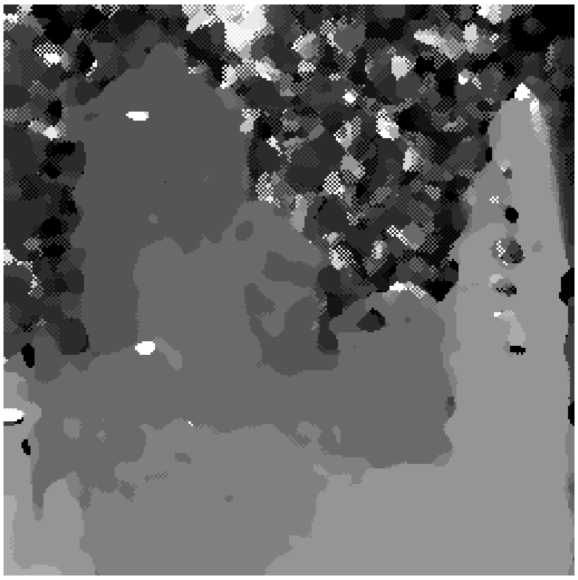

Figure 7: Castle disparity image - Beltrami flow, 20 iterations.

[4] W. E. L. Grimson, Computational experiments with a feature based stereo algorithm, *IEEE Trans. PAMI*, Vol. 7, No. 1, pp. 17-34, January 1985.

[5] T. Kanade and M. Okutomi, A stereo Matching Algorithm with an Adaptive Windows: Theory and Experiment, *IEEE Trans. PAMI*, vol. 16, no. 9, pp. 920-932, September 1994.

[6] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, PMF: A stereo correspondence algorithm using a disparity gradient limit, *Perception*, vol. 14, pp. 449-470, 1981.

[7] A. M. Polyakov, Quantum geometry of bosonic strings, *Physics Letters B*, 103B, pp. 207-210, 1981.

[8] R. Kimmel, R. Malladi and N. Sochen, Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images, LBNL, UC and TAUHEP Report LBNL-40490, UC-405, June 1997.

[9] E. Kreyszig, Differential Geometry, Dover Publications, Inc., New York, 1991.

[10] Geometric–Driven Diffusion in Computer Vision, editor B. M. ter Haar Romeny, Kluwer Academic Publishers, The Netherlands, 1994.

[11] N. Sochen, R. Kimmel and R. Malladi, A general framework for low level vision, *IEEE Trans. IP*, vol. 7, no. 3, pp. 310-318, March 1998

[12] D. Scharstein and D. Szeliski, Stereo matching with non-linear diffusion, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, pp. 343-350, June 1996.

[13] D. Scharstein and D. Szeliski, Stereo matching with non-linear diffusion, *International Journal of Computer Vision*, 28(2): pp. 155-174, June/July 1998.