Non-Isotropic Regularization of the Correspondence Space in Stereo-vision

Rami Ben-Ari Nir Sochen
Department of Applied Mathematics
Tel Aviv University
Ramat-Aviv, Tel-Aviv 69978, Israel
{ramiben,sochen}@post.tau.ac.il

Abstract

The correspondence problem in stereo vision is notoriously difficult. In many approaches a noisy solution is extracted from the correspondence space. Various sophisticated regularization techniques are applied then on this noisy solution. We study here the possibility to denoise the correspondence/correlation space before *extracting the solution, by a non-linear and non-isotropic scheme. We show that this methods preserves edges (depth discontinuities) well and overcomes some of the problems encountered in previous approaches.*

1. Introduction

Over the years, numerous algorithms for stereo vision have been proposed, which use different strategies. Some techniques are based on direct evaluation of depth from two or more images of the scene [2]. In Other approaches the correspondence problem is solved and the depth is calculated from the disparity evaluation.

The correspondence problem in stereo vision is notoriously difficult task. There are feature based, area based or energy based approaches to this problem. In the feature based techniques [1] the disparity is evaluated only for selected points extracted from the images. While the feature based strategy leads to sparse resolution, a dense resolution can be achieved by an area based approach. A general survey for different area based approaches can be found in [3].

Given two intensity images I_L and I_R, taken from two parallel cameras located on the x axis, the correspondence problem can be formulated by (when a correspondence measure is assigned to the right image):

$$\widehat{d}(x,y) = arg \min_{d(\cdot,\cdot)\in D} \psi(I_R(x,y), I_L(x+d(x,y),y)),$$

where $\widehat{d}$ is the evaluated disparity, ψ is a cost function and D presents the range of $d(\cdot,\cdot)$. Although this Formulation is general, it can not be directly applied.

Most area-based approaches are implied by the following three stages: 1) Definition of potential function as a measure for correspondence, based on a local (pixelwise) cost function. 2) Regularization of the potential function. 3) Extraction of the disparity map by minimization of the potential function. In the classical SSD (Sum of Squared Differences) the first and the second stages are combined into one. The focus of this paper is in the second stage i.e the regularization of the potential function (correspondence space).

Many previous works incorporating the regularization of the correspondence space attempted to find the optimal size of the support region. A support region can either be two dimensional at a fixed disparity (favoring piecewise constant disparity surfaces) or three dimensional in $x-y-d$ space, where (x,y) are the spatial point coordinates and d is the disparity. Two or three dimensional Gaussian convolution or the SSD scheme present a well known linear regularizations. The SSD scheme is based on a fixed support region defined by the window size. The Gaussian convolution can be approximated also by a fixed support region where the effective support depends on the Gaussian variance. If the window size or the Gaussian width are too small, mismatch is usually obtained due to ambiguities and noise. In the other hand, if the window size or the Gaussian width are too big, loss of detail and blurring of the depth discontinuities are caused. In order to cope with this problem Kanade and Okutomi proposed a non-linear scheme with varying support region employing adaptive window size [4].

A general formulation for an area based technique was suggested by Scharstein and Szeliski in [3]. The authors proposed a probabilistic scheme based on iterative process with *non-linear* but *isotropic* regularization of the correspondence space. The regularization was based on support with variable size. This approach was compared in [3] with several other approaches including the adaptive window scheme [4], and achieved improved results.

Here we introduce a new approach based on the three stage process mentioned above where the Beltrami framework (BFW) is employed for regularization of the correspondence space (first applied in [8]). Under the BFW we present a *non-linear* and *non-isotropic* regularization, leading to more reliable disparity evaluations and better preserving the discontinuities. We compare the performance of our scheme with the simple 2-D Gaussian convolution and the probabilistic scheme suggested in [3], for both synthetic and real images.

2. Regularization by Gaussian convolution

Regularization of the correspondence space can be carried out by convolution with two or three dimensional Gaussian. The three dimensional convolution includes convolution in the disparity axis, leading directly to the undesired smoothing of the discontinuities. Therefore as regularization by Gaussian convolution we choused the two dimensional Gaussian, acting on the spatial coordinates. The convolution can be implemented using local iterative diffusion (Linear Scale Space) or directly by discrete convolution:

$$E(x, y, d) = G_\sigma * E_0(x, y, d), \tag{1}$$

Where E_0 is the initial potential function (before regularization), G_σ is a normalized Gaussian kernel with variance σ^2 and the $*$ is the convolution operator.

3. The Beltrami Framework

The Beltrami framework has already been employed for non-linear diffusion in computer vision [6, 7].

Lets denote by (Σ, g) the three dimensional potential function manifold $X = (x, y, d, E(x, y, d))$ and its metric. This manifold is embedded in $\mathbb{R}^4$ with its Euclidian metric. The functional S attaches a real number to a map $\mathbf{X} : \Sigma \to \mathbb{R}^4$:

$$S[X^i, g_{\mu\nu}] = \int \mathsf{d}^3\sigma \sqrt{g} g^{\mu\nu} \partial_\mu X^i \partial_\nu X^j \delta_{ij}(\mathbf{X}), \tag{2}$$

where g is the determinant of the potential function metric, $g^{\mu\nu}$ is the inverse of the metric on the potential function manifold, the range of the indices $\mu, \nu = 1, 2, 3$ and $i, j = 1, 2, 3, 4$ and δ_{ij} (Kroneker function) presents the metric of the Euclidian embedding space. This functional, for two dimensional manifold, was first proposed by Polyakov in the context of high energy physics [5]. We minimize the functional in (2) with respect to the embedding map ($\mathbf{X}$), employing the induced metric of the image manifold (see [7]). Our functional (2) under the choice of induced metric is simply the area of the image manifold $(x, y, d, E(x, y, d))$:

$$S(E) = \int \sqrt{g}\ \mathsf{d}x\mathsf{d}y\mathsf{d}d, \tag{3}$$

Where g is the determinant of the induced metric matrix:

$$g = \sqrt{1 + |\nabla E|^2}$$

The minimization of the functional (3) drive the correspondence space towards a minimal volume manifold. Modifying only the E values of the manifolds is an edge preserving operation [6, 7].

4. Application of Beltrami Framework in Stereo Vision

Let the local cost function be the contaminated Gaussian model, as used by Scharstein and Szeliski in [3]:

$$\rho(u) = -\log[\epsilon + (1 - \epsilon)\exp(-\frac{u^2}{2\sigma^2})], \tag{4}$$

Where u is the variable and ϵ and σ are parameters. This robust statistics function presents a compromise between binary metric (where a constant cost is given for every non-correlation) and quadratic metric. The cost function is quadratic in the interval determined by σ and reaches a saturation level outside that interval.

When a correspondence measure is assigned to each pixel of the right image, the initial potential function will be:

$$E_0(x, y, d) = \rho(I_R(x, y) - I_L(x + d, y)). \tag{5}$$

The disparity map is evaluated from the potential function by minimization:

$$\hat{d}(x, y) = arg \min_{d \in (d_{min}, d_{max})} E(x, y, d). \tag{6}$$

The initial potential function is usually highly noised and the disparity extraction from this function will obtain a noisy disparity map. In this approach we employ the BFW for regularization. The regularization process is obtained by minimization of the functional $S(E)$ in (3). Using standard methods in the calculus of variations the Euler-Lagrange equation for to the correspondence is:

$$-\frac{1}{\sqrt{g}}\frac{\delta S}{\delta E} = \frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}g^{\mu\nu}\partial_\nu E). \quad (7)$$

The diffusion is obtained as a gradient descent minimization. This diffusion is a non-linear, non-isotropic diffusion preserving the discontinuities in the potential function (see [7]).

An important point is the choice of the relative scale in the embedded space. For example, the disparity is not on equal footing as x and y. We specify here the relative scale of the disparity, d, with respect to the spatial coordinates (x, y) and assign it as β. The gradient decent flow then emerges as:

$$\begin{aligned} E_t \quad = \quad & \frac{1}{\tilde{g}^2}\{\tilde{g}[(E_{xx} + E_{yy})\beta^2 + E_{dd}] \quad (8) \\ & -\beta^4(E_{xx}E_x^2 + E_{yy}E_y^2) + E_{dd}E_d^2 \\ & -2\beta^2(\beta^2 E_x E_y E_{xy} + E_x E_d E_{xd} + E_y E_d E_{yd}]\}. \end{aligned}$$

Where $\tilde{g} = \beta^2(1 + E_x^2 + E_y^2) + E_d^2$.

This flow has low sensitivity to number of iterations even though theoretically the process converges to constant potential as $t \to \infty$. This property is due to relatively slow regularization of the edges in the Beltrami flow.

The disparity map is evaluated from the regularized potential function by (6). The desired property of preserving the disparity discontinuities (edges) is achieved by preserving the discontinuities in the potential function.

5. Results

In this section we numerically evaluate the performance of our scheme and compare it to regularization by Gaussian convolution and the probabilistic model introduced in [3]. Two examples are presented. One pair of synthetic images with ground truth and one real image pair.

First we tested the schemes on the pair of synthetic 256x256 images, shown in figure 1, having a predetermined disparity pattern. The images are graylevel and we employ pseudo-color to better visualize them. The disparity range is -14 to 9 pixels (based on the right image).

The input images were contaminated by added gaussian noise with variance $\sigma^2_{Noise} = 5 \cdot 10^{-5}$ (for image intensities normalized to $[0, 1]$). Figure (2) shows the recovered disparity maps.

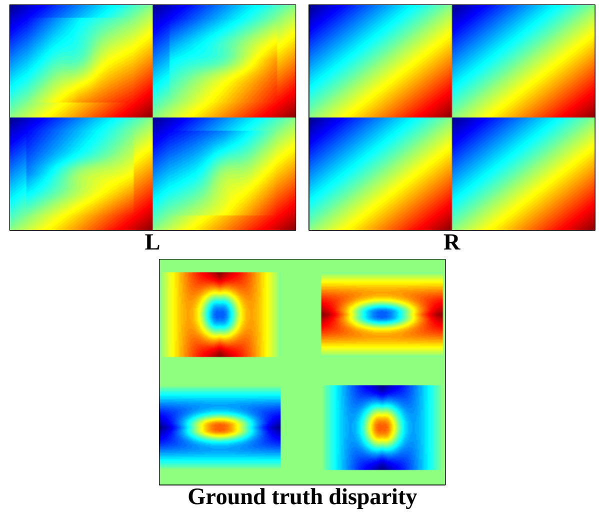

Figure 1. Top: The Synth pair, Down: Ground truth disparity

It can be seen that the results from the BFW scheme are the most reliable (based on ground truth), qualitatively and quantitatively. The BFW outperforms the other schemes by smoothing the regions with continuous disparity variations while preserving the depth discontinuities.

We also tested our scheme on real images. The results for the 286x421 CMU's Castle images are shown in figure (4).

In the Gaussian convolution scheme the boundaries are highly blurred and in some regions different objects are almost joined together (such as the tower at the right side of the image). The probabilistic model better preserves the discontinuities compared to Gaussian convolution. Still there is noticeable noise nearby the edges and within the objects, compared to the BFW results. The noise at the background observed in all the methods is due to lack of texture.

6. Conclusions

In this paper, we have demonstrated a new approach, employing the BFW for non-linear and non-isotropic diffusion of the correspondence space in stereo vision. The results obtained show improved regularization while preserving the discontinuities. It is due to this feature that more reliable disparity images are achieved.

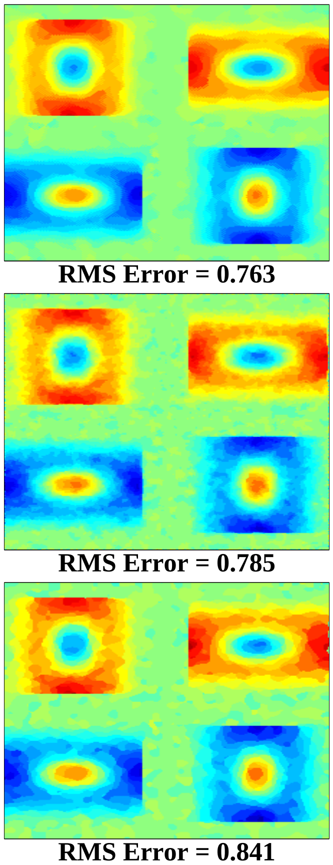

Figure 2. Reconstructed disparity : Top: Beltrami Frame Work, Middle: Probabilistic Model, Down: Gaussian Convolution

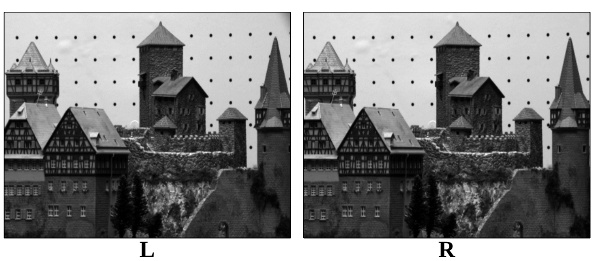

Figure 3. The Castle pair

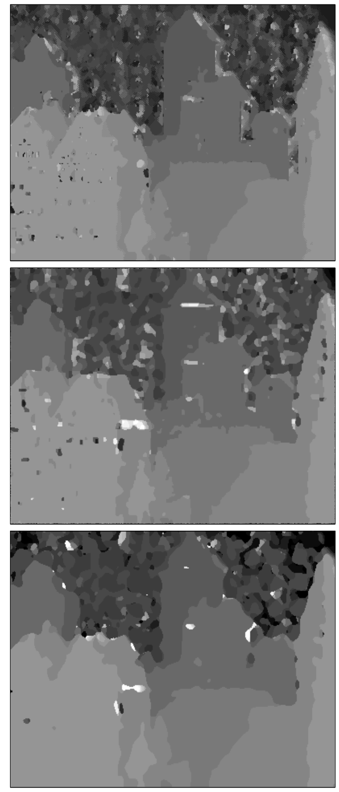

Figure 4. Reconstructed disparity : Top: Beltrami Frame Work, Middle: Probabilistic Model, Down: Gaussian Convolution

References

[1] U. R. Dhondand and J. K. Aggarwal. Srtucture from stereo – a review. *IEEE Tr. on systems, man and cybernetics* ,19(6):1489–1510, November/December 1989.

[2] L. Robert and R. Deriche. Dense Depth Map Reconstruction: A Minimization and Regularization Approach which Preserves discontinuities. *Lecture notes in Computer Science 1064* ,439–451, Springer–Verlag, Berlin 1996.

[3] D. Scharstein and R. Szeliski. Stereo matching with non–linear diffusion. *International Journal of Computer vision* ,28(2):155–174, June/July 1996.

[4] T. Kanade and M. Okutomi. A Stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Tr. on PAMI* ,16(9):920–932, September 1994.

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

[6] N. Sochen and R. Kimmel and R. Malladi. From high energy physics to low level vision. *Report, LBNL, UC Berkeley LBNL 39243 August* ,Presented in ONR workshop, UCLA, Sept. 5 1996.

[7] N. Sochen and R. Kimmel and R. Malladi. A general framework for low level vision. *IEEE Trans. on Image Processing* ,7 (1998) 310-318.

[8] A. Ran. Differential Geometry Techniques in Stereo Vision. *M.Sc. Thesis, The Technion-Israel Institute of Technology* , December 1999, In Hebrew.