Edge-preserving Simultaneous Joint Motion-Disparity Estimation

Dongbo Min, Kwanghoon Sohn
Digital Image Media Lab.
Yonsei University, Seoul, Korea
forevertin@diml.yonsei.ac.kr, khsohn@yonsei.ac.kr

Abstract

We propose an energy-based joint motion and disparity estimation algorithm with an anisotropic diffusion operator to yield correct and dense displacement vectors. The model estimates the left and right motions simultaneously in order to increase accuracy. We use the Euler-Lagrange equation with variational methods and solve the equation with the finite difference method (FDM). Then, the method computes the initial disparity in the current frame with joint estimation constraint, and regularizes this disparity by using our energy model. Experimental results show that the proposed algorithm provides accurate motion-disparity maps, and preserve the discontinuities of the object boundaries well.

1 Introduction

There has been considerable interest in recovering 3D motion flow in image sequences. Most research has involved 3D voxel data sets or image sequences from a monocular camera [1][2]. The voxel data sets are used in specific fields such as medical imaging. Estimating 3D motions from a monocular sequence is limited only when the objects show rigid motions and only contain simple depth information. 3D motion interpretation from stereo image sequences has recently been studied and used to achieve better results and applications [3]. However, most of these algorithms compute motion and depth information separately and do not consider the constraints between motion and disparity in the stereoscopic image sequences. One of the most pressing problems in 3D motion estimation is to locate corresponding points in the images. The resulting motion and disparity can be converted into a 3D motion flow system which consists of x, y and z motion parameters. A number of studies have been reported on the correspondence problem [4]. In order to produce smooth disparity fields while preserving the discontinuities resulting from the boundaries, we propose the energy model to regularize the fields. Much research has been performed in the field of edge-preserving regularization. For example, the regularization method used by Horn and Schunck introduces the edge-preserving smoothing term to compute the optical flow [5]. In addition, Nagel and Enkelmann modified the model to improve edge-preserving smoothing performance [6]. In this paper, we propose an energy model that is useful for correspondence estimation in stereo image sequences. The model also presents an efficient way to solve the energy minimization problem. To reduce computational load and improve performance, we estimate the disparity in the current frame using joint estimation constraints [7], and it is used as the initial disparity in the frame. Moreover, by including the relation in the energy model, we propose the simultaneous joint energy model which can estimate left and right motions simultaneously. In the energy model, we can acquire more correct displacement vectors and reduce the number of displacement vectors to be found.

2. Simultaneous estimation with regularization

2.1. Joint estimation constraint

Joint estimation is an efficient and accurate way to estimate motion and disparity in stereo image sequences. A coherence condition between motion and disparity in stereo sequences may be expressed as a linear combination of four vectors (two motion vectors and two disparity vectors) in two successive frame pairs, as shown in Fig. 1 and Eq. (1).

$$d_2(m_{l2}) = -u_l(m_{l2}) + d_1(m_{l1}) + u_r(m_{r1}) \qquad (1)$$

2.2. Energy-based motion and disparity estimation

General motion and disparity fields should be smooth in the object area while preserving discontinuities at the object's boundaries in the frame. In order to preserve discontinuities and overcome a classic ill-posed problem, we add the regularization term proposed by Nagel-Enkelmann [6][8]. We can estimate the motion and disparity fields by

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

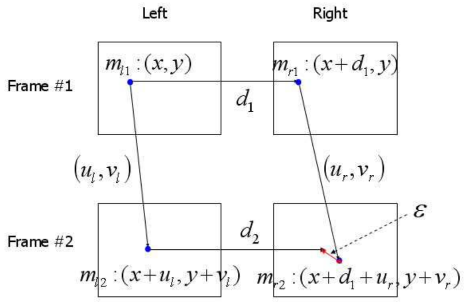

Figure 1. Simultaneous estimation model

minimizing the energy model, which consists of fidelity and regularization terms.

$$\begin{aligned} E_D(d) &= \int_\Omega (I_{l1}(x,y) - I_{r1}(x+d,y))^2 dxdy \\ &\quad + \lambda \int_\Omega (\nabla d)^T D(\nabla I_{l1})(\nabla d) dxdy \\ E_M(u,v) &= \int_\Omega (I_{l1}(x,y) - I_{l2}(x+u,y+v))^2 dxdy \\ &\quad + \lambda \int_\Omega trace((\nabla h)^T D(\nabla I_{l1})(\nabla h)) dxdy \end{aligned} \quad (2)$$

where $h(x,y) = (u(x,y), v(x,y))$

$$D(\nabla I) = \frac{1}{|\nabla I|^2 + 2\sigma^2} \left[\begin{pmatrix} \frac{\partial I}{\partial y} \\ -\frac{\partial I}{\partial x} \end{pmatrix} \begin{pmatrix} \frac{\partial I}{\partial y} \\ -\frac{\partial I}{\partial x} \end{pmatrix}^T + \sigma^2 I \right] \quad (3)$$

E_D, E_M refer to the energy functional of motion and disparity, respectively. Ω is an image plane, λ is a weighting factor. $D(\nabla I_{l1})$ is an anisotropic linear operator, which is a regularized projection matrix in the perpendicular aspect of ∇I_{l1}. An energy model that uses the diffusion operator inhibits blurring of the fields across the boundaries of I_1.

2.3. Simultaneous joint estimation model

By applying the joint constraint to the energy model, we propose the simultaneous joint estimation model. The model can acquire more correct vectors and reduce the number of vectors to be found. The model for simultaneous joint estimation is shown in Fig. 1. To apply the constraints to the motion and disparity in the energy model, we define the new data term E_J as follows:

$$\begin{aligned} E_J(u_l, v_l, u_r, v_r; d_1) &= \int_\Omega (I_{l2}(x+u_l, y+v_l) \\ &\quad - I_{r2}(x+d_1+u_r, y+v_r))^2 dxdy \end{aligned} \quad (4)$$

The energy functional is the cost between the point $(x + u_l, y + v_l)$ estimated by the left motion from the (x, y) and the point $(x+d_1+u_r, y+v_r)$ estimated by the right motion from the $(x+d_1, y)$. The more correct the joint estimation, the smaller the energy functional E_J. By adding the term, we propose the simultaneous joint energy model which can estimate left and right motions simultaneously. The simultaneous joint estimation model is as follows.

$$\begin{aligned} E_D(d_1) &= E_{DP}(d_1) \\ E_{SM}(u_l, v_l, u_r, v_r; d_1) &= E_{ML}(u_l, v_l) + E_{MR}(u_r, v_r) \\ &\quad + E_J(u_l, v_l, u_r, v_r; d_1) \end{aligned} \quad (5)$$

E_{DP}, E_{DC} are the energy functional of disparity in the previous and current frames, and E_{ML}, E_{MR} are the energy functional of motion in left and right image sequences in Eq. (2). E_{SM} refers to the energy functional of left and right motions in the simultaneous joint estimation model. The functional also consists of the data term and the smoothing term. The data term refers to the data term for the left and right motions, and the data term E_J for the joint estimation constraint. The smoothing term refers to the smoothing term for the left and right motions. The smoothing term is the same as that in Eq. (2), apart from the fact that the smoothing term for the right motion is $(x + d_1, y)$. This is because the four corresponding points starting from (x, y) are the same points in the 3D space. Simultaneous joint estimation should be executed using the corresponding points. By adding the E_J, the error generated by joint estimation is reduced, and we can acquire a more correct solution. Since the estimation uses a disparity vector in the previous frame $d_1(x, y)$, the disparity vector must be estimated correctly at first. In the model, since the joint estimation constraint is considered, the value of the vertical disparity in the current frame is reduced. Therefore, we can obtain the y-motion in the right sequence as follows:

$$\begin{aligned} d_{2y}(x+u_l, y+v_l) &= v_r(x+d_1, y) - v_l(x,y) \cong 0 \\ v_r(x+d_1, y) &= v_l(x,y) \end{aligned} \quad (6)$$

In general, the local minimum problem is one of the most serious problems when using energy-based methods because of the non-convexity of the functional. Therefore, in order to minimize the local minimum problem, initial field estimation should be performed before the proposed estimation is applied. We use hierarchical area-based motion estimation and a region-dividing technique for the initial motion and disparity field in the previous frame [9]. For the initial disparity field in the current frame, we simply use Eq. (1), which is also refined by the proposed energy model.

The simultaneous joint estimation is performed as follows:

1. Compute the initial motion and disparity vectors of the previous frame.
2. Estimate the disparity vector of the previous frame $d_1(x, y)$ using the initial disparity vector.
3. Estimate the left and right motion vectors using the initial motion vector and regularized disparity vector.

4. Compute the initial disparity vector of the current frame by using Eq. (1).
5. Estimate the disparity of current frame using initial disparity computed in step 4.

The minimization of Eq. (5) yields the following associated Euler-Lagrange equation with Neumann boundary conditions. We obtain the solutions to the Euler-Lagrange equations by calculating the asymptotic state (t→ ∞) of the parabolic system. The equation form for disparity in left and right is same.

$$\begin{aligned}
\frac{\partial d_1(m_{l1})}{\partial t} &= \lambda div(D(\nabla I_{l1}(m_{l1}))\nabla d_1(m_{l1})) \\
&+ (I_1(m_{l1}) - I_2(m_{r1}))\frac{\partial I_2(m_{l1})}{\partial x} \\
\frac{\partial u_l(m_{l1})}{\partial t} &= \lambda div\left(D(\nabla I_{l1}(m_{l1}))\nabla u_l(m_{l1})\right) \\
&+ (I_{l1}(m_{l1}) - 2I_{l2}(m_{l2}) + I_{r2}(m_{r2}))\frac{\partial I_{l2}(m_{l2})}{\partial x} \\
\frac{\partial v_l(m_{l1})}{\partial t} &= \lambda div\left(D(\nabla I_{l1}(m_{l1}))\nabla v_l(m_{l1})\right) \\
&+ (I_{l1}(m_{l1}) - 2I_{l2}(m_{l2}) + I_{r2}(m_{r2}))\frac{\partial I_{l2}(m_{l2})}{\partial y} \\
\frac{\partial u_r(m_{r1})}{\partial t} &= \lambda div\left(D(\nabla I_{l1}(m_{r1}))\nabla u_r(m_{r1})\right) \\
&+ (I_{r1}(m_{r1}) - 2I_{r2}(m_{r2}) + I_{l2}(m_{l2}))\frac{\partial I_{r2}(m_{r2})}{\partial x}
\end{aligned} \tag{7}$$

We also discretize Eq. (7) using a finite difference method. All spatial derivatives are approximated by forward differences, and the computationally expensive solution of the nonlinear system is avoided by using the first-order Taylor expansion. The final solution can be found in a recursive manner.

3. Experimental Results

The experiment was performed on the stereoscopic sequences, which is "Boy" of size 320×240 and "Man" of size 256× 256 in Fig. 2. The stereoscopic sequences "Boy" were captured with a Digiclops^{TM} of Point Gray Research Inc. The parameters used in the experiment for motion and disparity estimations are shown in Table 1. The values of the weighting factor λ were defined differently to the motion and disparity estimations in each model. Fig. 3 shows the disparity fields in the previous and current frames, the x and y motions in the "Boy" and "Man" images. We can see that the simultaneous model estimates smooth and edge-preserving displacement vectors. Moreover, because the initial disparity vector was computed using accurate disparity and motion vectors, the initial disparity vector acquired by joint estimation converges faster to the true solution than

(a) "Boy" (b) "Man"

Figure 2. Test sequences in previous frame

Table 1. Parameters used in joint estimation

Parameter	Values
Weighting factor(disparity, motion)	λ_D,λ_M=2000,10000
Number of iteration	T=600

the original estimation. Although "Man" images have very large disparity, maximum value of which is 64 pixels, the model yields a correct disparity map and preserves the discontinuities. Moreover, though the motion of the man is large, the model yields a correct motion map and the initial wrong disparity vector produced by the occlusion of motion is eliminated in the regularization process. Fig. 4 shows the energy of the regularization term of the disparity and motion according to λ. In the disparity, as λ increases, the convergence rate also increases, and it shows a maximum value when λ=2000. However, when λ is 3000, the regularization term diverges and this causes the solution to become trapped in the local minima. In the motion, as λ increases, the convergence rate also increases. Large λ causes the motion to become over-smoothed because the model overemphasizes the regularization term, thus we define λ as 10000.

4. Conclusion

We proposed an energy-based joint correspondence estimation algorithm to be used in stereo image sequences, and confirmed the performance of the model by applying it to several stereo image sequences. At first, we used an anisotropic diffusion operator to improve edge-preserving performance. Secondly, to increase efficiency, we estimated the motion and disparity using joint constraints. Since the model did not require time-consuming initial disparity estimation, the processing time was reduced. Finally, we proposed the joint data term to increase reliability and estimated the motion vectors in the left and right sequences simultaneously. As a conclusion, the proposed method showed good edge-preserving performance and a fast convergence rate.

Acknowledgement

"This research was supported by the MIC, Korea, under the ITRC support program supervised by the IITA" (IITA-

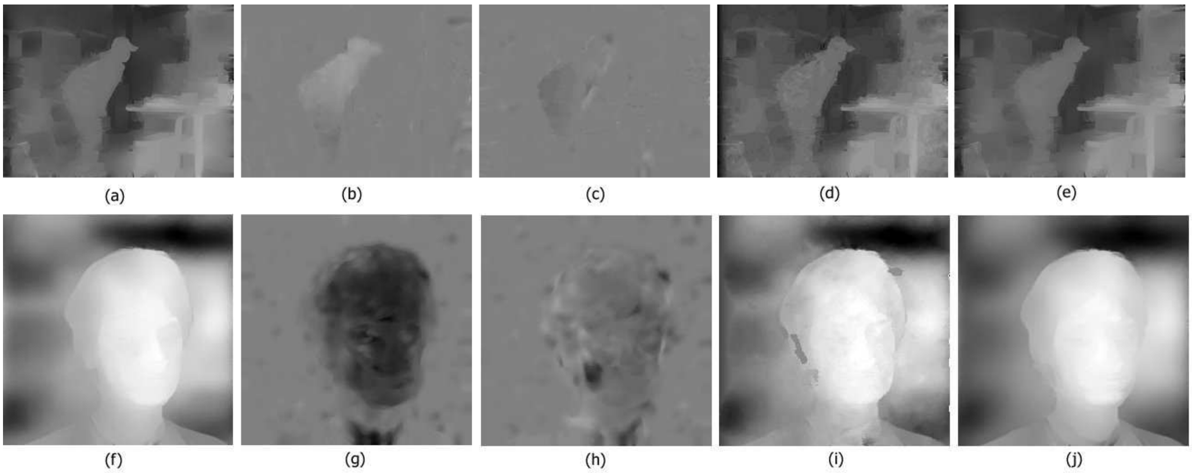

Figure 3. Disparity and Motion maps in "Boy" and "Man"images: (a) (f) disparity in the previous frame, (b) (g) x motion and (c) (h) y motions in the left image sequences (d) (i) Initial disparity and (e) (j) regularized disparity in current frame

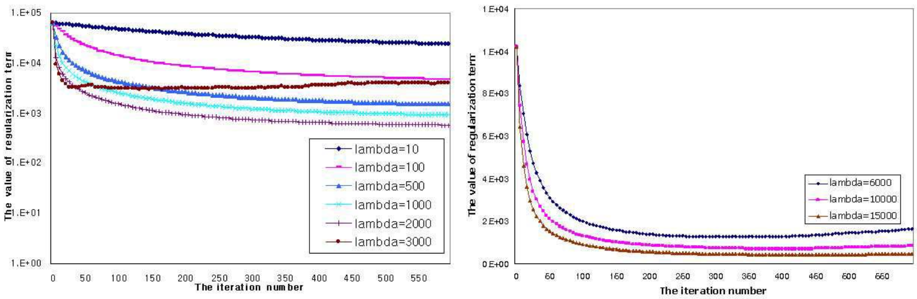

Figure 4. Energy of regularization term of disparity and motion varying with lambda in "Boy" images

2005-(C1090-0502-0027))

References

[1] M. Kruger, A. Pesavento, H. Ermert, K.M. Hiltawsky, L. Heuser and A. Jensen, "Ultrasonic strain imaging of the female breast using phase root seeking and three-dimensional optical flow," *Proc. IEEE Ultrasonic Symp.*, pp. 1757-1760, 1998.

[2] S. Vedula, S. Baker, P. Rander, R. Collins, T. Kanade, "Three-Dimensional Scene Flow," *IEEE Trans. PAMI*, vol. 27(3), pp. 475-480, 2005.

[3] Y. Zhang and C. Kambhamettu, "On 3-D scene flow and structure recovery from multiview image sequences," *IEEE Trans. Systems, Man and Cybernetics, Part B*, 33, pp. 592-606, 2003.

[4] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," *IJCV*, vol. 47, pp. 7-42, 2002.

[5] B. Horn, B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[6] H. Nagel, W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. PAMI*, vol. 8, pp. 565-593, 1986.

[7] M. Izquierdo, "Stereo Matching for Enhanced Telepresence in Three-Dimensional Videocommunications," *IEEE Trans. On CSVT*, vol. 7(4), pp. 629-643, 1997.

[8] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-space Based Approach," *J. of Visual Comm. And Image Representation*, vol. 13, pp. 3-21, 2002.

[9] H. Kim, Y. Choe, and K. Sohn, "Disparity estimation using a region-dividing technique and energy-based regularization," *Optical Eng.*, vol. 43(8) pp. 1882-1890, 2004.