

Edge-preserving joint motion-disparity estimation in stereo image sequences

Dongbo Min, Hansung Kim, Kwanghoon Sohn*

Department of Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul, 120-749, Korea

Received 15 February 2005; received in revised form 6 August 2005; accepted 24 October 2005

Abstract

An energy-based joint motion and disparity estimation algorithm with an anisotropic diffusion operator is proposed to yield correct and dense displacement vectors. We propose two energy models; the joint estimation model and the simultaneous joint estimation model. In the joint estimation model, we compute the initial disparity in the current frame with the joint estimation constraint, using the left and right motions and the disparity in the previous frame; therefore, the model is prevented from being trapped in the local minima. Then, we regularize this disparity by using our proposed energy model. In the simultaneous joint estimation model, we propose an energy functional that consists of fidelity and smoothing terms for the left and right motions and the joint data terms. We estimate the left and right motions simultaneously in order to increase correctness. We use the Euler–Lagrange equation with variational methods and solve the equation with the finite difference method (FDM) to minimize the energy model. Experimental results show that the proposed algorithm provides accurate motion-disparity maps that reflect the constraints of motion and disparity, and preserve the discontinuities of the object boundaries well.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Energy-based joint estimation; Anisotropic diffusion operator; Euler–lagrange equation; FDM; Simultaneous joint estimation

1. Introduction

There has been considerable interest in recovering 3D motion flow in image sequences. Most research has involved 3D voxel data sets [1,2] or image sequences from a monocular camera [3,4]. 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 [5]. 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. Estimating motion and disparity involves finding the corresponding points between two stereoscopic or

*Corresponding author. Tel./fax: +82 2 2123 2879.

E-mail address: khsohn@yonsei.ac.kr (K. Sohn).

temporally sequential 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 [6,16,17]: some of these studies have used feature-based approaches, energy-based approaches, dynamic programming-based approaches, and area-based approaches. In this paper, we use energy-based estimation, which yields accurate and dense displacement vector fields and shows good flexibility for several constraints [6–9].

The disparity or motion displacement vectors vary smoothly inside an object and change abruptly on the object boundaries. This phenomenon leads to the smoothness constraint. Because the surface of an object is usually continuous, the neighborhood values of the disparity or motion vectors in the objects generally correlate with each other. On the other hand, the value of the displacement vector changes radically on the boundary of an object. 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 [10]. In addition, Nagel and Enkelmann modified the regularization model to improve edge-preserving smoothing performance [11].

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—by discretizing the partial differential equation (PDE) with the finite difference method (FDM). Finally, the correct displacement vector can be found in a recursive manner by updating the vectors. However, the variational methods are sensitive to the initial data, therefore the proper initial data should be used to prevent the solution from being trapped in the local minima. The estimation of motion and disparity in stereoscopic image sequences requires high computational complexity, especially when the energy-based approach is used to acquire dense and accurate solutions. To overcome this problem and improve estimation performance, we estimate motion and disparity jointly using joint estimation constraints, which express the relation of four displacement vectors (two motion vectors and two disparity vectors) as a linear combination [12]. Using this relation, the disparity in the current frame is computed from the other three vectors, and it is used as the initial disparity in the frame. This leads to a reduction of complexity and there is also less probability of getting trapped in the local minima. Moreover, by including the relation in the energy model, we define a new energy functional, called by ‘joint data term’. By adding the term, 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.

The layout of this paper is as follows. Sections 2 and 3 explain joint estimation with regularization and simultaneous joint estimation with regularization, respectively. We describe the estimation models and energy-based estimation algorithms in each section. Finally, we present our experimental results and confirm the performance of the energy model in Section 4, and summarize the algorithm and provide suggestions for future work in Section 5.

2. Joint estimation with regularization

2.1. Joint estimation model

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. For a sampling position $(ptx1,pty1)$ at the previous frame, the relationship is as follows [12,13]:

$$\|d_1(ptx1,pty1) + u_{r1}(ptx1 + d_1(ptx1,pty1),pty1) - d_2(x,y) - u_{l1}(ptx1,pty1)\| < \varepsilon, \quad (1)$$

where ε is the error. This should be zero ideally when joint estimation is accurate. When $\varepsilon = 0$, this relationship between the existing motion and disparity vectors can be used to calculate the disparity vector in the current frame with the left and right motion vectors and the disparity vector in the previous frame, as shown in Eq. (2)

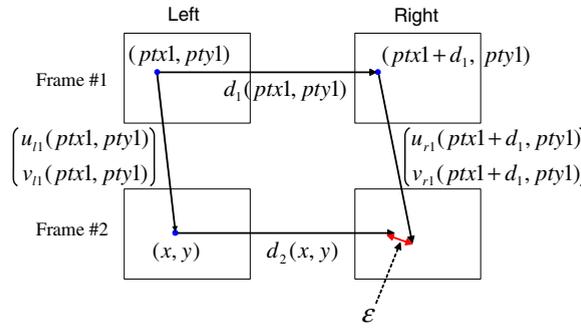


Fig. 1. Joint estimation model.

[12,13]. This method may seriously yield an accumulation of errors, which result from perturbations in the vectors. For this reason, the method is used only for finding the initial disparity vectors.

$$d_2(x, y) = -u_{l1}(ptx1, pty1) + d_1(ptx1, pty1) + u_{r1}(ptx1 + d_1(ptx1, pty1), pty1). \tag{2}$$

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 [7,8,11]. We can estimate the motion and disparity fields by minimizing the energy model, which consists of fidelity and regularization terms:

$$E_M(u, v) = \int_{\Omega} (I_{l1}(x, y) - I_{l2}(x + u, y + v))^2 dx dy + \lambda \int_{\Omega} \text{trace}((\nabla h)^T D(\nabla I_{l1})(\nabla h)) dx dy, \\ E_D(d) = \int_{\Omega} (I_{l1}(x, y) - I_{r1}(x + d, y))^2 dx dy + \lambda \int_{\Omega} (\nabla d)^T D(\nabla I_{l1})(\nabla d) dx dy, \tag{3}$$

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

E_M, E_D refer to the energy functional of motion and disparity, respectively. Ω is an image plane, λ is a weighting factor. $D(\nabla I_{l1})$, shown in Eq. (4) and proposed by Nagel and Enkelmann, is an anisotropic linear operator, which is a regularized projection matrix in the perpendicular aspect of ∇I_{l1} . To analyze the diffusion operator $D(\nabla I_{l1})$, we refer to [7,8]. σ is the anisotropic diffusion constant and represents the rate of anisotropic diffusion. An energy model that uses the diffusion operator inhibits blurring of the fields across the boundaries of I_1 where $|\nabla I_{l1}| \gg \sigma$:

$$D(\nabla I_{l1}) = \frac{1}{|\nabla I_{l1}|^2 + 2\sigma^2} \left[\begin{pmatrix} \frac{\partial I_{l1}}{\partial y} \\ -\frac{\partial I_{l1}}{\partial x} \end{pmatrix} \begin{pmatrix} \frac{\partial I_{l1}}{\partial y} \\ -\frac{\partial I_{l1}}{\partial x} \end{pmatrix}^T + \sigma^2 I \right]. \tag{4}$$

The model reveals a diffusion tensor that resembles the one used for anisotropic diffusion filtering. $D(\nabla I_{l1})$ shows the eigenvectors $v_1 = \nabla I_{l1}, v_2 = \nabla I_{l1}^\perp$ and the corresponding eigenvalues are

$$\sigma_1 = \frac{\sigma^2}{|\nabla I_{l1}|^2 + 2\sigma^2}, \\ \sigma_2 = \frac{|\nabla I_{l1}|^2 + \sigma^2}{|\nabla I_{l1}|^2 + 2\sigma^2}. \tag{5}$$

In objects where $|\nabla I_{l1}|$ approaches 0, both σ_1 and σ_2 approach 1/2, while on the boundaries where $|\nabla I_{l1}|$ approaches ∞ , σ_1 and σ_2 approach 0 and 1, respectively. In other words, the model shows isotropic behavior

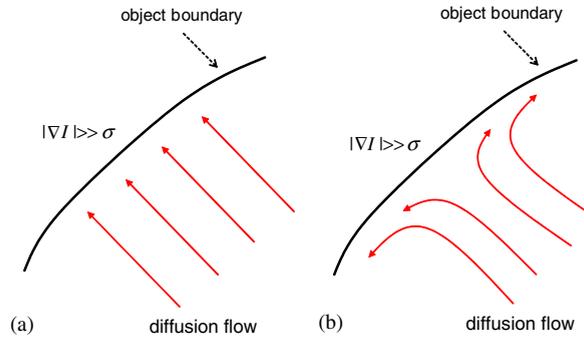


Fig. 2. The flow of diffusion process: the isotropic operator diffuses in the direction parallel to ∇d proportional to $g(|\nabla I|)$. The anisotropic operator considers the magnitude and direction of an edge: (a) The flow of isotropic diffusion; (b) the flow of anisotropic diffusion.

in objects, but on the boundaries the model smoothes dependently to not only the magnitude but also the direction of the edge. This means that it performs anisotropic behavior along the edges, as shown in Fig. 2. Therefore, the determination of σ is very important. The selection of σ depends strongly on the characteristic of an image. Alvarez et al. proposed the formulation of σ as follows:

$$s = \int_0^\sigma H_{|\nabla I_{I1}|}(z) dz, \tag{6}$$

where $H_{|\nabla I_{I1}|}(z)$ represents the normalized histogram of $|\nabla I_{I1}|$ and s is the isotropy fraction. When s approaches 0, the diffusion operator becomes anisotropic at all locations, and when s approaches 1, it leads to isotropic diffusion everywhere. With the normalization of σ , the energy model has the characteristic that is invariant to the gray level transform.

The minimization of Eq. (3) yields the following associated Euler–Lagrange equation with Neumann boundary conditions:

$$\begin{aligned} \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla u_l(x, y)) + (I_{I1}(x, y) - I_{I2}(x + u_l, y + v_l)) \frac{\partial I_{I2}(x + u_l, y + v_l)}{\partial x} &= 0, \\ \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla v_l(x, y)) + (I_{I1}(x, y) - I_{I2}(x + u_l, y + v_l)) \frac{\partial I_{I2}(x + u_l, y + v_l)}{\partial y} &= 0, \\ \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla d(x, y)) + (I_{I1}(x, y) - I_{I2}(x + d, y)) \frac{\partial I_{I2}(x + d, y)}{\partial x} &= 0. \end{aligned} \tag{7}$$

We obtain solutions to the Euler–Lagrange equation by calculating the asymptotic state ($t \rightarrow \infty$) of the parabolic system:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla u_l(x, y)) + (I_{I1}(x, y) - I_{I2}(x + u_l, y + v_l)) \frac{\partial I_{I2}(x + u_l, y + v_l)}{\partial x}, \\ \frac{\partial v}{\partial t} &= \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla v_l(x, y)) + (I_{I1}(x, y) - I_{I2}(x + u_l, y + v_l)) \frac{\partial I_{I2}(x + u_l, y + v_l)}{\partial y}, \\ \frac{\partial d}{\partial t} &= \lambda \operatorname{div}(D(\nabla I_{I1}(x, y))\nabla d(x, y)) + (I_{I1}(x, y) - I_{I1}(x + d, y)) \frac{\partial I_{I2}(x + d, y)}{\partial x}. \end{aligned} \tag{8}$$

This PDE corresponds to the nonlinear diffusion equation with an additional reaction term [14]. $D(\nabla I_{I1}(x, y))$ is a diffusivity function, which plays a role in preserving discontinuities. This function reduces smoothing on the object boundaries to preserve their discontinuities. This model works better than isotropic diffusion, because anisotropic diffusion considers the direction as well as the magnitude of the edge.

In general, the local minimum problem is one of the most serious problems when using energy-based methods because of the nonconvexity of the functional. We cannot expect unique solutions to Eq. (8), since there can be multiple local minima. Therefore, in order to minimize the local minimum problem, initial field estimation should be performed before the proposed estimation is applied. We use hierarchical block-based

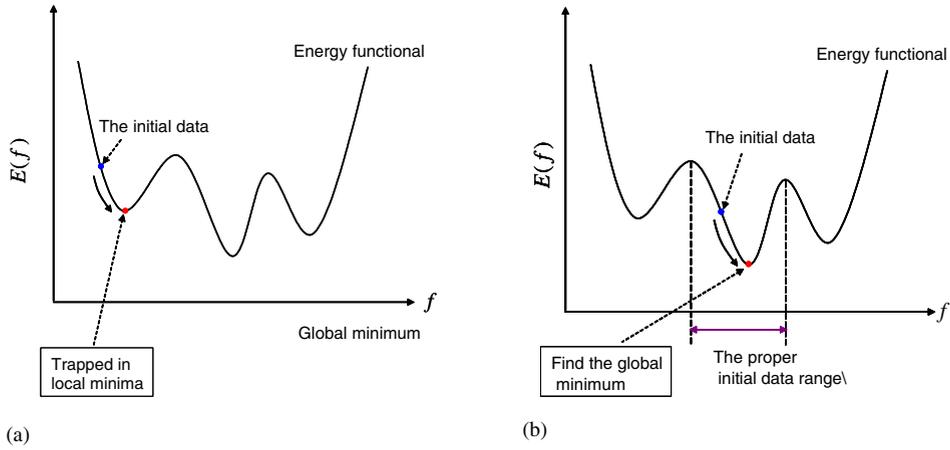


Fig. 3. Ill-posed and well-posed problems for initial data: (a) the case trapped in local minima; (b) the case finding the global minimum.

motion estimation for the initial motion field, and a region-dividing technique for the initial disparity field in the previous frame [15]. The computed motion and disparity is used as the initial solution to Eq. (7). As shown in Fig. 3, if the initial data is adjacent to the true solution, it is less probable that the equation is trapped in the local minimum. For the initial disparity field in the current frame, we simply use Eq. (2), which is also refined by the proposed energy model.

2.3. Numerical scheme

We discretize Eq. (8) to find a solution with the FDM. All spatial and temporal derivatives are approximated by forward differences and the semi-implicit scheme. The computationally more expensive solution of nonlinear systems is avoided by using the first-order Taylor expansion in an implicit discretization, as shown in

$$I_{l2}(i + u_{i,j}^{k+1}, j + v_{i,j}^{k+1}) = I_{l2}(i + u_{i,j}^k, j + v_{i,j}^k) + (u_{i,j}^{k+1} - u_{i,j}^k)I_{l2,x}(i + u_{i,j}^k, j + v_{i,j}^k) + (v_{i,j}^{k+1} - v_{i,j}^k)I_{l2,y}(i + u_{i,j}^k, j + v_{i,j}^k),$$

$$I_{l2}(i + d_{i,j}^{k+1}, j) = I_{l2}(i + d_{i,j}^k, j) + (d_{i,j}^{k+1} - d_{i,j}^k)I_{l2,x}(i + d_{i,j}^k, j). \quad (9)$$

Let

$$D(\nabla I_1) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad (10)$$

then $\nabla \cdot (D(\nabla I)\nabla u)$ can be expressed in the discrete form as follows:

$$\nabla \cdot (D_{i,j}(\nabla I)\nabla u_{i,j}) = \frac{\partial}{\partial x} \left(a_{i,j} \left(\frac{u_{i,j+1}^k - u_{i,j}^k}{h_1} \right) + b_{i,j} \left(\frac{u_{i+1,j}^k - u_{i,j}^k}{h_2} \right) \right) + \frac{\partial}{\partial y} \left(b_{i,j} \left(\frac{u_{i,j+1}^k - u_{i,j}^k}{h_1} \right) + c_{i,j} \left(\frac{u_{i+1,j}^k - u_{i,j}^k}{h_2} \right) \right). \quad (11)$$

In Eq. (11), h_1 and h_2 are the pixel sizes in the x and y directions, respectively.

As a result, Eq. (8) is discretized as follows:

$$\begin{aligned} \frac{u_{i,j}^{k+1} - u_{i,j}^k}{\tau} &= \lambda \nabla \cdot (D_{i,j}(\nabla I_{l1}) \nabla u_{i,j}) + \left\{ \left(I_{l1}(i,j) - I_{l2}(i + u_{i,j}^k, j + v_{i,j}^k) \right) - (u_{i,j}^{k+1} - u_{i,j}^k) I_{l2,x}(i + u_{i,j}^k, j + v_{i,j}^k) \right. \\ &\quad \left. - (v_{i,j}^{k+1} - v_{i,j}^k) I_{l2,y}(i + u_{i,j}^k, j + v_{i,j}^k) \right\} I_{l2,x}(i + u_{i,j}^k, j + v_{i,j}^k), \\ \frac{v_{i,j}^{k+1} - v_{i,j}^k}{\tau} &= \lambda \nabla \cdot (D_{i,j}(\nabla I_{l1}) \nabla v_{i,j}) + \left\{ \left(I_{l1}(i,j) - I_{l2}(i + u_{i,j}^k, j + v_{i,j}^k) \right) - (u_{i,j}^{k+1} - u_{i,j}^k) I_{l2,x}(i + u_{i,j}^k, j + v_{i,j}^k) \right. \\ &\quad \left. - (v_{i,j}^{k+1} - v_{i,j}^k) I_{l2,y}(i + u_{i,j}^k, j + v_{i,j}^k) \right\} I_{l2,y}(i + u_{i,j}^k, j + v_{i,j}^k), \\ \frac{d_{i,j}^{k+1} - d_{i,j}^k}{\tau} &= \lambda \nabla \cdot \nabla \cdot (D_{i,j}(\nabla I_{l1}) \nabla d_{i,j}) + \left\{ \left(I_{l1}(i,j) - I_{r1}(i + d_{i,j}^k, j) \right) - (d_{i,j}^{k+1} - d_{i,j}^k) I_{r1,x}(i + d_{i,j}^k, j) \right\} \\ &\quad \times I_{r1,x}(i + d_{i,j}^k, j), \end{aligned} \tag{12}$$

where τ is the time step size.

3. Simultaneous joint estimation with regularization

3.1. Model for simultaneous joint estimation

In Section 3, the joint model uses the relationship between the motion and disparity vectors to compute the initial disparity vector in the current frame. In this section, by applying the joint constraint to the energy model, we can acquire more correct displacement vectors and reduce the number of displacement vectors to be found. The model for simultaneous joint estimation is shown in Fig. 4. The starting point (x, y) for joint estimation exists in the left image in the previous frame, because the simultaneous joint model estimates the motion of the left and right sequences simultaneously.

3.2. Energy-based motion and disparity estimation

To apply the constraints to the motion and disparity in the energy model, we define the new data term $E_J(u, v)$ as follows:

$$E_J(u, v) = \int_{\Omega} (I_{l2}(x + u_l, y + v_l) - I_{r2}(x + d_1 + u_r, y + v_r))^2 dx dy. \tag{13}$$

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 $E_J(u, v)$. By adding the joint term, the energy model for

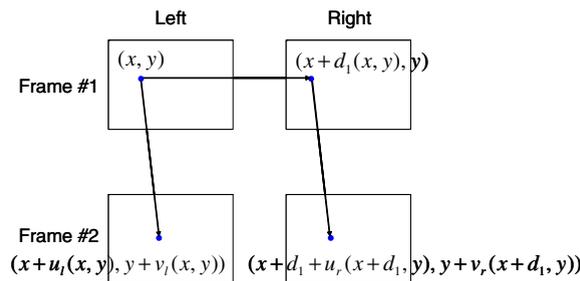


Fig. 4. Simultaneous joint estimation model.

the motion field can be defined as follows:

$$\begin{aligned}
E_{SM}(u_l, v_l, u_r, v_r) = & \int_{\Omega} (I_{l1}(x, y) - I_{l2}(x + u_l, y + v_l))^2 + (I_{r1}(x + d_1, y) - I_{r2}(x + d_1 + u_r, y + v_r))^2 \\
& + (I_{l2}(x + u_l, y + v_l) - I_{r2}(x + d_1 + u_r, y + v_r))^2 dx dy \\
& + \lambda_1 \int_{\Omega} \text{trace}(\nabla h_l)^T D(\nabla I_{l1})(\nabla h_l) dx dy \\
& + \lambda_2 \int_{\Omega} \text{trace}(\nabla h_r(x + d_1, y))^T D(\nabla I_{r1})(\nabla h_r(x + d_1, y)) dx dy,
\end{aligned} \tag{14}$$

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

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 $E_J(u, v)$ 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 Section 3, 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 $E_J(u, v)$, 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:

$$d_{y2}(x + u_l, y + v_l) = v_r(x + d_1, y) - v_l(x, y) \cong 0, \tag{15}$$

$$v_r(x + d_1, y) = v_l(x, y). \tag{16}$$

The minimization of Eq. (14) yields the following associated Euler–Lagrange equation with Neumann boundary conditions:

$$\begin{aligned}
\lambda \operatorname{div}(D(\nabla I_{l1}(x, y))\nabla u_l(x, y)) + (I_{l1}(x, y) - 2I_{l2}(x + u_l, y + v_l) + I_{r2}(x + d_1 + u_r, y + v_r)) \frac{\partial I_{l2}(x + u_l, y + v_l)}{\partial x} &= 0, \\
\lambda \operatorname{div}(D(\nabla I_{l1}(x, y))\nabla v_l(x, y)) + (I_{l1}(x, y) - 2I_{l2}(x + u_l, y + v_l) + I_{r2}(x + d_1 + u_r, y + v_r)) \frac{\partial I_{l2}(x + u_l, y + v_l)}{\partial y} &= 0, \\
\lambda \operatorname{div}(D(\nabla I_{l1}(x + d_1, y))\nabla u_r(x + d_1, y)) + (I_{r1}(x + d_1, y) - 2I_{r2}(x + d_1 + u_r, y + v_r) \\
+ I_{l2}(x + u_l, y + v_l)) \frac{\partial I_{r2}(x + d_1 + u_r, y + v_r)}{\partial x} &= 0.
\end{aligned} \tag{17}$$

We obtain the solutions to the Euler–Lagrange equations by calculating the asymptotic state ($t \rightarrow \infty$) of the parabolic system.

$$\begin{aligned}
\frac{\partial u_l}{\partial t} = & \lambda \operatorname{div}(D(\nabla I_{l1}(x, y))\nabla u_l(x, y)) + (I_{l1}(x, y) - 2I_{l2}(x + u_l, y + v_l) \\
& + I_{r2}(x + d_1 + u_r, y + v_r)) \frac{\partial I_{l2}(x + u_l, y + v_l)}{\partial x}, \\
\frac{\partial v_l}{\partial t} = & \lambda \operatorname{div}(D(\nabla I_{l1}(x, y))\nabla v_l(x, y)) + (I_{l1}(x, y) - 2I_{l2}(x + u_l, y + v_l) \\
& + I_{r2}(x + d_1 + u_r, y + v_r)) \frac{\partial I_{l2}(x + u_l, y + v_l)}{\partial y},
\end{aligned}$$

$$\begin{aligned} \frac{\partial u_r(x + d_1, y)}{\partial t} = & \lambda \operatorname{div}(D(\nabla I_{l1}(x + d_1, y))\nabla u_r(x + d_1, y)) + (I_{r1}(x + d_1, y) - 2I_{r2}(x + d_1 + u_r, y + v_r) \\ & + I_{l2}(x + u_l, y + v_l)) \frac{\partial I_{r2}(x + d_1 + u_r, y + v_r)}{\partial x}. \end{aligned} \quad (18)$$

The simultaneous joint estimation is performed as follows:

1. Compute the initial motion and disparity vectors of the previous frame using hierarchical block-based motion estimation and a region-dividing disparity estimation.
2. Estimate the disparity vector of the previous frame $d_1(x, y)$ using the initial disparity vector computed in step 1.
3. Estimate the left and right motion vectors using the initial motion vector computed in step 1 and the regularized disparity vector computed in step 2.
4. Compute the backward motion vector of the left frame.
5. Compute the initial disparity vector of the current frame.
6. Estimate the disparity of current frame using initial disparity computed in step 5.

3.3. Numerical scheme

We also discretize Eq. (18) using a FDM. The first-order Taylor expansion used in an implicit discretization is shown as

$$\begin{aligned} I_{l2}(i + u_{l,i,j}^{k+1}, j + v_{l,i,j}^{k+1}) = & I_{l2}(i + u_{l,i,j}^k, j + v_{l,i,j}^k) + (u_{l,i,j}^{k+1} - u_{l,i,j}^k)I_{l2,x}(i + u_{l,i,j}^k, j + v_{l,i,j}^k) \\ & + (v_{l,i,j}^{k+1} - v_{l,i,j}^k)I_{l2,y}(i + u_{l,i,j}^k, j + v_{l,i,j}^k), \\ I_{r2}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^{k+1}, j + v_{r,i+d_{1,i,j},j}^{k+1}) = & I_{r2}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^k, j + v_{r,i+d_{1,i,j},j}^k) + (u_{r,i+d_{1,i,j},j}^{k+1} - u_{r,i+d_{1,i,j},j}^k)I_{r2,x}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^k, \\ & + v_{r,i+d_{1,i,j},j}^k) + (v_{r,i+d_{1,i,j},j}^{k+1} - v_{r,i+d_{1,i,j},j}^k)I_{r2,y}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^k, j + v_{r,i+d_{1,i,j},j}^k). \end{aligned} \quad (19)$$

The final solution can be found in a recursive manner by using the following equation:

$$\begin{aligned} \frac{u_{l,i,j}^{k+1} - u_{l,i,j}^k}{\tau} = & \lambda \nabla \cdot (D_{i,j}(\nabla I_{l1})\nabla u_{l,i,j}) + \{I_{l1}(i, j) - 2I_{l2}(i + u_{l,i,j}^{k+1}, j + v_{l,i,j}^{k+1}) \\ & + I_{r2}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^{k+1}, j + v_{r,i+d_{1,i,j},j}^{k+1})\}I_{l2,x}(i + u_{l,i,j}^k, j + v_{l,i,j}^k), \\ \frac{v_{l,i,j}^{k+1} - v_{l,i,j}^k}{\tau} = & \lambda \nabla \cdot (D_{i,j}(\nabla I_{l1})\nabla v_{l,i,j}) + \{I_{l1}(i, j) - 2I_{l2}(i + u_{l,i,j}^{k+1}, j + v_{l,i,j}^{k+1}) \\ & + I_{r2}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^{k+1}, j + v_{r,i+d_{1,i,j},j}^{k+1})\}I_{l2,y}(i + u_{l,i,j}^k, j + v_{l,i,j}^k), \\ \frac{u_{r,i+d_{1,i,j},j}^{k+1} - u_{r,i+d_{1,i,j},j}^k}{\tau} = & \lambda \nabla \cdot (D_{i,j}(\nabla I_{r1})\nabla u_{r,i+d_{1,i,j},j}) + \{I_{r1}(i + d_{1,i,j}, j) + I_{l2}(i + u_{l,i,j}^{k+1}, j + v_{l,i,j}^{k+1}) \\ & - 2I_{r2}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^{k+1}, j + v_{r,i+d_{1,i,j},j}^{k+1})\}I_{r2,x}(i + d_{1,i,j} + u_{r,i+d_{1,i,j},j}^k, j + v_{r,i+d_{1,i,j},j}^k). \end{aligned} \quad (20)$$

4. Experimental results

The experiment was performed on the stereoscopic sequences, which is “Boy” of size 320×240 and “Man” of size 256×256 and “Aqua” of size 320×256 in Fig. 5. The stereoscopic sequences “Boy” were captured with a DigiclopsTM of Point Gray Research Inc., which provides rectified stereo sequences from parallel stereo

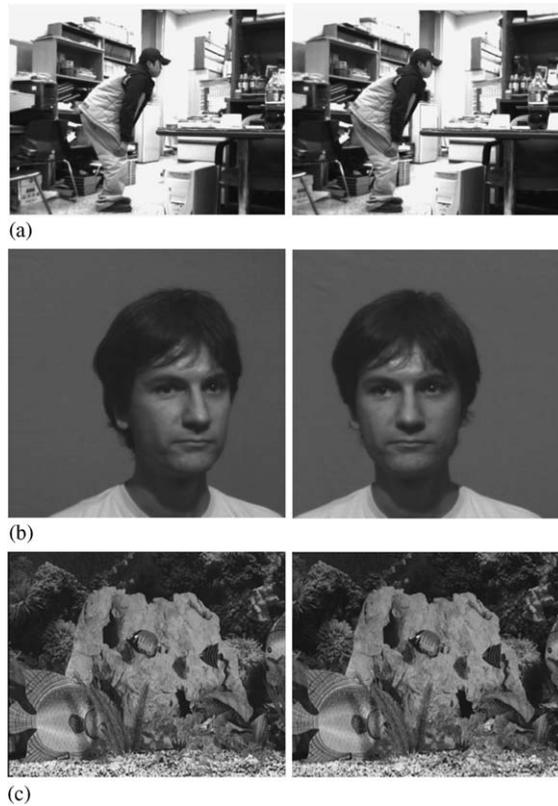


Fig. 5. Test image sequences in previous frames: (a) “Boy” images; (b) “Man” images; (c) “Aqua” images.

Table 1
Parameters used in joint estimation

Parameter	Values
Weighting factor (disparity)	$\lambda = 2000$
Weighting factor (motion)	
Joint model	$\lambda = 6000$
Simultaneous model	$\lambda_1, \lambda_2 = 10,000$
Gradient step size	$\delta_1 = 3/\delta_d = 1$
Time step size	$\tau = 0.0001$
Isotropy fraction	$s = 0.3$
Number of iteration	$T = 600$

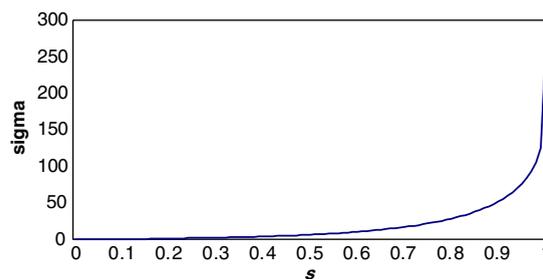


Fig. 6. Distribution function of σ according to isotropy fraction s in “Boy” images.

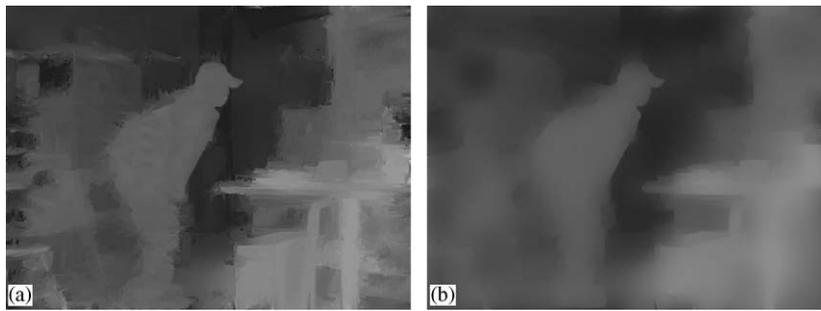


Fig. 7. Disparity map in previous frame of “Boy” images according to s : (a) when $s = 0.1$; (b) when $s = 0.7$.

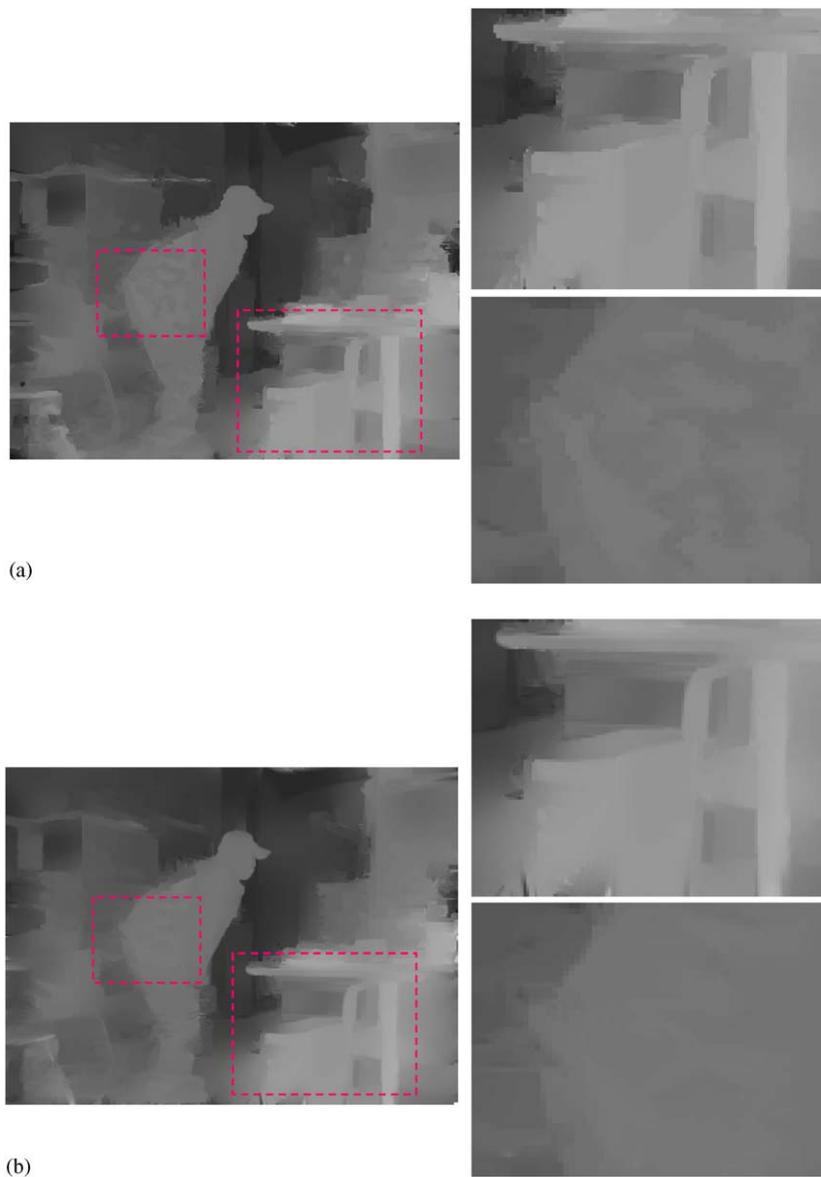


Fig. 8. Comparison with disparity map using isotropic and anisotropic operator of “Boy” images: (a) Geman & McClure’s isotropic operator; (b) anisotropic operator.

cameras. The focal length of the DigiclopsTM is 6 mm, and the baseline distance between the two lenses is 10 cm. In order to visualize motion and disparity vectors, we used gray level images which consist of values between 0 and 255 for disparity maps and are defined by $128 + 8u$ and $128 + 8v$ for x and y motion maps, respectively.

The parameters used in the experiment for motion and disparity estimations are shown in Table 1. Most of these parameters were selected empirically and intuitively. The values of the weighting factor λ were defined differently to the motion and disparity estimations in each model. λ was 2000 in the disparity model, 6000 in the motion model, and 10,000 in the simultaneous motion model. We analyzed the influence of the weighting factor in the motion and disparity vectors. The step size τ was 0.0001, because larger values sometimes cause divergent solutions to the equations. In the case of the gradient step size, we applied different sizes to the gradient of an image and that of the fields, because the values of the fields are more sensitive to the results than those of the images. The isotropy fraction was 0.3, which means that the energy model showed anisotropic diffusion.

To evaluate the performance of the model with the anisotropic diffusion operator, we compared the results to those found when using the isotropic diffusion operator. The diffusivity function, proposed by Geman & McClure, was used as the isotropic diffusion operator [14,18].

4.1. Performance analysis of the anisotropic diffusion operator

Fig. 6 shows how σ changes for the value of s in Eq. (6). The σ is the anisotropic diffusion constant, and s is the isotropy fraction. In regions where $|\nabla I| \gg \sigma$, the model restrains the blurring of the fields across the boundaries of I . The amount of the blurring is determined based on s , which shows a value between 0 and 1. In the experiment, we defined s as 0.3. This means that the regularization term diffuses isotropically at 30% of all image locations. If s is too small or too large, the displacement vector can either be under-smoothed or over-smoothed, as shown in Fig. 7. When $s = 0.1$, the discontinuities are preserved but the disparity field is not smooth enough. When $s = 0.7$, the disparity field is over-smoothed and the discontinuities are not preserved well.

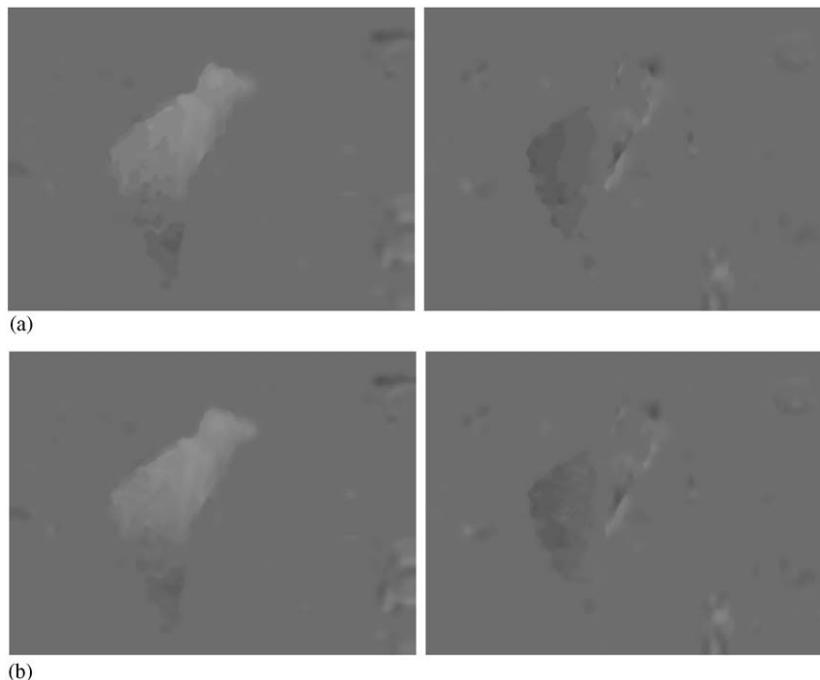


Fig. 9. Comparison with x and y motion map using isotropic and anisotropic operator of “Boy” images (only left sequences): (a) Geman & McClure’s isotropic operator; (b) anisotropic operator.

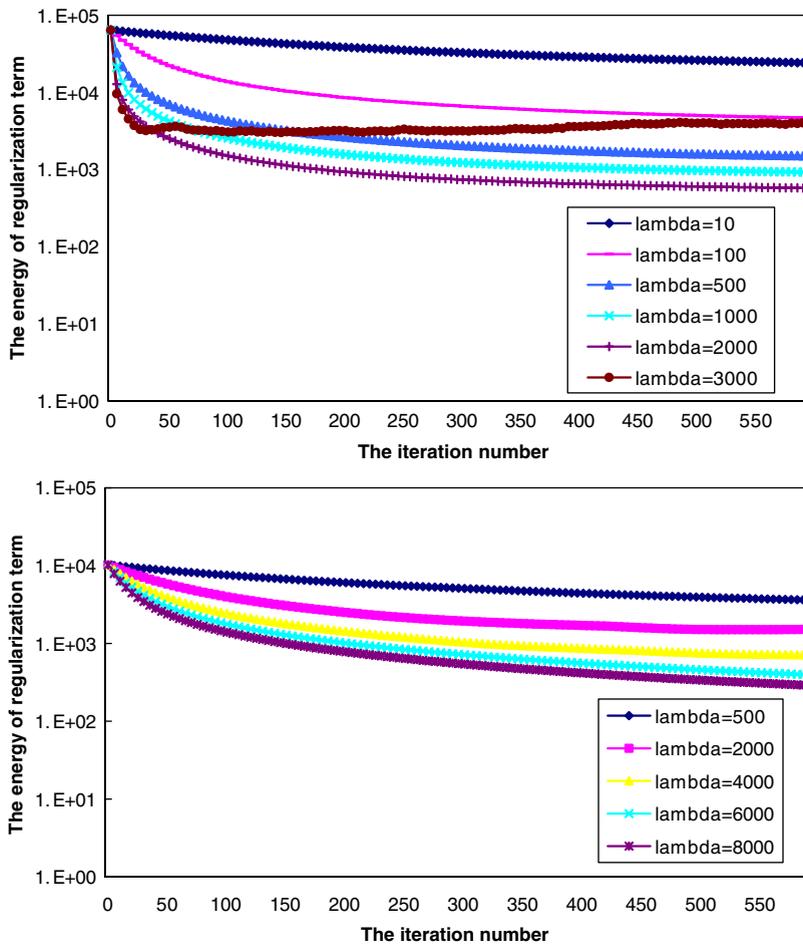


Fig. 10. Energy of regularization term of disparity and motion varying with λ in “Boy” images: (a) energy of regularization term of disparity; (b) energy of regularization term of motion.

Fig. 8 shows the disparity field in the previous frame, estimated with energy models that used isotropic and anisotropic operators. We found that the energy model that used an anisotropic operator showed more accurate performance in the estimated disparity fields and discontinuity localizations. These characteristics can be seen in the disparity field of the back of the man, and in the table and chair in the right part. Fig. 9 shows the x and y motions of the model when using isotropic and anisotropic operators. Compared with that of the isotropic diffusion model, the motion around the cap shows a more distinct difference.

Fig. 10(a) shows the energy of the regularization term of the disparity field according to λ . In order to see the difference of the convergence rate according to the λ , we draw behaviors of energy as a graph on the logarithmic scale. Since the difference in the data term is insignificant to the λ , we only show the graph of the regularization term. As λ increases, the convergence rate also increases, and it shows a maximum value when $\lambda = 2000$. However, when λ is 3000, the regularization term diverges and this causes the solution to become trapped in the local minima. For small or large λ , the convergence rate is reduced because the model overemphasizes the data term or the regularization term. Fig. 10(b) shows the energy of the regularization term for motion according to λ . This graph is also drawn on the logarithmic scale. As λ increases, the convergence rate also increases. The energy of regularization term in the motion model is smaller than that in the disparity model because the magnitude of motion is relatively small. A large λ causes the motion to become over-smoothed because the model overemphasizes the regularization term. For this reason, we define λ as 6000.

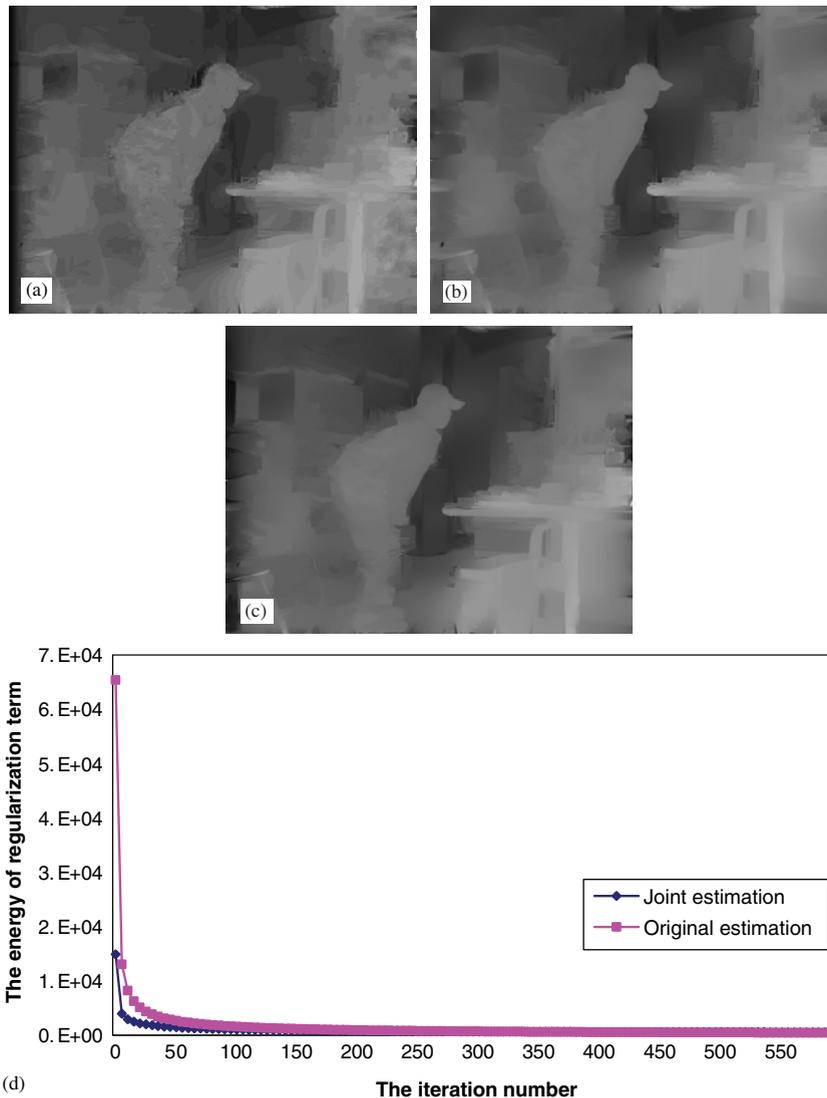


Fig. 11. Initial disparity, regularized disparity of joint and original estimation in “Boy” images: (a) initial disparity; (b) regularized disparity of joint estimation; (c) disparity of original estimation; (d) energy of regularization term of joint and original estimation.

4.2. Joint estimation with regularization

Given the disparity in the previous frame and the motion vectors in the left and right sequences, the initial disparity in the current frame is computed using the relation of the motion and disparity vectors in the stereo sequences. Fig. 11 shows the initial disparity and the regularized disparity vectors in “Boy” images. The regularized disparity vectors are similar to those of the original estimation, which used disparity vectors (computed by the region-dividing technique) as the initial data. Moreover, because the initial disparity vector was computed using accurate disparity and motion vectors, the initial disparity vector acquired by joint estimation is better than that of the original estimation. In Fig. 11(d), an error in the initial state is smaller than that obtained with the original estimation, and this leads to the conclusion that the model converges faster to the minimum point.

Fig. 12 shows the motion and disparity of “Man” images. Actually, the “Man” images were captured by a toed-in camera system so that their epipolar lines are not exactly parallel, but we applied the same horizontal

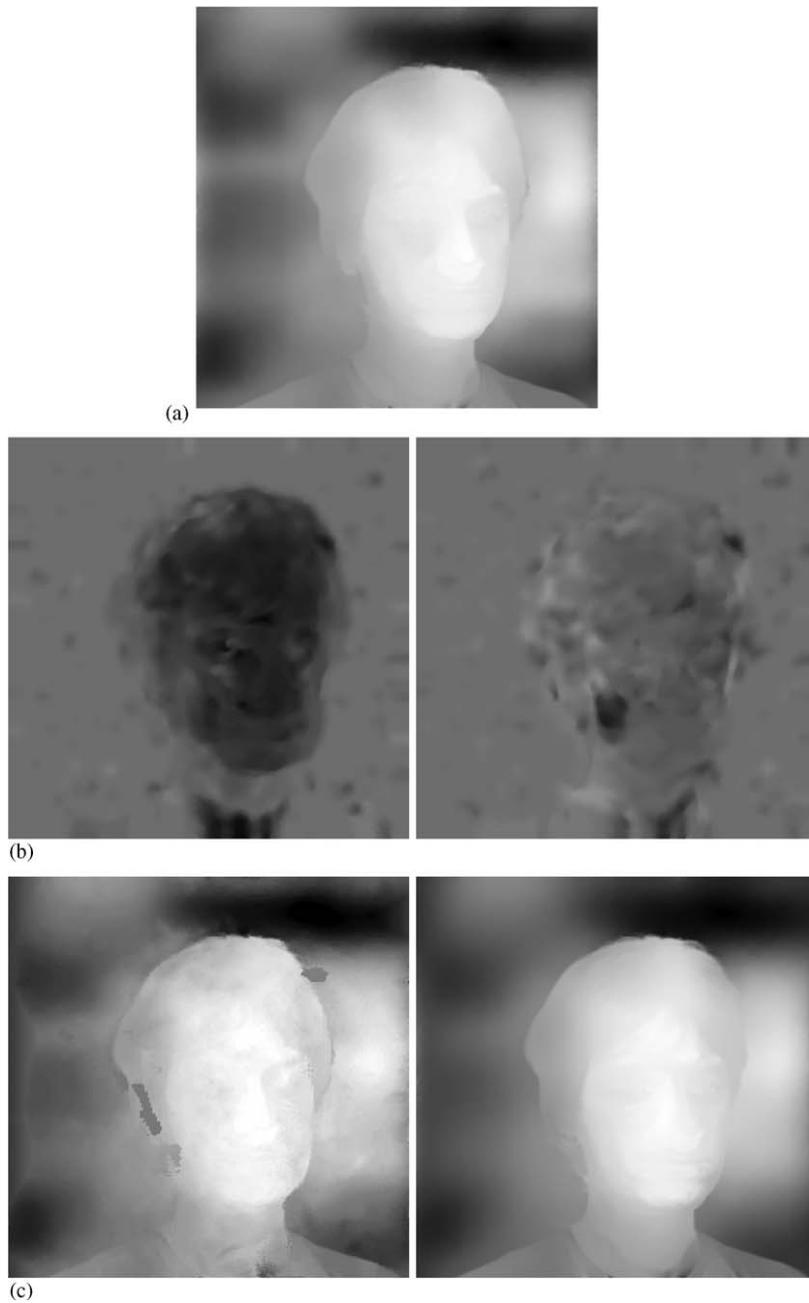


Fig. 12. Disparity and motion (for left sequence only) map in “Man” images: (a) disparity of previous frame; (b) x and y motion of left sequence; (c) initial and regularized disparity of current frame.

scanline search for simplicity because the images are small and the main object is placed at the center of the image where epipolar line distortion is not serious. Since the background region is completely textureless, this leads to the conclusion that the energy model is trapped in the local minima. This influences the initial disparity vector in the current frame, but the regularized disparity is also estimated in the region sporadically. However, though “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

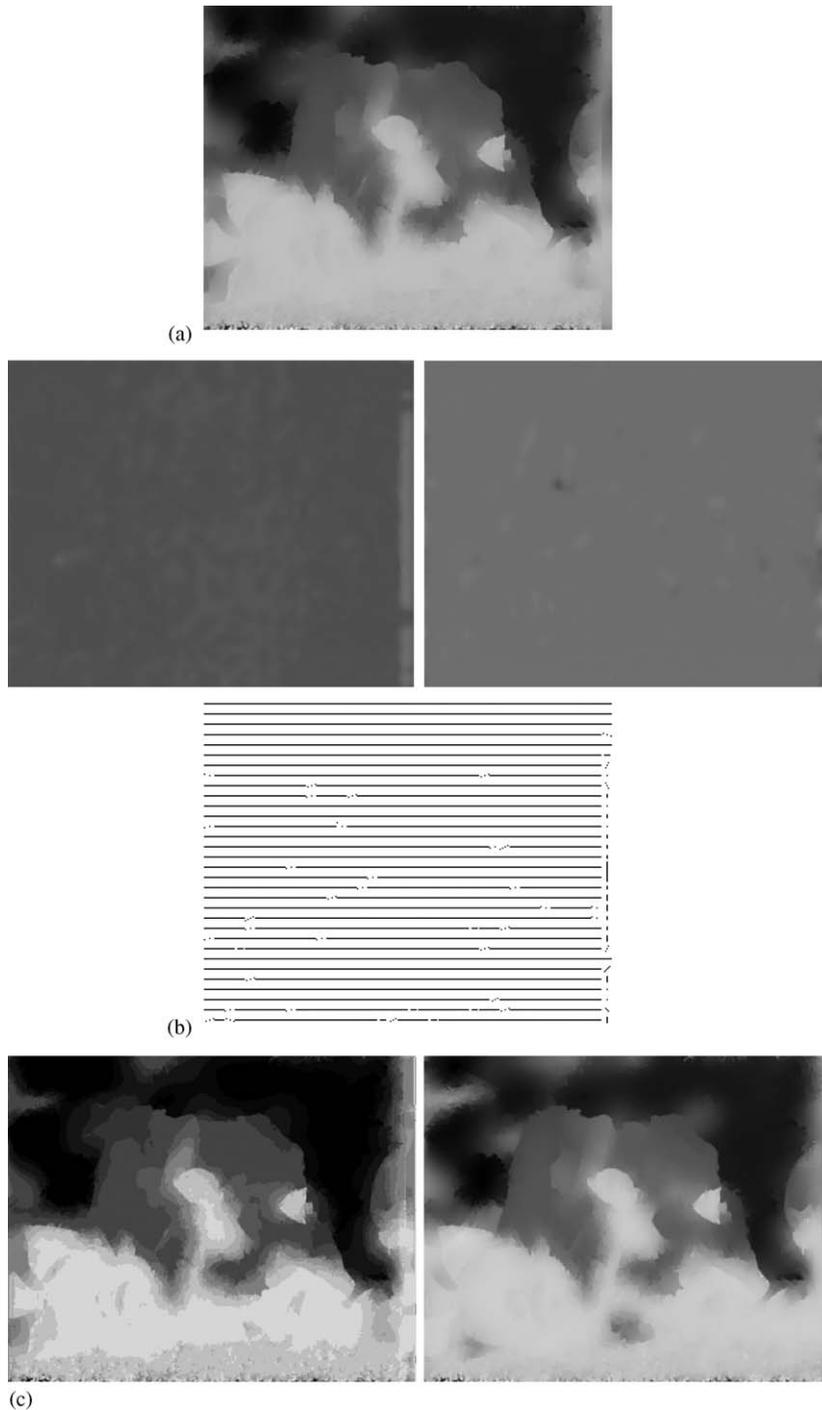


Fig. 13. Disparity and motion (for left sequence only) map in “Aqua” images: (a) disparity of previous frame; (b) x and y motion of left sequence; (c) initial and regularized disparity of current frame.

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. 13 shows the motion and disparity of “Aqua” images. The visualization method was same as Fig. 12, and vector plot of motion map was added. In these sequences, there is a global camera panning and small fish

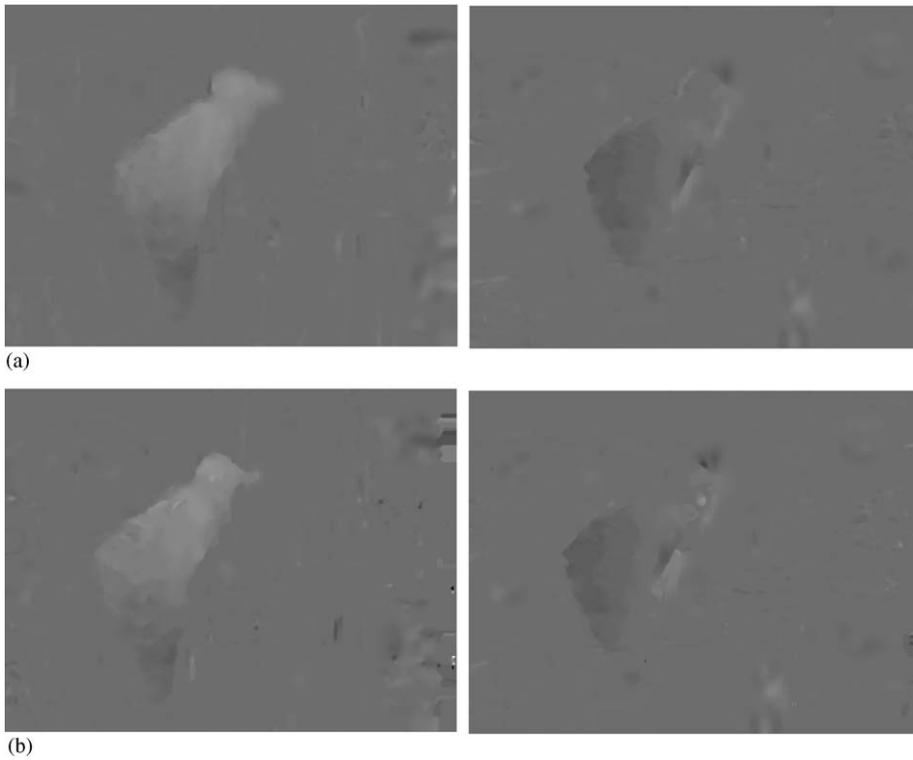


Fig. 14. x and y motion of simultaneous model in “Boy” images: (a) x and y motion of left sequence in simultaneous model; (b) x and y motion of right sequence in simultaneous model.

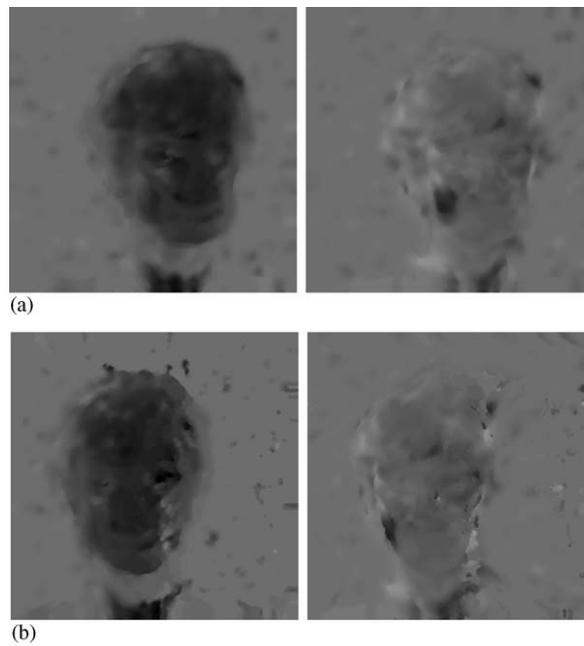


Fig. 15. x and y motion of simultaneous model in “Man” images: (a) x and y motion of left sequence in simultaneous model; (b) x and y motion of right sequence in simultaneous model.

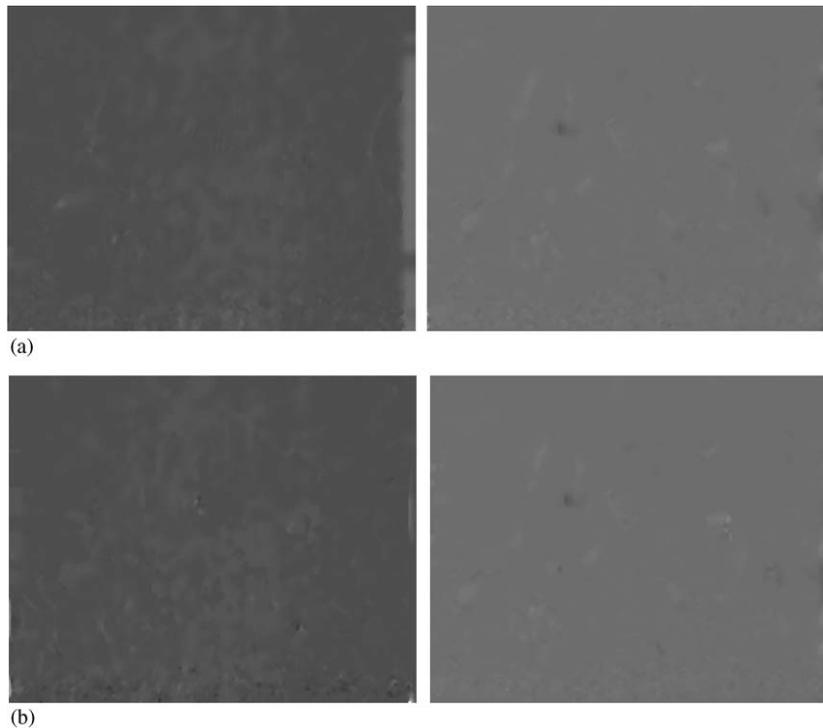


Fig. 16. x and y motion of simultaneous model in “Aqua” images: (a) x and y motion of left sequence in simultaneous model; (b) x and y motion of right sequence in simultaneous model.

motion. The vector plot of motion map shows that both background and object move in x -axis direction. In disparity map, we found that the energy model showed accurate performance in the estimated disparity fields and discontinuity localizations. These characteristics can be seen in the discontinuities of the disparity field between rock and fishes. Moreover, the initial wrong disparity vector produced by joint estimation process was eliminated in the regularization process, for example, at the bottom of the image.

4.3. Simultaneous joint estimation with regularization

In this model, motions in the left and right sequences are simultaneously estimated. By adding the joint term to the energy model, errors generated from joint estimation are reduced, so we are able to compute the y -motion in the right sequence from that in the left sequence directly. Since this model uses the disparity vector in the previous frame, the disparity vector should be correctly estimated at first. After the motion vectors are simultaneously estimated, the remaining process is the same as when using only the joint model (see Section 3).

Figs. 14–16 show the left and right motions in “Boy”, “Man” and “Aqua” images, respectively. The y -motion in the right sequence is similar to that in the joint model, and this means that the y -motion in the right sequence can be computed using that in the left sequence. We can see that the simultaneous model also estimates smooth and edge-preserving displacement vectors.

Fig. 17(a) shows the energy of the regularization term of the simultaneous model according to λ . 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 10,000. As repeating this iteration scheme, the energy of the data term increases because of the smoothing effect of the regularization term in Fig. 17(b). The smoothing effect reduces the similarity between the corresponding points, so this leads to an increase of the energy of the data term. However, the overall energy decreases for the reduction of energy of the regularization term. In the simultaneous model, the energy of the data term of the right sequence is larger

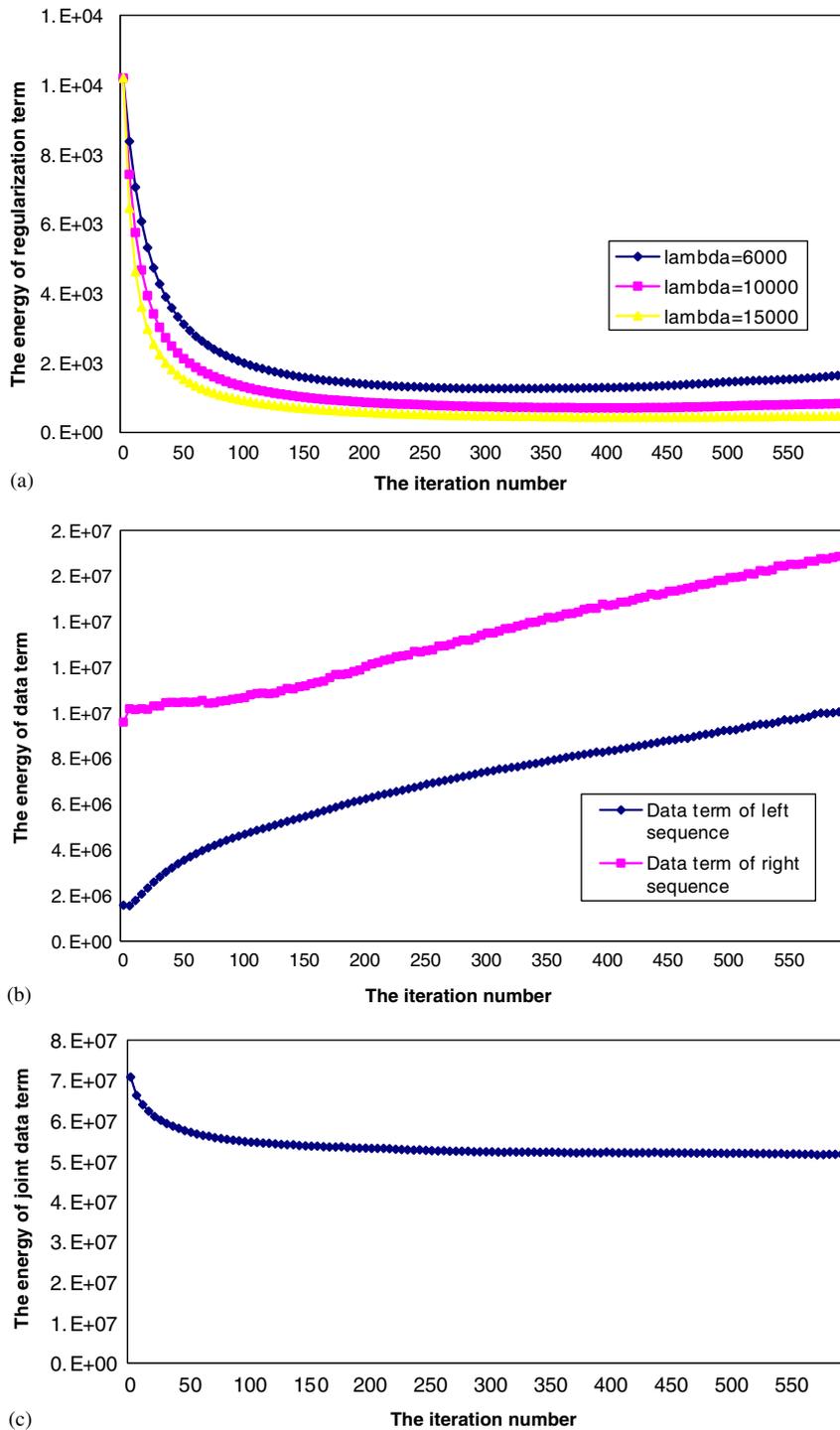


Fig. 17. Energy of regularization and data term of motion in simultaneous model in “Boy” images: (a) energy of regularization term of motion; (b) energy of data term of left and right sequence ($\lambda = 10,000$); (c) energy of joint term ($\lambda = 10,000$).

than that in the left sequence, because the left motion vector is estimated for point (x, y) but the right motion vector is estimated for point $(x + d, y)$. In Fig. 17(c), the energy of the joint data term decreases, showing a reduction in the perturbations from the joint estimation.

5. Conclusions

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 real stereo image sequences. The energy-based estimation model consists of a data term and a regularization term, and the weighting factor λ controls the amount of smoothing. At first, we used an anisotropic diffusion operator, which considers both the magnitude and the direction of an edge to improve edge-preserving performance. Secondly, to increase efficiency, we estimated the motion and disparity jointly using the relation of the motion and disparity vectors in the stereo image sequences. Since the model did not require time-consuming initial disparity estimation, the processing time was reduced. Moreover, since the energy of the initial disparity vector obtained with joint estimation was lower than that of the initial disparity vector obtained with conventional techniques, we acquired a higher convergence rate. Finally, in the simultaneous joint estimation model, we proposed the joint data term to increase reliability and estimated the motion vectors in the left and right sequences simultaneously. With the joint data term, the error of the vertical disparity vector in the current frame was reduced, and thus we were able to obtain the y -motion in the right sequence from that in the left sequence.

As a conclusion, the proposed edge-preserving joint motion and disparity estimation showed good edge-preserving performance and a fast convergence rate. A more correct displacement vector can be obtained by using an anisotropic diffusion operator, and it is possible to increase efficiency and accuracy by using the joint estimation model. However, since parameters such as weighting factor, time step size, and the isotropy fraction were acquired experimentally, they were not robust to the characteristic of an image. Therefore, we need more research on parameter estimation. In addition, another numerical method is necessary in order to improve the convergence rate.

Acknowledgments

This work is financially supported by the Ministry of Education and Human Resources Development (MOE), the Ministry of Commerce, Industry and Energy (MOCIE) and the Ministry of Labor (MOLAB) through the fostering project of the Lab of Excellency, and is in part supported by the MIC (Ministry of Information and Communication), Korea under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment)" (IITA-2005-(C1090-0502-0027)).

References

- [1] G.J. Klein, Ronald, H. Huesman, A 3D optical flow approach to addition of deformable PET volumes, Proc. IEEE Motion Non-Rigid Articulated Objects (1997) 136–144.
- [2] M. Krüger, A. Pesavento, H. Ermert, K.M. Hiltawsky, L. Heuser, A. Jensen, Ultrasonic strain imaging of the female breast using phase root seeking and three-dimensional optical flow, Proc. IEEE Ultrasonic Symp. (1998) 1757–1760.
- [3] S. Vedula, S. Baker, P. Rander, R. Collins, T. Kanade, Three-dimensional scene flow, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 475–480.
- [4] L. Zhou, C. Kambhamettu, D. Goldgof, Extracting nonrigid motion and 3-D structure of hurricanes from satellite image sequences without correspondences, Proc. IEEE CVPR (1999) 280–285.
- [5] Y. Zhang, C. Kambhamettu, On 3-D scene flow and structure recovery from multiview image sequences, IEEE Trans. Syst. Man Cybern. Part B 33 (2003) 592–606.
- [6] J. Cox, L. Hingorani, B. Rao, A maximum likelihood stereo algorithm, Comput. Vision Image Understand. 63 (3) (1996) 542–576.
- [7] L. Alvarez, R. Deriche, J. Sanchez, J. Weickert, Dense disparity map estimation respecting image discontinuities: a PDE and scale-space based approach, J. Visual Comm. Image Represent. 13 (2002) 3–21.
- [8] L. Alvarez, J. Weickert, J. Sanchez, Reliable estimation of dense optical flow fields with large displacements, IJCV 39 (1) (2000) 41–56.
- [9] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.
- [10] B. Horn, B. Schunck, Determining optical flow, Artif. Intell. 17 (1981) 185–203.
- [11] H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 8 (1986) 565–593.

- [12] M. Izquierdo, Stereo matching for enhanced telepresence in three-dimensional videocommunications, *IEEE Trans. Circ. Syst.* 7 (4) (1997) 629–643.
- [13] I. Patras, N. Alvertos, G. Tziritas, Joint disparity and motion field estimation in stereoscopic image sequences, *Proc. IEEE ICPR* (1996) 359–363.
- [14] J. Weickert, A review of nonlinear diffusion filtering, *Lecture Notes in Computer Science*, vol. 1252, 1997, pp. 3–28.
- [15] H. Kim, Y. Choe, K. Sohn, Disparity estimation using a region-dividing technique and energy-based regularization, *Opt. Eng.* 43 (8) (2004) 1882–1890.
- [16] M. Brown, G. Hager, D. Burschka, Advances in computational stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (8) (2003) 993–1008.
- [17] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *IJCV* 47 (2002) 7–42.
- [18] S. Geman, D.E. McClure, Bayesian image analysis: an application to single photon emission tomography, *Proc. Stat. Comput. Sect.* (1985) 12–18.