A VARIATIONAL FRAMEWORK FOR SIMULTANEOUS MOTION AND DISPARITY ESTIMATION IN A SEQUENCE OF STEREO IMAGES

Wided Miled, Béatrice Pesquet-Popescu and Wael Chérif

TELECOM ParisTech, Signal and Image Processing Department
46 rue Barrault, 75634 Paris Cédex 13, France
e-mail: {miled, pesquet, cherif}@telecom-paristech.fr

ABSTRACT

In this paper, we present a variational framework for joint disparity and motion estimation in a sequence of stereo images. The problem involves the estimation of four dense fields: two motion fields and two disparity fields. In order to reduce computational complexity and improve estimation accuracy, the two motion fields, for the left and right sequences, and the disparity field of the current stereo pair are jointly estimated, using the *stereo-motion consistency* constraint. In the proposed variational framework, the joint estimation problem is formulated as a convex programming problem in which a convex objective function is minimized under specific convex constraints. This minimization is achieved using an efficient parallel block-iterative algorithm. Experimental results involving real stereo sequences indicate the feasibility and robustness of our approach.

Index Terms— Stereo sequences, Disparity, Motion, Joint estimation, Convex optimization, Regularization

1. INTRODUCTION

The recovery of depth and motion information from a sequence of stereo images is a very common task in several applications in computer vision, including 3D tracking, stereo video coding, 3D scene interpretation and 3D television. Motion estimation uses two consecutive frames from a video sequence, whereas disparity estimation is performed within a pair of stereo images taken from distinct viewpoints. Finding an accurate correspondence between points in two stereoscopic or temporal sequence images is the most important and difficult step in both depth and motion estimation.

The problem of establishing spatial or temporal correspondences between pixels has been investigated for many years [1, 2]. A number of studies have therefore been reported including feature-based, area-based, and energy-based approaches. Feature-based approaches are those which use invariant geometric primitives and match extracted salient features, such as edges, corners or regions. They provide accurate results, as the features are discriminant and have many attributes, but provide sparse displacement results. Area-based approaches match image pixels based on their positions and intensity values. They offer the advantage of directly generating dense displacement fields by correlation over local windows, but often fail around discontinuities and in textureless areas. The energy-based approaches are mainly based on optimizing a global energy function, which is typically the sum of a data term and a smoothness term. These global approaches also produce dense displacement results, but are more accurate than area-based approaches, particularly in the challenging image regions like occlusions. Recently, many powerful global stereo and motion matching algorithms have been developed based on dynamic programming [3, 4], graph cuts [5] or belief propagation [6]. Variational approaches have also been very effective for solving the correspondence problem globally [7, 8, 9].

The estimation of motion and disparity in stereo image sequences has a high computational cost, especially when a global approach is used to compute dense and accurate solutions. One way to overcome this problem is to jointly estimate disparity and motion fields using the *stereo-motion consistency* constraint, which relates the four displacement fields (two motion fields and two disparity fields) involved in each two consecutive stereo frames (see Figure 1). Based on this constraint, the disparity in the current frame can be deduced from the estimated left and right motions and the disparity in the previous frame. This results in a reduction of complexity as well as an improvement of estimation performance. Several approaches have recently been proposed for combining stereo and motion analysis within a sequence of stereo images [10, 11, 12]. In [13], the joint estimation was performed on a multi-resolution pyramid of images using an anisotropic diffusion regularization to preserve image boundaries. In [14], the authors proposed a multi-scale iterative relaxation algorithm to first calculate the disparity field of the first stereoscopic pair. Using the computed disparity field and the consistency constraint, the two motion fields are estimated together with a partially constructed current disparity field, which is refined later using the same multi-scale relaxation algorithm. In [15], an edge-preserving regularization algorithm that simultaneously calculates dense disparity and motion fields is proposed. The authors use the Euler-Lagrange equations within a variational framework to minimize a global edge-preserving energy function. Although interesting results were reported, the discretization of the PDE, using a finite difference method, is a difficult and numerically instable task.

In this work, we propose a variational optimization method for jointly estimating dense disparity and motion vector fields from two consecutive stereo frames. Based on the *stereo-motion consistency* constraint, a global energy function is minimized under various convex constraints, to simultaneously estimate left and right motion fields. The disparity of the current stereo pair is implicitly constructed by applying the joint consistency constraint, and is refined later using the dense disparity estimation method we proposed in [9]. Since motion fields vary smoothly in homogeneous regions and change abruptly around object boundaries, we use an edge preserving regularizing constraints based on the Total Variation measure, which has already proven to be very useful in image recovery and denoising problems [17], so motivating its use in the field of variational stereo [9] and optical flow methods [7]. Within the proposed set theoretic framework, the joint estimation problem is solved using a parallel block iterative decomposition method, which provides dense and accurate displacement fields and offers great flexibility for incorporating several a priori constraints.

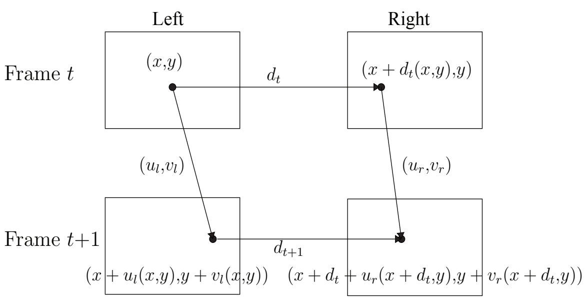


Fig. 1. The stereo-motion consistency constraint.

The outline of the paper is as follows. In Section 2 the relation between motion and disparity is presented. In Section 3, we describe the simultaneous estimation framework we propose. Section 4 presents experimental results, and Section 5 concludes the work.

2. JOINT ESTIMATION MODEL

A sequence of stereo images, obtained from two cameras separated by a fixed baseline distance, shows the temporal evolution of a 3D scene from two slightly different viewpoints. In order to allow for an accurate stereo sequence analysis, it is essential to exploit the spatio-temporal relationships that exist between the different images of the sequence. Joint estimation of disparity and motion displacement fields is an efficient way to benefit from these relationships, leading therefore to improving results while reducing the computational cost.

Consider two consecutive stereo image pairs, denoted by I_l^t, I_r^t, I_l^{t+1} and I_r^{t+1}, which are, respectively, the left and right views of the previous and current frames of a stereo image sequence. The stereo pairs are assumed to be rectified, so that the geometry of the cameras can be considered as horizontal epipolar [16]. Let $\mathbf{v}_l = (u_l, v_l)$ and $\mathbf{v}_r = (u_r, v_r)$ be the left and right motion fields, and $\mathbf{d}_t$ and $\mathbf{d}_{t+1}$ designate disparity vector fields of the stereo image pair at t and $t+1$. If these fields relate to the projections of the same physical point in the scene, the following constraint must hold:

$$\mathbf{d}_t + \mathbf{v}_r - \mathbf{d}_{t+1} - \mathbf{v}_l = 0. \tag{1}$$

This constraint, illustrated in Figure 1, establishes the relationship between motion vectors and disparity vectors. Assuming that the spatial point is projected to the pixel $s = (x, y)$ on frame I_l^t, Eq. (1) can be rewritten as follows:

$$\begin{cases} d_{t+1}(s + \mathbf{v}_l(s)) \simeq u_r(x + d_t(s), y) + d_t(s) - u_l(s), \\ v_r(x + d_t(s), y) \simeq v_l(s). \end{cases} \tag{2}$$

Using the above constraint, the disparity field obtained at time t can be used to simultaneously estimate left and right motion fields. The disparity field at time $t+1$ is then implicitly constructed using the three computed fields and the first equality in (2). However, as this equality is only approximate because of occlusions and accumulation of errors, we only use it to provide an initial disparity field, which will be refined later through the convex optimization approach we proposed in [9]. Furthermore, according to Eq. (2), the vertical motion in the right sequence can be deduced directly from that in the left sequence. So, by applying the stereo-motion consistency constraint, we improve estimation accuracy and reduce the number of displacement vectors to be computed.

3. SIMULTANEOUS MOTION AND DISPARITY ESTIMATION

The joint estimation framework we propose in this work consists of estimating the left and right motion vectors and the disparity at time $t+1$. The disparity at time t is considered as known, that is previously estimated. The left and right motion vectors are simultaneously estimated using a convex programming approach, in which a quadratic objective function is minimized subject to specific convex constraints.

3.1. Energy model for joint motion estimation

Assuming that the four corresponding points, that are the projections of the same spatial point, have identical intensity values, the left and right motion fields can be computed by minimizing the following cost function:

$$\begin{aligned} \tilde{J}(\mathbf{v}_l, \mathbf{v}_r) = & \sum_{(x,y)\in\mathcal{D}} [I_l^t(x,y) - I_l^{t+1}(x+u_l, y+v_l)]^2 \\ & + \sum_{(x,y)\in\mathcal{D}} [I_r^t(x+d_t), y) - I_r^{t+1}(x+d_t+u_r, y+v_l)]^2 \\ & + \sum_{(x,y)\in\mathcal{D}} [I_l^{t+1}(x+u_l, y+v_l) - I_r^{t+1}(x+d_t+u_r, y+v_l)]^2 , \end{aligned}$$

where $\mathcal{D} \subset \mathbb{N}^2$ is the image support. This function consists of three data terms: the first two for left and right motion fields and the latter one for the joint estimation constraint. These expressions are nonconvex with respect to the displacement fields $\mathbf{v}_l$ and $\mathbf{v}_r$. Thus, to avoid a non-convex minimization, similarly to [1], we consider a Taylor expansion of the non-linear terms $I_l^{t+1}(x+u_l, y+v_l)$ and $I_r^{t+1}(x+d_t+u_r, y+v_l)$ around initial estimates $\bar{\mathbf{v}}_l = (\bar{u}_l, \bar{v}_l)$ and $\bar{\mathbf{v}}_r = (\bar{u}_r, \bar{v}_r)$, respectively, as follows:

$$\begin{aligned} I_l^{t+1}(x+u_l, y+v_l) \simeq & \; I_l^{t+1}(x+\bar{u}_l, y+\bar{v}_l) \\ & + (u_l - \bar{u}_l) I_l^{t+1,x} + (v_l - \bar{v}_l) I_l^{t+1,y} , \\ I_r^{t+1}(x+d_t+u_r, y+v_r) \simeq & \; I_r^{t+1}(x+d_t+\bar{u}_r, y+\bar{v}_r) \\ & + (u_r - \bar{u}_r) I_r^{t+1,x} + (v_r - \bar{v}_r) I_r^{t+1,y} , \end{aligned}$$

where $I_l^{t+1,x}$, $I_l^{t+1,y}$, $I_r^{t+1,x}$ and $I_r^{t+1,y}$ are, respectively, the horizontal and vertical gradient of the warped motion compensated left and right images. Our goal is to simultaneously recover the three components u_l, v_l and u_r, v_r being directly deduced from v_l using Eq. (2). Thus, by setting $w = (u_l, v_l, u_r)^\top$ and using the above linearizations, we end up with the following quadratic criterion to be minimized:

$$J_{\mathcal{D}}(w) = \sum_{i=1}^{3} \sum_{s\in\mathcal{D}} [L_i(s)\, w(s) - r_i(s)]^2 , \tag{3}$$

where
$$\begin{cases} L_1 = [I_l^{t+1,x},\ I_l^{t+1,y},\ 0] \\ L_2 = [0,\ I_r^{t+1,y},\ I_r^{t+1,x}] \\ L_3 = [I_l^{t+1,x},\ I_l^{t+1,y} - I_r^{t+1,y},\ I_r^{t+1,x}], \end{cases}$$

and
$$\begin{cases} r_1 = -I_l^{t+1} + \bar{u}_l I_l^{t+1,x} + \bar{v}_l I_l^{t+1,y} + I_l^t \\ r_2 = -I_r^{t+1} + \bar{u}_r I_r^{t+1,x} + \bar{v}_r I_r^{t+1,y} + I_r^t \\ r_3 = -I_l^{t+1} + I_r^{t+1} + \bar{u}_l I_l^{t+1,x} + \bar{v}_l I_l^{t+1,y} \\ \quad\;\; - \bar{u}_r I_r^{t+1,x} - \bar{v}_r I_r^{t+1,y}. \end{cases}$$

Optimizing the criterion (3), known as the *data fidelity* term in the inverse problem literature, aims at obtaining the best estimate of the vector parameters w knowing $\{L_{(i)}\}_i$ and $\{r_{(i)}\}_i$. However, the optimization problem that is solely based on the data fidelity objective function admits an infinite number of solutions due to the fact that three variables have to be determined for each pixel and that the components of $\{L_{(i)}\}_i$ may simultaneously vanish. This problem is therefore ill-posed, and, in order to get satisfactory solutions, it is necessary to consider additional constraints derived from prior knowledge. In this work, we seek to efficiently describe available constraints as closed convex sets in a Hilbert space $\mathcal{H}$ such as to formulate the problem, within a set theoretic framework, as follows:

$$\text{Find } u \in S = \bigcap_{i=1}^{m} S_i \text{ such that } J(u) = \inf J(S) \ , \tag{4}$$

where the objective $J : \mathcal{H} \to (-\infty, +\infty]$ is a convex function and the constraint sets $(S_i)_{1 \le i \le m}$ are closed convex sets of $\mathcal{H}$. Constraint sets can generally be modelled as level sets of continuous convex functions.

3.2. Convex constraints on motion vectors

The construction of convex constraints is derived here from the properties of the estimated fields. An example of possible prior knowledge is the range of motion values. Given a set of candidate motion vectors, we can impose minimal and maximal amplitudes on the amount of allowed horizontal and vertical motion, denoted respectively by $u_{\min}$, $u_{\max}$, $v_{\min}$ and $v_{\max}$. The constraint sets associated with this information are

$$S_1 \quad = \{w \in \mathcal{H} \mid u_{\min} \le u_l \le u_{\max}\} \ , \tag{5}$$

$$S_2 \quad = \{w \in \mathcal{H} \mid v_{\min} \le v_l \le v_{\max}\} \ , \tag{6}$$

$$S_3 \quad = \{w \in \mathcal{H} \mid u_{\min} \le u_r \le u_{\max}\} \ . \tag{7}$$

Furthermore, motion vectors should be smooth in homogeneous areas while keeping sharp edges [7]. The classical Tikhonov regularization [19], used in many ill-posed problems, tends to oversmooth discontinuities [1]. In this work, we circumvent the problem by using a total variation (tv) regularization constraint [17]. Basically, we introduce a bound on the integral of the norm of the spatial gradient whose effect is to smooth homogeneous regions in the motion field while preserving edges. Imposing an upper bound on the total variation allows to efficiently restrict the solution to the constraint sets

$$S_4 = \{w \in \mathcal{H} \mid \text{tv}(u_l) \le \tau_{u_l}\} \ , \tag{8}$$

$$S_5 = \{w \in \mathcal{H} \mid \text{tv}(v_l) \le \tau_{v_l}\} \ , \tag{9}$$

$$S_6 = \{w \in \mathcal{H} \mid \text{tv}(u_r) \le \tau_{u_r}\} \ , \tag{10}$$

where τ_{u_l}, τ_{v_l} and τ_{u_r} are positive constants that can be estimated from prior experiments and image databases.

The problem of motion estimation can finally be formulated as jointly finding the left and right motion fields which minimize the energy function (3) subject to the constraints $(S_i)_{1 \le i \le 6}$. Many powerful optimization algorithms have been proposed to solve this convex feasibility problem. For the current work, we employ the constrained quadratic minimization method developed in [18] and particularly well adapted to our needs. However, due to space limitation, we will not describe the algorithm but the reader is referred to [18, 9] for more details. By applying this algorithm, we obtain the two dense motion fields, and we can construct the initial disparity field for the second stereoscopic pair by using Eq. (2). The obtained disparity is sufficiently accurate to serve as a starting point for the convex programming approach we propose in [9] for dense disparity estimation.

4. EXPERIMENTAL RESULTS

We evaluated the proposed method on the real stereo image sequences "Outdoor" and "Aqua", for which the original left images of frames 44 and 1 are shown in Figures 2(a) and 2(b), respectively.

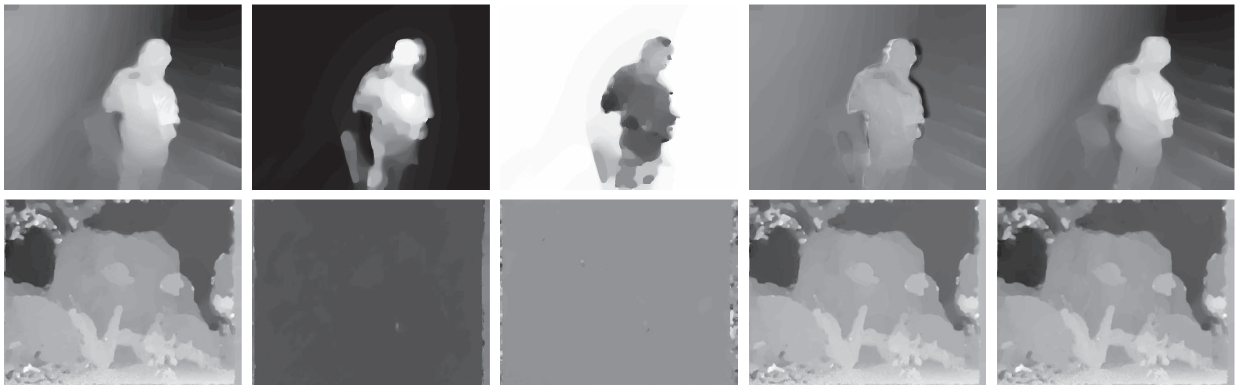
(a) (b)

Fig. 2. Left images for (a) "Outdoor" and (b) "Aqua" Stereo sequences.

The sequence "Outdoor[1]" shows first a static scene containing a background wall, a staircase and a uniform panel, then two persons enter the scene. In the sequence "Aqua", there is a global camera panning and a small horizontal fish motion. First of all, the disparity in the first frame is estimated using the method in [9] (see Figure 3(a)). The left and right motion vectors are then jointly estimated using the framework described in Section 3. A block based method is used to produce initial estimates for the dense motion fields, by looking for pixels in the search range with the maximum correlation between each block, of size 9×9 centered at the pixel of interest, in the previous frame and displaced blocks in the current frame. The horizontal and vertical motions of the left "Outdoor" and "Aqua" sequences are shown in Figures 3(b) and 3(c), respectively. We notice from these figures that our method allows to obtain consistent and smooth displacement vectors while preserving discontinuities around object boundaries. In the sequence "Outdoor", the unified motion of the background and the independent motion of the person are clearly distinguished. Moreover, the unified horizontal motion of both background and objects in the scene "Aqua" is also clearly perceived from computed motion vectors. Using the disparity in the previous frame, the left and right motion vectors and the constraint between motion and disparity, we compute the initial disparity of the current frame (see Figure 3(d)), which is refined later using the constrained quadratic minimization method proposed in [9]. The final estimated disparity field is shown in Figure 3(e). As expected, initial matching errors produced by the occlusions of motion are greatly reduced by using the refinement stage, which also guarantees that the obtained disparity field satisfies the imposed constraints, especially the disparity range constraint.

As the initial disparity is computed using accurate disparity and motion fields, the joint disparity estimation is better than that of the separate estimation where the initial disparity is obtained from a block-based correlation technique. Figure 4 shows the PSNR plots for the prediction of the current left images for frames 44 to 53 of the "Outdoor" sequence. The left images are predicted from right images through the current estimated disparity fields. The reconstruction errors obtained using the proposed joint estimation algorithm are compared with those obtained by applying a separate disparity estimation and a block-matching method. As can be seen from the curves in Figure 4, the joint estimation method performs well and better than the direct estimation and the block-matching disparity compensation. An additional benefit from the joint estimation model is the reduction of computational load by about 30 to 40 percent, since we have reduced the number of displacement vectors to be estimated and saved the time-consuming initial disparity computation. Notice that our current implementation was completely written in Matlab code and so more efficient implementations can be written in C. In addition, on a parallel architecture, we can exploit the parallel

[1]http://www.vision.deis.unibo.it/smatt/stereo.htm



(a) (b) (c) (d) (e)

Fig. 3. Disparity and motion fields in "Outdoor" (top) and "Aqua" (bottom) stereo sequences: (a) disparity in the previous frame (b) horizontal left motion (c) vertical left motion, (d) initial disparity and (e) final constrained disparity in the current frame.

structure of the algorithm, where subgradient projections on the different constraint sets may be computed in parallel, to further reduce the computational time.

5. CONCLUSION

In this paper, a new method for the joint estimation of motion and disparity in stereo image sequences was investigated. At first, a robust and efficient optimization algorithm was developed to simultaneously estimate accurate left and right motion vectors. Within a convex set theoretic framework, this algorithm minimizes a quadratic convex objective function subject to some appropriate convex constraints. Secondly, the disparity field at the current frame was estimated using the consistency constraint, the left and right motion vectors and the disparity field obtained at the previous frame. The proposed method has given promising performance results while reducing the computational cost.

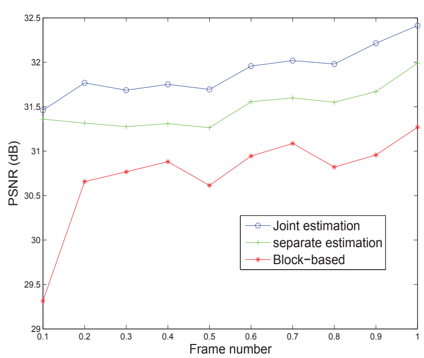


Fig. 4. PSNR plot of the predicted current left images of "Outdoor" sequence using the joint estimation model, the separate constrained estimation and the block-based correlation.

6. REFERENCES

[1] K.P. Horn and G. Schunck, "Determining optical flow," *Artificial Intelligence,* no. 7, pp. 185-203, 1981.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, pp. 7–42, Apr. 2002.

[3] O. Veksler, "Stereo correspondence by dynamic programming on a tree," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, San Diego, USA, Jun. 20-25, 2005, vol. 2, pp. 384–390.

[4] M. Mozerov, V. Kober and T.S. Choi, "Motion estimation with a dynamic programming optimization operator, in *Proc. Int. Conf. Image Process.*, Rochester, USA, Sep. 22-25, 2002, vol. 2, pp. 269–272.

[5] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. Int. Conf. Comput. Vis.,* Vancouver, BC, Canada, Jul. 9-12, 2001, vol. 2, pp. 508–515.

[6] G. Boccignone, A. Marcelli, P. Napoletano and M. Ferraro, "Motion Estimation via Belief Propagation," in *Proc. Int. Conf. Image Analysis and Process.,* Modena, Italy, Sept. 10-14, 2007, pp. 55-60.

[7] G. Aubert, R. and Deriche and P. Kornprobst, "Computing Optical Flow via Variational Techniques," *SIAM Journal on Numerical Analysis,* vol. 60, no. 1, pp. 156-182, Dec. 1999.

[8] N. Slesareva, A. Bruhn and J. Weickert, "Optic flow goes stereo: a variational method for estimating discontinuity-preserving dense disparity maps," in *27th DAGM Symposium*, Vienna, Austria, Aug. 31 - Sep. 2, 2005, vol. 3663, pp. 33–40.

[9] W. Miled, J. C. Pesquet and M. Parent, "Disparity map estimation using a total variation bound," in *Proc. 3rd Canadian Conf. Comput. Robot Vis.,* Quebec, Canada, Jun. 7-9, 2006, pp. 48–55.

[10] A. Tamtaoui and C. Labit, "Coherent disparity and motion compensation in 3DTV image sequencecoding schemes," in *Proc. Int. Conf. Acoustics, Speech, and Signal Process.,* Toronto, Canada, Apr. 14-17, 1991, vol. 4, pp. 2845–2848.

[11] J. Liu and R. Skerjanc, "Stereo and motion correspondence in a sequence of stereo images," *Signal Processing: Image Communication,* vol. 5, pp. 305–318, 1993.

[12] W. Yang, K. Ngan, J. Lim and K. Sohn, "Joint Motion and Disparity Fields Estimation for Stereoscopic Video Sequences," *Signal Processing: Image Communication,* vol. 20, no. 3, pp. 265-276, Mar. 2005.

[13] H. Weiler, A. Mitiche and A. Mansouri, "Boundary preserving joint estimation of optical flow and disparity in a sequence of stereoscopic images, " *Int. Conf. on Visualisation, Imaging, and Image Process.,* Malaga, Spain, pp. 102-106, Sep. 2003.

[14] I. Patras, N. Alvertos and G. Tziritas, "Joint disparity and motion field estimation in stereoscopic image sequences," in *Proc. Int. Conf. Pattern Recognition,* Vienna, Austria, Aug. 25-29, 1996, vol. 1, pp. 359-363.

[15] D. Min, H. Kim and K. Sohn, "Preserving joint motion-disparity estimation in stereo image sequences," *Signal Processing: Image Communication,* vol. 21, no. 3, pp. 252-271, Mar. 2006.

[16] A. Fusiello, E. Trucco and A. Verri, "A compact algorithm for rectification of stereo pairs, " *Machine Vis. Appl.*, vol. 12, no. 1, pp. 16–22, 2000.

[17] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259–268, 1992.

[18] P. L. Combettes, "A block iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Process.*, vol. 51, pp. 1771–1782, Jul. 2003.

[19] A. N. Tikhonov and A. Y. Arsenin, "Solution of ill-posed problems," *John Wiley and Sons*, Washington D.C., 1977.