Joint disparity and motion field estimation in stereoscopic image sequences

I. Patras, N. Alvertos and G. Tziritas
Institute of Computer Science
Foundation for Research and Technology-Hellas, and,
Department of Computer Science, University of Crete
P.O. Box 1470, Heraklion, Greece
E-mail: patras@csd.uch.gr, alvertos@csd.uch.gr, tziritas@csi.forth.gr

Abstract

This work aims at determining, given two stereoscopic image sequences, at any time instant two dense velocity fields, for the left and the right sequence, and the disparity field. The disparity field of the previous stereoscopic pair is considered as known. Thus at the initial time instant the disparity field of the first stereoscopic pair is estimated. For both problems multiscale iterative relaxation algorithms are used. Results are given with real stereoscopic data.

1. Introduction

There are three general approaches regarding dynamic stereo vision which are found in the existing bibliography [11]. The first one consists of initially solving the stereoscopic problem, which leads to the static determination of objects, followed by the correspondence of these objects in time. Point correspondences are considered [5], [8], [7], as well as boundary segment correspondences [4], [13]. The second general approach evaluates independently the two 2-D velocity fields in the sequence of stereoscopic image pairs, and then determines and uses the stereoscopic relations which exist between the two velocity fields [6], [12]. Finally, the third approach uses a joint estimation of the two 2-D velocity fields taking advantage of their stereoscopic relation without seeking the complete 3-D reconstruction of the depicted objects [10].

As it was shown through the previous bibliographic reference, the existing solutions do not utilize the stereoscopic and motion relations simultaneously; instead, they consider the problem in sequential stages. It is known, however, that positions at each time instant are connected with displacements and, furthermore, the relations which connect the rate of change of the stereoscopic disparities with the velocity fields are known. This work aims at an integrated solution to the problem of dynamic stereoscopic vision. A simultaneous estimation of the velocity and disparity fields in a dense structure, as opposed to most of the existing methods, where only sparse image descriptions are given, is proposed. At any time instant two motion fields (for the left and right image sequences) and one disparity field are computed. The disparity field of the previous stereoscopic pair is considered as known, that is previously estimated. A cost function, which contains known equations regarding velocity and disparity fields in relation to image intensity, and which also constraints the different fields to be adaptatively smooth, is constructed. Minimization of the cost function results into estimation of the velocity and disparity fields. This minimization can be achieved using an iterative relaxation algorithm based on the gradient of the cost function.

The analysis is based on a converging (fixating) stereoscopic optical system, with an angle θ between the two optical axes, and with B the distance between the two focal points and f the same focal length for each camera. For a 3-D point, whose perspective projections in the right and left image respectively are (x_r, y_r) and (x_l, y_l), the disparity vector $\vec{d}$ is defined as

$$\vec{d} = (x_r - x_l, y_r - y_l)$$

Assuming that θ is very small, then $\frac{y_l}{y_r} \approx 1$; that is, the y-coordinate of $\vec{d}$ is almost zero. For the remainder of this work it is accepted that all previous assumptions hold true, thus $\vec{d}$ is a 1-D vector along the x-axis. The depth is then related to the two corresponding coordinates by

$$Z = -B \frac{\tan(\phi - \alpha)\tan(\phi + \beta)}{\tan(\phi - \alpha) + \tan(\phi + \beta)} \qquad (1)$$

where $\phi = \pi/2 - \theta$, $\alpha = \arctan(x_l/f)$ and $\beta = \arctan(x_r/f)$. If $\theta = 0$, we obtain the lateral model, in which case $\vec{d}$ is exactly 1-D.

In Section 2 we present a regularization method for obtaining a smooth disparity field from a stereoscopic pair

using diffusion adaptive functions, Also a method for detecting occlusions is proposed. In Section 3 the simultaneous estimation of the two motion fields and the current disparity field from two successive stereoscopic pairs is presented. Section 4 contains experimental results with real stereoscopic image sequences. Then some conclusions are given, as well as directions of future work.

2. Disparity Field Estimation

Solution to the stereoscopic problem consists of determining a dense disparity field δ through which every point (x_l, y_l) in the left image is matched to a point $(x_l + \delta, y_l)$ in the right image. Using the optical-flow preservation principle, it is also true that $I_l(x_l, y_l) = I_r(x_r, y_r)$. However, since intensity measurements are not exact and all hypotheses are not absolute, the following cost function, including a smoothness constraint, is minimized [3]

$$\sum_{(i,j)} (I_r(i+\delta, j) - I_l(i,j))^2 + \lambda \sum_{(i,j)} \sum_{p \in \mathcal{N}_{ij}} g(\delta - \delta_p)$$

where $\mathcal{N}_{ij} = \{(i-1,j),(i+1,j),(i,j-1),(i,j+1)\}$ is the 4-point neighborhood of (i,j). The dependence of δ on (i,j) is omitted for simplifying the notation. $g(.)$ is a diffusion adaptive function [9], which, if it is carefully chosen $g(.)$, it may regularize the solution and at the same time preserve the discontinuites. In that framework the *interaction function* $h(.)$ which is defined such that : $g'(x) = xh(x)$ determines the interaction between neighboring pixels. In this work $g(.)$ and $h(.)$ were chosen to be : $g(x) = \gamma|x| - \gamma^2 \ln(1 + \frac{|x|}{\gamma})$ and $h(x) = \frac{1}{1+\frac{|x|}{\gamma}}$. λ is a weight coefficient which determines to what degree estimation of the field is influenced by the smoothing operator. Minimization of this quantity results into the following set of equations:

$$(I_r(i+\delta, j) - I_l(i,j))\, I_{rx}(i+\delta, j) + \lambda\alpha(\delta - \bar{\delta}) = 0 \quad (2)$$

where

$$\bar{\delta} = \frac{1}{\alpha} \sum_{p \in \mathcal{N}_{ij}} h(\delta - \delta_p)\delta_p \quad \text{and} \quad \alpha = \sum_{p \in \mathcal{N}_{ij}} h(\delta - \delta_p)$$

Assuming that the magnitude of the field is relatively small and image intensity varies smoothly, the following relations hold true:

$$\begin{aligned} I_{rx}(i+\delta, j) &= I_{rx}(i+\bar{\delta}, j) \\ I_r(i+\delta, j) &= I_r(i+\bar{\delta}, j) + (\delta - \bar{\delta}) I_{rx}(i+\bar{\delta}, j) \end{aligned}$$

Considering the above, Eq. (2) becomes:

$$(\overline{\Delta I_{rl}} + (\delta - \bar{\delta}) I_{rx}(i+\bar{\delta}, j)) I_{rx}(i+\bar{\delta}, j) + \lambda\alpha(\delta - \bar{\delta}) = 0 \quad (3)$$

where $\overline{\Delta I_{rl}} = I_r(i+\bar{\delta}, j) - I_l(i,j)$. The solution at the k^{th} iteration is given by the relation:

$$\delta^k = \bar{\delta}^{k-1} - \frac{\overline{\Delta I_{rl}}\, I_{rx}}{\lambda\alpha + (I_{rx})^2} \quad (4)$$

where δ^k is the disparity field estimated at the k iteration. The algorithm is terminated, when the percentage of diminishment of the average correction $\mathrm{E}\{|\delta^k - \delta^{k-1}|\}$ becomes less than a threshold.

The previously described algorithm, as a gradient-descent algorithm, can estimate successfully only fields of small disparities. Otherwise, it requires good initial conditions, so that it will not be entrapped and converge to a local minimum. Thus, it is insufficient for real data, where large disparity values are possible and no prior general knowledge of the scene depth is available.

Consequently, a coarse-to-fine multiscale method in a pyramidal form is implemented, where in the upper levels the algorithm is applied to images of submultiple dimensions of the original ones [2]. Those images are the result of reduction by low-pass filtering and subsampling. An immediate result of this reduction is the scale change on the magnitude of the field to be estimated. The algorithm is applied at the various levels of the pyramid, from the top to bottom, and the disparity field which is estimated at the coarser level l constitutes the initial estimation at the subsequent finer level $l-1$. In this way what we actually have to estimate at level $l-1$ is the difference $\delta - \hat{\delta}^l$ between the real disparity field and the coarse estimation that we obtained at the previous level.

The value of parameter γ is also changed at the various levels of the pyramid. At coarser levels where there is lack of detail, due to the low-pass filtering and subsampling, larger values of γ impose a "harder" smoothing, while at finer levels of detail the discontinuites are more carefully preserved.

Via the estimated dense disparity field, a matching scheme between the points of the stereoscopic pair of images is obtained. However, some of these matches might be incorrect, either due to stereo occlusions, or due to errors in the disparity estimation process. In a post-processing step we try to detect these areas using error confidence measures.

The objective is the construction of a binary map Φ_{ij} which denotes, if the match between point (i,j) at the left image and $(i+\delta, j)$ at the right image is correct. The error confidence measure used is the mean square displaced frame difference (DFD) between the 3×3 blocks centered at (i,j) and $(i+\delta, j)$. Firstly, conflicts between matches are removed, that is the situation where several points are matched with the same point at the right image. A situation like that arises at left occluded areas, that is areas "seen" only by the left camera. Only the match with the smallest confidence error measure is considered as correct ($\Phi_{ij} = 1$).

The other matches are declared false ($\Phi_{ij} = 0$). Next, false matches are declared at every point at the left image that - via the disparity vector assigned to it - corresponds to a point outside the right image. Finally, false matches are declared at every point at the left image that the error confidence measure assigned to it is above a certain threshold.

3. Simultaneous Motion and Disparity Estimation

The second stage of this work consists of a simultaneous estimation of the two velocity fields and the disparity field of the second stereoscopic image pair. The aim is to determine for a point in the left frame at t a displacement vector (u_l, v_l) giving its corresponding point in the left frame at $t+1$, and for a point in the right frame at t a displacement vector (u_r, v_r) giving its corresponding point in the right frame at $t+1$. To estimate the motion and the second disparity fields, the correspondence between points in the first stereoscopic image pair is used, as derived in the first stage by evaluating the field δ_t.

The following relations among the components of the fields to be estimated hold [12], when there is a correct match between points (i, j) and (i', j')

$$v_r(i', j') = v_l(i, j) \quad \text{and}$$

$$\delta_{t+1}(i + u_l, j + v_l) = u_r(i', j') - u_l(i, j) + \delta_t(i, j) \quad (5)$$

Therefore, at the points where the match is correct to completely determine the requested fields it is sufficient to evaluate their three components u_r, u_l and v_l. For these points Eq. (5) implies that we can implicitly construct the dense disparity field δ_{t+1}.

As with the solution to the stereoscopic problem, the minimization of the squared deviation from the image intensity preservation principle and a smoothness constraint for the estimated fields are considered. Let us note
$\Delta I_l = I_l(i, j, t) - I_l(i + u_l, j + v_l, t + 1)$
$\Delta I_r = I_r(i', j', t) - I_r(i' + u_r, j' + v_l, t + 1)$
$\Delta I_{rl} = I_r(i' + u_r, j' + v_l, t + 1) - I_l(i + u_l, j + v_l, t + 1)$
The total quantity to be minimized is, for points correctly matched,

$$\sum_{(i,j)}((\Delta I_l)^2 + (\Delta I_r)^2 + (\Delta I_{rl})^2 + \lambda \sum_{p \in \mathcal{N}_{(i,j)}} g(u_r - u_r^p)) + \lambda \sum_{(i,j)} \sum_{p \in \mathcal{N}_{(i,j)}} ((g(u_l - u_l^p) + g(v_l - v_l^p)) \quad (6)$$

The first term is the mean square DFD between the two left images, the second term the mean square DFD between the two right images. With the third term we try to minimize the mean square DFD for the second stereoscopic pair. The last terms refer to the smoothing of the velocity fields.

Let us define an interpolation operation on the u_l field using the interaction function $h(.)$ as follows

$$\bar{u}_l = \frac{\sum_{p \in \mathcal{N}_{ij}} h(u_l - u_l^p) u_l^p}{\sum_{p \in N_{ij}} h(u_l - u_l^p)}$$

and in the same way on v_l and u_r.

Assuming that the fields to be estimated are small in magnitude and that the intensities vary smoothly, the following approximations is used, at time instant $t+1$

$$I_l(i + u_l, j + v_l) \approx I_l(i + \bar{u}_l, j + \bar{v}_l) + (u - \bar{u}_l) I_{lx}(i + \bar{u}_l, j + \bar{v}_l) + (v - \bar{v}_l) I_{ly}(i + \bar{u}_l, j + \bar{v}_l)$$

and the same for $I_r(i + u_r, j + v_r)$.

Let us siplify the notation of the above derivatives by omitting to explicitly indicate the point location. For example,

$$I_{rx} = I_{rx}(i' + \bar{u}_l, j' + \bar{v}_l, t + 1)$$

Let us also note $\overline{\Delta I}_l$ the value of ΔI_l given above if $(u_l, v_l) = (\bar{u}_l, \bar{v}_l)$, and in the same way $\overline{\Delta I}_r$ and $\overline{\Delta I}_{rl}$. Finally, we note $\alpha_l^u = \sum_{p \in \mathcal{N}_{ij}} h(u_l - u_l^p)$, and in the same way α_l^v and α_r^u.

Finally, for the points where the match is correct, we obtain after some simplification the following solution

$$\begin{bmatrix} u_l^k \\ u_r^k \\ v_l^k \end{bmatrix} = \begin{bmatrix} \bar{u}_l^{k-1} \\ \bar{u}_r^{k-1} \\ \bar{v}_l^{k-1} \end{bmatrix} + \begin{bmatrix} \frac{I_{lx}(\overline{\Delta I}_l - \overline{\Delta I}_{rl})}{2I_{lx}^2 + \lambda \alpha_l^u} \\ \frac{I_{rx}(\overline{\Delta I}_r + \overline{\Delta I}_{rl})}{2I_{rx}^2 + \lambda \alpha_r^u} \\ \frac{I_{ly}(\overline{\Delta I}_l - \overline{\Delta I}_{rl}) + I_{ry}(\overline{\Delta I}_r + \overline{\Delta I}_{rl})}{I_{ly}^2 + I_{ry}^2 + (I_{ly} - I_{ry})^2 + \lambda \alpha_l^v} \end{bmatrix} \quad (7)$$

For the points where the match is false, the resulting equations are similar to that obtained in monocular motion analysis [3]. The solution for the left motion field is :

$$\begin{bmatrix} u_l \\ v_l \end{bmatrix} = \begin{bmatrix} \bar{u}_l \\ \bar{v}_l \end{bmatrix} + \frac{\overline{\Delta I}_l}{\lambda \alpha_l^u \alpha_l^v + \alpha_l^v I_{lx}^2 + \alpha_l^u I_{ly}^2} \begin{bmatrix} \alpha_l^v I_{lx} \\ \alpha_l^u I_{ly} \end{bmatrix} \quad (8)$$

and a similar solution is obtained for the right motion field.

Once the motion fields are estimated, then a disparity field for the second stereoscopic pair of images can be partially constructed using Eq. (5). For each point (i, j) in the left image at time t for which $\Phi_{ij} = 1$, we find the corresponding point $(i + u^l, j + v^l)$ at the left image at time $t+1$ and we assign to it the disparity value that Eq. (5) implies.

This process leaves a number of points with no disparity vector assigned to them. These points are, either points that the motion has deccluded, or points whose correspondences at left image at time t are stereo occluded or false matches.

It should be noted that we should not declare these points stereo occluded or false matches without trying to find their correspondences at the right image at time $t+1$. Even areas that were stereo occluded at time t, might, at time $t+1$, (due to camera motion for example) have a correspondence at the right image.

For these points we apply the algorithm described at Section 2 with a large value parameter γ in order to get a coarse estimation of their disparity values. The resulting complete dense disparity field δ_{t+1}, is further improved by applying the algorithm described at Section 2 for all the points of the second stereoscopic pair. The algorithm is applied at the lowest level of the multiscale approach (*i.e.* finest detail) for a small number of iterations and with a small value of γ (to preserve the discontinuites).

4. Experimental Results

The algorithm was tested in a number of real stereoscopic sequences, namely the 'tunnel' and the 'train'. The left images of these sequences are depicted in Fig. 1.

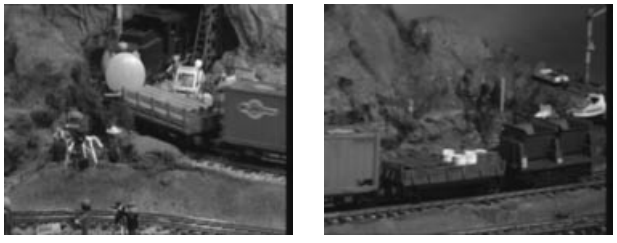

Figure 1. Left images at first time instant

Qualitatively the motion of scene objects consists of :

tunnel A train motion from right to left. The camera remains still.

train A train motion from left to right, and a camera motion from right to left.

4.1. Stereoscopic Problem Solutions

The results for this subsection were obtained with the method described in Section 2. The aim was the construction of dense disparity field to be used to apply the algorithm described in Section 3 on the rest of the stereoscopic sequences. A dense disparity field for each of these sequences appear in Fig. 2. Both were obtained for $\lambda = 500$ and γ initial value 500. The value of γ was reduced, as we descended the pyramid in the multiscale approach by a factor of 2 at each level. The number of levels in the multiscale approach was 4, so at the lowest level we had $\gamma = 31.25$. Furthermore at that last level after a small number of iterations (*i.e.* 20) we set $\gamma = 0.1$. The final DFDs were 65.3 and 22.1 for 'tunnel' and 'train' respectively.

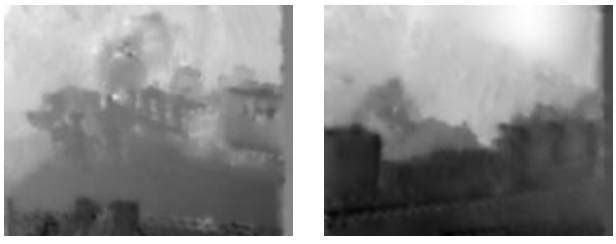

Figure 2. Estimated depth maps at first time instant

4.2. Stereoscopy and Motion

Implementation of the second stage with real data consists of two stereoscopic sequences. The result of the first stage (*i.e.*, correspondence between image points in the first pair) is assumed given and at any time instant the estimated disparity field δ_t was used for the joint motion and disparity estimation between frames at the next time instant. Results from this phase are summarized in Table 1, where LL refers to the mean square DFD between the two left images, RR to the mean square DFD between the two right images and RL to the mean square DFD between the images of the stereoscopic pair. For the evaluation of the last quantity only points where Eq. (5) applies are taken into account.

Sequence	Frame	LL	RR	RL
tunnel	27	5.51	5.36	29.36
	28	4.28	4.21	22.61
	29	3.94	3.83	19.04
	30	3.44	3.19	19.24
train	1	3.93	3.26	14.68
	2	4.80	3.85	13.04
	3	6.18	5.80	14.31
	4	6.69	7.85	15.82
	5	8.79	10.86	16.61
	6	8.02	8.16	15.09

Table 1. Stereoscopy and motion on real data

The left velocity fields for the two stereoscopic sequences are presented in Fig. 3 and 4.

In Fig. 5 the last disparity field at each sequence is depicted. We should note that large areas have just been deccluded due to the motion, however our method gets good estimations for them as well.

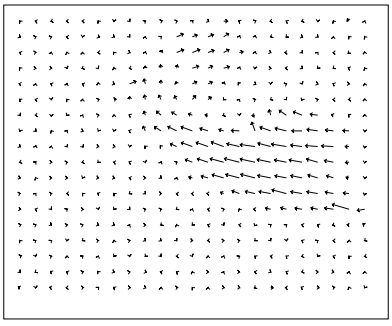

Figure 3. Estimated motion field for *tunnel* sequence

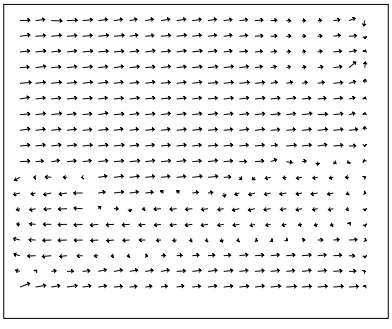

Figure 4. Estimated motion field for *train* sequence

5. Conclusions

An integrated approach to the problem of dynamic stereoscopic vision was proposed, where velocity and disparity fields in a dense structure are estimated simultaneously. In both stages of the scheme, where initially the dense disparity field of the stereoscopic pair is evaluated followed by estimation of the two dense velocity and disparity fields of the second stereoscopic pair, convergence is achieved using a multiscale iterative relaxation algorithm. Experimental results were presented for real data. Specifically, the approach was applied to real image sequences, where both the object and the binocular system are moving. Since theoretical and experimental approaches to this problem assume only a simple stereoscopic model (*i.e.*, converging or lateral), it would be useful to examine other models such as the axial [1], the telescopic, or a general one, where the two optical systems are related with non-zero rotations, so that it is possible to confront the more general case, where the geometry of the stereoscopic model is varying with time. This optical-system dynamic behavior finds application in robotics, where autonomous mechanisms are desired.

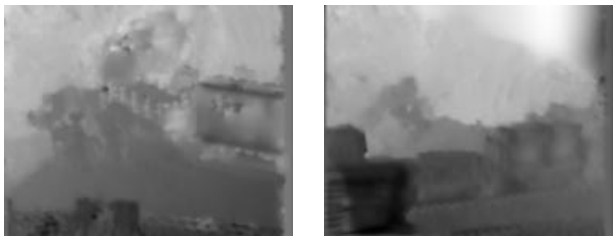

Figure 5. Estimated depth maps at last time instant

References

[1] N. Alvertos, D. Brazakovic, and R. C. Gonzalez. Camera geometries for image matching in 3-d machine vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-11(9):pp. 897–915, Sept 1989.

[2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int. J. of Computer vision*, 2:pp. 283–310, 1989.

[3] B. Horn. *Robot vision*. MIT Press, 1986.

[4] Y. C. Kim and J. K. Aggarwal. Determining object motion in a sequence of stereo images. *IEEE J. of Robotics and Automation*, 3(6):pp. 599–614, Dec 1987.

[5] M. K. Leung and T. S. Huang. An integrated approach to 3-d motion analysis and object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(10):pp. 1075–1084, Oct. 1991.

[6] A. Mitiche. On kineopsis and computation of structure and motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(1):pp. 109–112, Jan. 1986.

[7] A. Mitiche and P. Bouthemy. Tracking modelled objects using binocular images. *Computer Vision, Graphics and Image Processing*, 32:pp. 384–396, 1985.

[8] A. N. Netravali and *et al*. Algebraic methods in 3-d motion estimation from two-view point correspondences. *Int. J. of Imaging Systems and Technology*, 1:pp. 78–99, 1989.

[9] S.Z.Li. On discontinuity-adaptive smoothness priors in computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-17(6):pp. 576–586, June 1995.

[10] A. Tamtaoui and C. Labit. Constrained disparity and motion estimators for 3-dtv image sequence coding. *Signal Processing: Image Communication*, 4:pp. 45–54, 1991.

[11] G. Tziritas and C. Labit. *Motion analysis for image sequence coding*. Elsevier, 1994.

[12] A. M. Waxman and J. H. Duncan. Binocular image flows: steps forward stereo-motion fusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(6):pp. 715–729, Nov 1986.

[13] Z. Zhang and O. Faugeras. Estimation of displacements from two 3-d frames obtained from stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-14(12):pp. 1141–1156, Dec. 1992.