FORTH-ICS / TR-157 December 1995

Joint disparity and motion field estimation in stereoscopic image sequences

Ioannis Patras, Nikos Alvertos and Georgios Tziritas[†]

Abstract

This work aims at determining four dense fields given two stereoscopic pairs of images in consecutive time instances: two dense velocity fields and two dense disparity fields. The proposed scheme is completed in two stages. During the first stage, the dense disparity field of the first stereoscopic pair is estimated using a multiscale iterative relaxation algorithm. During the second stage, again using a similar iterative algorithm, the two dense velocity fields and the disparity field of the second stereoscopic pair are estimated. The method has been implemented with both synthetic and real data.

[†]Institute of Computer Science, F.O.R.T.H., and, Computer Science Department, University of Crete, Greece,

E-mail: patras@csd.uch.gr, alvertos@csd.uch.gr, tziritas@csi.forth.gr

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. Leung and Huang [5] assume a scene which contains only one moving object with respect to the stationary receiver (camera), where stereoscopic correspondences are first determined for a sufficient number of points followed by evaluation of their three-dimensional motion. A similar approach to the problem is employed by Netravali *et al.* [9] and Mitiche and Bouthemy [8]. Kim and Aggarwal [4] use segments of boundaries, instead of points, for which initially the stereoscopic problem is solved. Motion is determined under the hypothesis that the depicted objects are rigid and, thus, their geometric characteristics remain unchanged in time. Similar is also the approach by Zhang and Faugeras [13] which is based on straight line segments with complete determination and appropriate description in the three-dimensional space. The second general approach evaluates independently the two two-dimensional velocity fields in the sequence of stereoscopic image pairs and then determines and uses the stereoscopic relations which exist between the two velocity fields. Mitiche [7] determines relations among positions and velocities of discrete points which correspond to the stereoscopic pairs. Solution to these equations results into the three-dimensional motion of a set of points which are assumed to belong to the same rigid body. Waxman and Duncan [12] determine relations between the 2-D velocity fields of the stereoscopic pair and define the binocular differential motion which is connected to the stereoscopic disparity. 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. In this case, where a complete 3-D motion and shape object description is not required, Tamtaoui and Labit [10] introduce stereoscopic relations between the velocity fields of the stereoscopic pair which utilize for the simultaneous evaluation of the two velocity fields using a recursive method. This approach seems to produce interesting results for applications in 3-D television.

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, that is, for all image points, 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. In addition, a method which has been shown to be effective in monocular motion analysis is used, given that it has been appropriately adapted to the requirements of the problem. Initially, a cost function is determined which constraints the different fields to be adaptatively smooth. This cost function also contains known equations regarding velocity and disparity fields in relation to image intensity, as well as relations between velocities and disparities depending on the geometrical model of the optical system. 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.

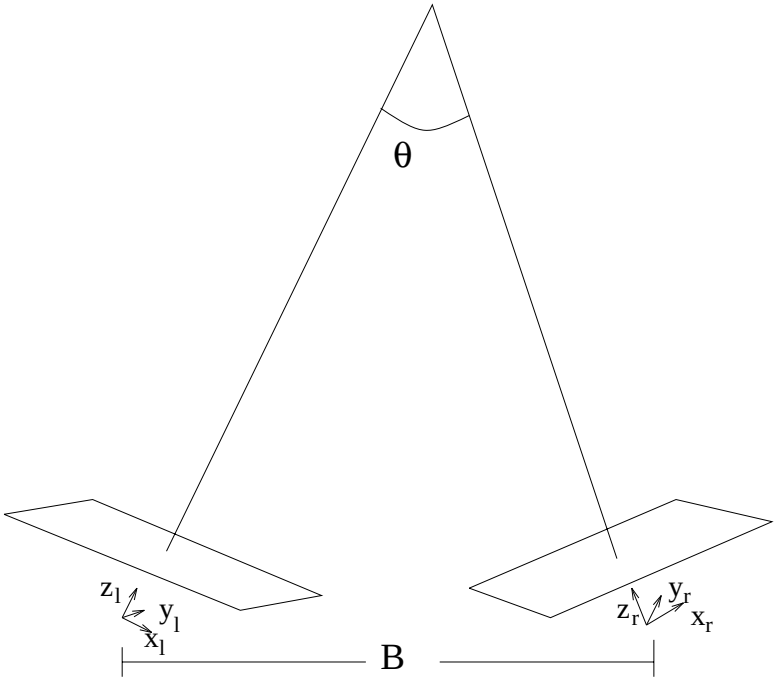


Figure 1: Converging Model

The analysis is based on a converging (fixating) stereoscopic optical system,

shown in Figure 1, where θ is the angle between the two optical axes, B is the distance between the two focal points and f is the same focal length for each camera. For a 3-D point $\mathbf{r}$ with left camera coordinates $r_l = (\ X_l\ ,\ Y_l\ ,\ Z_l\ )$ and right camera coordinates $r_r = (\ X_r\ ,\ Y_r\ ,\ Z_r\ )$ the following equality holds true:

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \\ Z_l \end{bmatrix} + \begin{bmatrix} -B\cos(\theta/2) \\ 0 \\ B\sin(\theta/2) \end{bmatrix} \tag{1}$$

If (x_r, y_r) and (x_l, y_l) are the perspective projections of $\mathbf{r}$ in the right and left image, respectively, the following relations among the coordinates of the same point $\mathbf{r}$ and its projections in the stereoscopic pair are obtained:

$$x_r \frac{Z_r}{f} = x_l \frac{Z_l}{f}\cos\theta + Z_l \sin\theta - B\cos\theta/2 \tag{2}$$

$$y_r \frac{Z_r}{f} = y_l \frac{Z_l}{f} \tag{3}$$

$$Z_r = -x_l \frac{Z_l}{f}\sin\theta + Z_l\cos\theta + B\sin\theta/2 \tag{4}$$

Examining the ratio $\frac{y_l}{y_r}$ it is shown:

$$\frac{y_l}{y_r} = \frac{Z_r}{Z_l} = \frac{\frac{x_l}{f}\sin\theta/2 + \cos\theta/2}{\frac{x_r}{f}\sin\theta/2 + \cos\theta/2}$$

Defining the disparity vector $\vec{d}$ as

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

and 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.

Eqs. (2 - 4) form the basis for 3-D scene reconstruction following the establishment of image-point correspondence in the stereoscopic pair. For every image-point pair $(x_r, y_r), (x_l, y_l)$, the solution of those equations provides a depth estimate (Z_l and Z_r) of the projected 3-D point. Depth can also be calculated using an intermediate coordinate system (the **cyclopean system**) which is obtained by rotating $<\ X_l\ ,\ Y_l\ ,\ Z_l\ >$ about the y axis with an angle $\theta/2$ and then translating $\frac{B}{2}\sin(\theta/4)$ along the x-axis. From the corresponding equations, a closed form solution for Z (depth) is determined [2]:

$$Z = -B\,\frac{\tan(\phi-\alpha)\tan(\phi+\beta)}{\tan(\phi-\alpha)+\tan(\phi+\beta)} \tag{5}$$

where $\phi = \pi/2 - \theta$, $\alpha = \arctan(x_l/f)$ and $\beta = \arctan(x_r/f)$.

If the angle $\theta = 0$, we obtain the lateral model. In this case, the equations relating image and 3-D point coordinates are:

$$x_r = x_l - B\frac{f}{Z} \quad (6)$$

$$y_r = y_l \quad (7)$$

where $Z = Z_r = Z_l$. It is also noted that the disparity vector $\vec{d}$ is 1-D along the x-axis.

In Section 2 we present a regularization method for obtaining a smooth disparity field from a stereoscopic pair using diffusion adaptive functions. 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 both synthetic and 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 $(x_l + \delta, y_l) = (x_r, y_r)$ 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 functional is minimized [3]:

$$e_i = \int\int (I^l(x,y) - I^r(x+\delta, y))^2 dxdy \quad (8)$$

In addition, it is required that Z, and thus δ, is varying smoothly. Consequently, a smooth-field functional is introduced for minimization:

$$e_s = \int\int (g(\delta_x) + g(\delta_y))dxdy \quad (9)$$

where $\delta_x = \frac{\partial \delta}{\partial x}$ and $\delta_y = \frac{\partial \delta}{\partial y}$ and $g(.)$ is a function which belongs to the **DAF** (Diffusion Adaptive Functions) as described in [6]. $g(.)$ as a member of the **DAF** family, is provided as an alternative to the quadratic regularizer $(.)^2$ which has the disadvantage that it imposes the smoothness constraint everywhere and leads to oversmoothing. On the other hand a carefully chosen $g_c(.)$ 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 $h(x) = \frac{1}{1+\frac{|x|}{\gamma}}$. For the discrete case the total quantity to be minimized is

given by:

$$\sum_i \sum_j (I^r(i+\delta_{i,j},j) - I^l(i,j))^2 + \lambda \sum_i \sum_j \sum_{p' \in N_{ij}} g(\delta_{ij} - \delta_{p'})$$

where $N_{ij} = \{(i-1,j),(i+1,j),(i,j-1),(i,j+1)\}$ is the 4-point neighborhood of (i,j). The sum $\sum_{p' \in N_{ij}} g(\delta_{ij} - \delta_{p'})$ is an approximation for e_s and can be extended to a neighborhood of more than 4 points (e.g., 3×3 where the diagonal terms are normalized by $\frac{\sqrt{2}}{2}$). 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 equation:

$$(I^r(i+\delta_{i,j},j) - I^l(i,j))\, \mathrm{I}^r_x(i+\delta_{i,j},j) + \lambda \sum_{p' \in N_{ij}} l_{p'}(\delta_{ij} - \bar{\delta}_{ij}) = 0 \qquad (10)$$

where $l_{p'} = h(\delta_{ij} - \delta_{p'})$ are the coefficients which determine the contribution of each of the neighborhood points in estimating δ_{ij}, and

$$\bar{\delta}_{ij} = \frac{\sum_{p' \in N_{ij}} l_{p'} \delta_{p'}}{\sum_{p' \in N_{ij}} l_{p'}}$$

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

$$\begin{aligned} I^r_x(x+\delta,y) &= I^r_x(x+\bar{\delta},y) \\ I^r(x_r,y_r) &= I^r(x_l+\delta,y_l) = I^r(x_l+\bar{\delta},y_l) + (\delta-\bar{\delta}) I^r_x(x_l+\bar{\delta},y_r) \end{aligned}$$

Considering the above, Eq. (10) becomes:

$$(\overline{\Delta I^{rl}} + (\delta-\bar{\delta}) I^r_x(x_l+\bar{\delta},y_r))\, I^r_x(x_l+\bar{\delta},y_r) + \lambda \sum_{p' \in N_{ij}} l_{p'}(\delta-\bar{\delta}) = 0 \qquad (11)$$

where $\overline{\Delta I^{rl}} = I^r(x_l+\bar{\delta},y_l) - I^l(i,j)$. In total, there are as many equations as there are points (*i.e.*, unknowns). The system is formulated into a tri-diagonal matrix, thus allowing use of the iterative Gauss-Seidel method. The solution at the $k^{\mathrm{t}h}$ iteration is given by the relation:

$$\delta^k_{i,j} = \bar{\delta}^{k-1}_{i,j} - \frac{\overline{\Delta I^{rl}} I^r_x}{\lambda \sum_{p' \in N_{ij}} l_{p'} + (\mathrm{I}^r_x)^2} \qquad (12)$$

where δ^{k-1} is the disparity field estimated at the $(k-1)$ iteration.

It is particularly important to determine a termination condition for the algorithm, so that convergence is ensured. For this reason, a convergence

measure is examined periodically (every ten iterations), which is the average correction in the direction of the slope:

$$\Delta_T = \sum_{(i,j)} \frac{\overline{|\Delta I^{rl}} I_x^r|}{\lambda \sum_{p' \in N_{ij}} l_{p'} + (I_x^r)^2}$$

The algorithm is terminated when the percentage diminishment of the Δ_T amount becomes less than a threshold. This condition on the diminishment Δ_T ratio is proven to be more stable and reliable for different stereoscopic pairs than a threshold condition on the Δ_T amount itself.

2.1 Multiscale Algorithm

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.

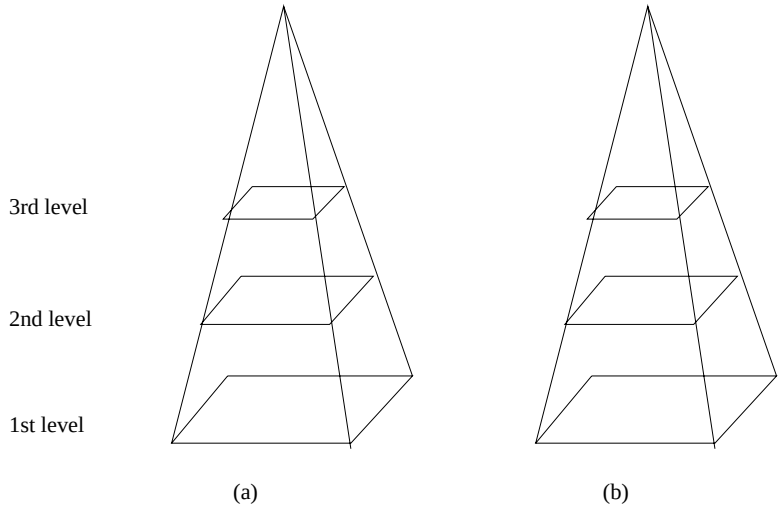


Figure 2: (a) Data pyramid and (b) Field pyramid

Consequently, a coarse-to-fine multiscale method in a pyramidal form (Fig. 2) is implemented, where in each level (l) the algorithm is applied to images of submultiple dimensions of the original ones. Those images are the result of reduction by low-pass filtering and subsampling. The image dimensions at level l are both reduced by a factor 2^l. Convergence of the algorithm at level l implies a dense disparity field where each of its 1-D vector $\delta^{(l)}$ corresponds to an initial image region of size $2^l \times 2^l$. The estimated at level l disparity field $\delta^{(l)}$ constitutes the initialization of the field $\delta^{(l-1)}$ to be estimated at level $l-1$. An immediate result of this reduction is the scale change on the magnitude of the field to be estimated. Because of the reduction, the

characteristics (features) to be matched in the two images differ by a smaller absolute value (number of pixels). If, for example, $(x_l,\ y_l)$ and $(x_l + \delta,\ y_l)$ are the coordinates of corresponding points at level (l), then at the higher level $(l+1)$ the coordinates will be $(\frac{x_l}{2},\ \frac{y_l}{2})$ and $(\frac{x_l}{2} + \frac{\delta}{2},\ \frac{y_l}{2})$, respectively. This means that at a higher level the disparity to be estimated has a submultiple magnitude, thus, making the algorithm more efficient. This way, at the higher levels a coarse estimation of the disparity field is achieved which becomes finer, in exactness and detail, at the lower levels of the pyramid. In the last level, the base of the pyramid, the algorithm is applied to the original images and its convergence provides the final disparity field.

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.

3.1 Optical Flow - Disparity Relations

Combining equations 6-7, the following relations among the components of the fields to be estimated are derived:

$$y_l = y_r \quad \Rightarrow \quad v_l = v_r \tag{13}$$

and

$$x_r = x_l + \delta \quad \Rightarrow \quad u_r = u_l + \delta^{t+1} - \delta^t \tag{14}$$

Solving the system of these two equations with respect to δ^{t+1} and v_r results into:

$$v_r = v_l \quad \text{and} \quad \delta^{t+1} = u_r - u_l + \delta^t \tag{15}$$

Therefore, to completely determine the requested fields it is sufficient to evaluate their three components u_r, u_l and v_l.

3.2 Proposed Solution

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. In the second stage, the aim is to determine two points in the second pair of images for every pair of points of the first stereoscopic image pair, so that these four corresponding points are the projections of the same point in space. The first point pair is the projection at time instant t while the second is at $t+1$. The principle of optical-flow preservation is used; that is, the projections in the four images of

the same feature of an object in space are of the same light intensity. Thus, the following relations result:

$$\begin{aligned} e_s^1 = I_{t+1}^r - I_t^r &= 0 \\ e_s^2 = I_{t+1}^l - I_t^l &= 0 \\ e_s^3 = I_{t+1}^r - I_{t+1}^l &= 0 \end{aligned}$$

where I_{t+1}^l and I_{t+1}^r are the intensities of the left and right points in the second stereoscopic image pair, while I_t^l and I_t^r are the intensities of the corresponding points in the first stereoscopic image pair.

As with the solution to the stereoscopic problem, the minimization of the squared deviation from the optical-flow preservation principle and the minimization of a smoothness measure for the estimated fields are considered. In addition to the fields for direct evaluation, a smoothness measure for the quantity $(u_r - u_l)$ is introduced; that is, the velocity difference in the two image sequence which in the case of the lateral model is $u_r - u_l = f\frac{V_z}{Z^2}$. Smoothness of this quantity is achieved with a coefficient μ in relation to the smoothness of fields u_r, u_l and v_l. The total quantity to be minimized is:

$$\begin{aligned} E &= \int\int((e_s^1)^2 + (e_s^2)^2 + (e_s^3)^2)dxdy + \lambda\mu \int\int \left(g(\frac{\partial(u_r - u_l)}{\partial x}) + g(\frac{\partial(u_r - u_l)}{\partial y}) \right) dxdy \\ &+ \lambda \int\int \left(g(\frac{\partial u_r}{\partial x}) + g(\frac{\partial u_r}{\partial y}) + g(\frac{\partial u_l}{\partial x}) + g(\frac{\partial u_l}{\partial y}) + g(\frac{\partial v_r}{\partial x}) + g(\frac{\partial v_r}{\partial y}) \right) dxdy \end{aligned} \quad (16)$$

In the discrete case this quantity becomes:

$$\begin{aligned} &\sum_{(i,j)} (\Delta I^l)^2 + (\Delta I^r)^2 + (\Delta I^{rl})^2 \\ &+\lambda \sum_{(i,j)} \sum_{p' \in N_{ij}} (g(u_{ij}^r - u_{p'}^r) + g(u_{ij}^l - u_{p'}^l) + g(v_{ij}^l - v_{p'}^l) + g((u_{ij}^r - u_{ij}^l) - (u_{p'}^r - u_{p'}^l)) \end{aligned}$$

where

$$\begin{aligned} \Delta I^{rl} &= I_{t+1}^r - I_{t+1}^l \\ \Delta I^r &= I_{t+1}^r - I_t^r \\ \Delta I^l &= I_{t+1}^l - I_t^l \end{aligned}$$

Minimizing this last quantity the following equations are obtained:

$$\Delta I^l \, I_x^l - \Delta I_{t+1}^{rl} \, I_x^l + \lambda \sum_{p' \in N_{ij}} l_{p'}^1 (u_{ij}^l - u_{p'}^l) + \mu \, l_{p'}^4 \, ((u_{ij}^l - u_{ij}^r) - (u_{p'}^l - u_{p'}^r)) = 0 \quad (17)$$

$$\Delta I^r \, I_x^r + \Delta I_{t+1}^{rl} \, I_x^r + \lambda \sum_{p' \in N_{ij}} l_{p'}^2 (u_{ij}^l - u_{p'}^l) + \mu \, l_{p'}^4 \, ((u_{ij}^r - u_{ij}^l) - (u_{p'}^r - u_{p'}^l)) = 0 \quad (18)$$

$$\Delta I^l \, I_x^l + \Delta I^r \, I_x^r + (\Delta I_{t+1}^{rl}) \, (I_x^r - I_x^l) + \lambda \sum_{p' \in N_{ij}} l_{p'}^3 (v_{ij}^l - v_{p'}^l) = 0 \quad (19)$$

where the coefficients $l^1_{p'}, l^2_{p'}, l^3_{p'}$ and $l^4_{p'}$ result from the smoothing of u^l, u^r, v^l and ($u^r - u^l$), respectively, and the derivatives of the light intensities refer to the second stereoscopic image pair (*i.e.*, at time instant $t+1$).

Let

$$\bar{u^l_{ij}} = \frac{\sum_{p' \in N_{ij}} l^1_{p'} u^l_{p'}}{\sum_{p' \in N_{ij}} l^1_{p'}} + \mu \frac{\sum_{p' \in N_{ij}} l^4_{p'} u^l_{p'}}{\sum_{p' \in N_{ij}} l^4_{p'}}$$

and similarly $\bar{u^r_{ij}}$, while let $\bar{v^l_{ij}}$ be:

$$\begin{aligned} \bar{v^l_{ij}} &= \frac{\sum_{p' \in N_{ij}} l^3_{p'} v^l_{p'}}{\sum_{p' \in N_{ij}} l^3_{p'}} \quad \text{and} \\ V^l_{ij} &= \sum_{p' \in N_{ij}} l^3_{p'} (v^l_{ij} - \bar{v^l_{ij}}) \end{aligned}$$

Assuming that the fields to be estimated are small in magnitude and that the intensities vary smoothly, the following approximations are used:

- approximation of the first derivative of the intensity:

$$I^r_x \triangleq \frac{\partial I^r_{t+1}(x_r + u_r, y_l + v_l)}{\partial x} \approx \frac{\partial I^r_{t+1}(x_r + \bar{u}_r, y_l + \bar{v}_l)}{\partial x}$$
$$I^r_y \triangleq \frac{\partial I^r_{t+1}(x_r + u_r, y_l + v_l)}{\partial y} \approx \frac{\partial I^r_{t+1}(x_r + \bar{u}_r, y_l + \bar{v}_l)}{\partial y}$$

  Similar approximations hold true for the intensities in the left image (I^l_x and I^l_y).

- approximation of the intensities at points $(x_r + u_r, y_l + v_l)$ and $(x_l + u_l, y_l + v_l)$ in the left and right image, respectively, with first order Taylor expansion:

$$I^r_{t+1}(x_r + u_r, y_l + v_l) \approx I^r_{t+1}(x_r + \bar{u}_r, y_l + \bar{v}_l) + (u_r - \overline{u}_r) I^r_x + (v_r - \overline{v}_r) I^r_y$$

$$I^l_{t+1}(x_l + u_l, y_l + v_l) \approx I^l_{t+1}(x_l + \overline{u_l}, y_l + \overline{v_l}) + (u_l - \overline{u}_l) I^l_x + (v_l - \overline{v}_l) I^l_y$$

The final form of the equations in the discrete grid are:

$$Q \begin{bmatrix} u^k_l - \overline{u}^{k-1}_l \\ u^k_r - \overline{u}^{k-1}_r \\ v^k_l - \overline{v}^{k-1}_l \end{bmatrix} = b \tag{20}$$

where

$$Q = \begin{bmatrix} 2(I^l_x)^2 + \lambda & -I^l_x I^r_x - \lambda\,\mu \sum_{p' \in N_{ij}} l^4_{p'} & I^l_x (2I^l_y - I^r_y) \\ -I^l_x I^r_x - \lambda\,\mu \sum_{p' \in N_{ij}} l^4_{p'} & 2(I^r_x)^2 + \lambda & I^r_x (2I^r_y - I^l_y) \\ I^l_x (2I^l_y - I^r_y) & I^r_x (2I^r_y - I^l_y) & \lambda + (I^l_y)^2 + (I^r_y)^2 + (I^r_y - I^l_y)^2 \end{bmatrix}$$

and

$$b = \begin{bmatrix} I_x^l(\Delta I^l - \Delta I^{rl}) + \lambda\,\mu\,(\sum\limits_{p' \in N_{ij}} l_{p'}^4\,((\mu-1)\dfrac{\sum\limits_{p' \in N_{ij}} l_{p'}^4 u_r}{\sum\limits_{p' \in N_{ij}} l_{p'}^4} + \dfrac{\sum\limits_{p' \in N_{ij}} l_{p'}^2 u_r}{\sum\limits_{p' \in N_{ij}} l_{p'}^2} \\ I_x^r(\Delta I^r + \Delta I^{rl}) + \lambda\,\mu\,(\sum\limits_{p' \in N_{ij}} l_{p'}^4\,((\mu-1)\dfrac{\sum\limits_{p' \in N_{ij}} l_{p'}^4 u_l}{\sum\limits_{p' \in N_{ij}} l_{p'}^4} + \dfrac{\sum\limits_{p' \in N_{ij}} l_{p'}^2 u_l}{\sum\limits_{p' \in N_{ij}} l_{p'}^2} \\ I_y^l(\Delta I^l - \Delta I^{rl}) + I_y^r(\Delta I^r + \Delta I^{rl}) \end{bmatrix}$$

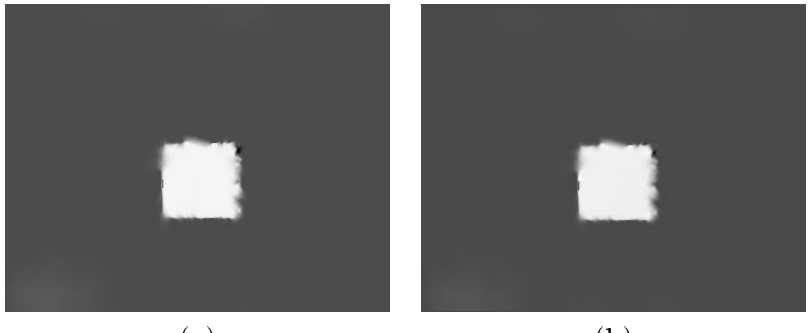
The above system is represented by a positive-definite matrix; thus, its solution is obtained through direct inversion.

4 Experimental Results

4.1 Stereoscopic Problem Solutions

This algorithm was tested with both synthetic and real data. In the synthetic case, the image of a square is given, with constant disparity $\delta = 5$ for the square and $\delta = 0$ for the background. The mean squared error in evaluating the field δ was 0.0809, and the mean squared difference between the right image and its reconstruction using the estimated field was 0.257869. The largest scale where the field was estimated was 2, while the number of iterations necessary for the algorithm to converge was 150. The value of the coefficient λ was 500.

A quantitative evaluation of the field δ is shown in Fig. 3(a), where the corresponding depth map is given.



(a) (b)

Figure 3: Depth map at time (a) t, and (b) $t + 1$

Real data consist of a stereoscopic pair shown in Fig. 4. For $\lambda = 500$, the resulting mean squared error on the intensities was 14.76. In addition, Fig. 5(a) shows a depth evaluation, as it results from Eq. (5), of the depicted objects.

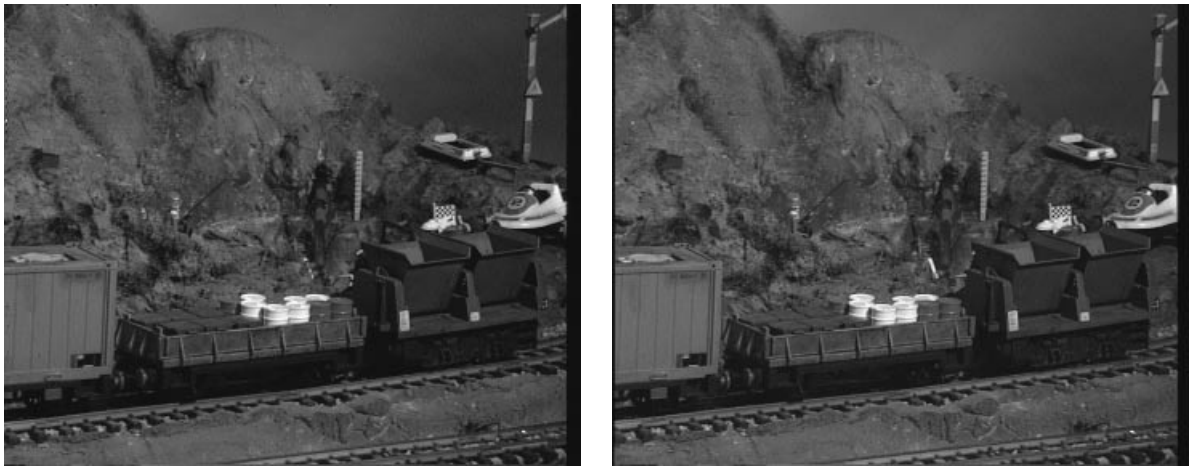


Figure 4: Stereoscopic image pair at time t

4.2 Stereoscopy and Motion

The results on synthetic data from the second stage of the work are concluded in Table 1. Tests on synthetic data were realized using two categories of images:

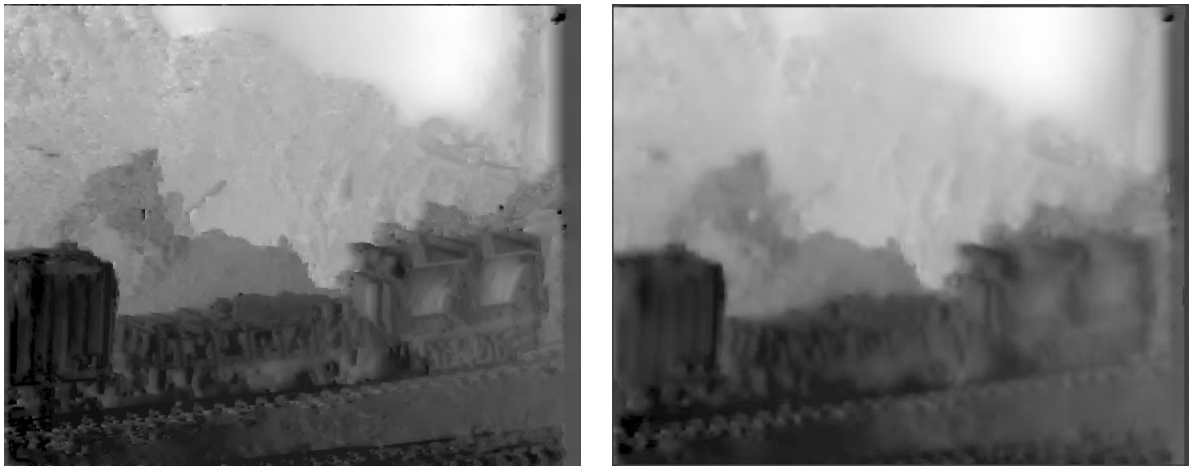
- The second stereoscopic image pair follows from the first by translating the components U_l, U_r and $V_l = V_r$ which are varied from the left top to the right bottom point of the image:
  U_l from 1.0 to 3.0
  U_r from 1.0 to 2.0
  $V_l = V_r = 2.0$ (constant)

- The second stereoscopic image pair follows from the first by translating the depicted square (where $\delta = 5$) as it results from the constant fields $U_l = 3.0$, $U_r = 3.0$, $V_l = V_r = 2.0$. The remaining image is translated by a vector $(-1, -1)$ for both the left and the right sequence.

Category	Iterations	λ	MSE			MSE on Intensities		
			U_l	U_r	V	LL	RR	RL
1	310	150	0.0320	0.01	0.008	1.16	1.67	1.00
2	320	150	0.147	0.143	0.318	1.78	1.76	0.951

Table 1: Experimental results on synthetic data

In Table 1 the MSE expresses the mean squared deviation for each of the estimated fields; that is, it results from $\sum\sum(\hat{u}_{ij} - U_{ij})^2$, where $\hat{u}$ is the



(a) (b)

Figure 5: Estimated depth map at time (a) t, and (b) $t+1$

estimated field, while U is the synthetically created field, on which production of the second stereoscopic image pair was based. In the same table MSE on intensities are given too, for the left images (LL), for the right images (RR) and for the stereoscopic pair at $t+1$ (RL). A qualitative evaluation of the field δ^{t+1} is shown in Fig. 3(b).

Implementation of the second stage with real data consists of two stereoscopic pairs, the first of which was the one used during the first stage of this work. The result of the first stage (*i.e.*, correspondence between image points in the first pair) is assumed given and is utilized for the simultaneous estimation of motion and disparity at the next time instant. Qualitatively, the motion of scene objects consists of two components:

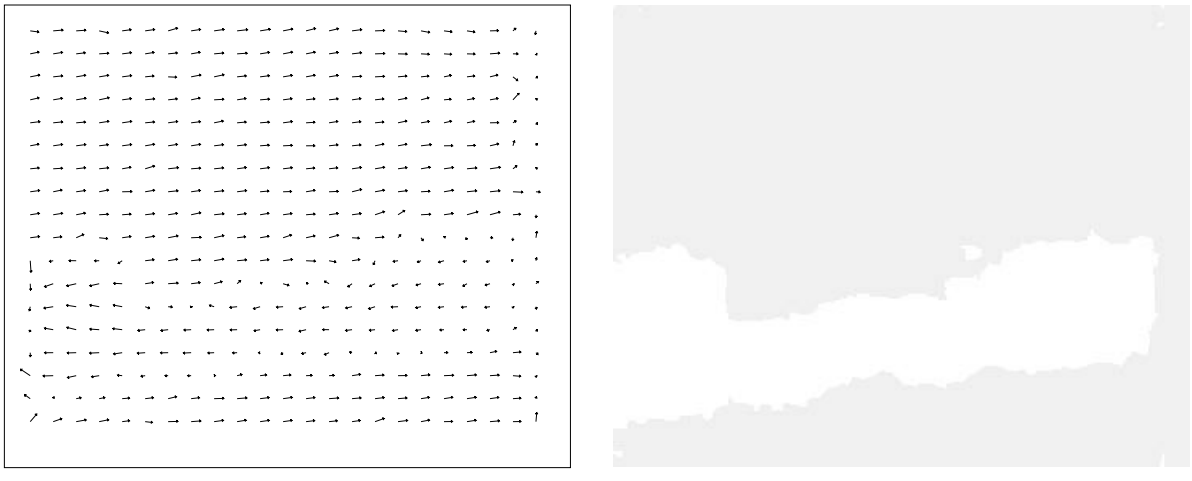
- a camera movement from right to left, which creates the optical effect of scene objects moving towards the opposite direction
- movement of the depicted train from right to left

Maximum level	λ	Iterations	MSE on Intensities		
			LL	RR	RL
3	150	730	3.64	3.34	15.51

Table 2: Stereoscopy and motion on real data

The estimated motion field for a scale depth equal to 3 and $\lambda = 150$ is shown in Fig. 6, where the unified motion of the background to the right,

and the independent motion of the train to the left, are clearly distinguished. Other numerical results are given in Table 2. The estimated motion fields were then segmented into regions using a K-Means type algorithm, in which the number of segments should be provided *a priori*. The distance between a segment s and a point p was defined to be the quantity: $\| \hat{U}_s - U_p \|$, where $\hat{U}_s$ is the mean of the velocity vectors at the points that belong to segment s and U_p the velocity vector at point p. The resulting segmentation on the left motion field is shown in Fig. 6, were the regions that correspond to the train and to the camera's motion are clearly distinguished. The evaluated depth map for the time instant $t+1$ is shown in Fig. 5(b).



(a) (b)

Figure 6: (a) Estimated dense velocity field for the left camera, and (b) corresponding segmentation

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 both synthetic and real data. Specifically, the approach was applied to an image sequence of a real scene, where both the object and the binocular system are moving. The estimated velocity field was then segmented using a K-Means type algorithm. Currently we are working on the reconstruction of the 3-D motion for the different segments using robust regression techniques. It would be also useful to incorporate the regression scheme into the segmentation phase, that is to segment the velocity field into regions that correspond to the same 3-D motion. 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.

References

[1] N. Alvertos, D. Brazakovic, and R. C. Gonzalez, "Camera geometries for image matching in 3-D machine vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-11, No. 9, pp. 897-915, Sept. 1989.

[2] E.Grosso and M.Tistarelli, " Active/Dynamic Stereo Vision," *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. 17, No. 9, pp. 868-879, Sept. 1995.

[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 Journal of Robotics and Automation*, Vol. 3, No. 6, pp. 599-614, Dec. 1987.

[5] M. K. Leung and T. S. Huang, "An integrated approach to 3D motion analysis and object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-13, No. 10, pp. 1075-1084, Oct 1991.

[6] S.Z.Li, "On Dincontinuity-Adaptive Smoothness Priors in Computer Vision", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. 17, No. 6, pp. 576-586, June 1995.

[7] A. Mitiche, "On kineopsis and computation of structure and motion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 1, pp. 109-112, Jan. 1986.

[8] A. Mitiche and P. Bouthemy, "Tracking modelled objects using binocular images", *Computer Vision, Graphics and Image Processing*, Vol. 32, pp. 384-396, 1985.

[9] A. N. Netravali, T. S. Huang *et al*, "Algebraic methods in 3D motion estimation from two-view point correspondences", *Int. Journal of Imaging Systems and Technology*, Vol. 1, pp. 78-99, 1989.

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

[11] G. Tziritas and C. Labit, *Motion analysis for image sequence coding*, Elsevier, pp. 366+xxiv, 1994.

[12] A. M. Waxman and J. H. Duncan, "Binocular image flows: steps forward stereo-motion fusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, pp. 715-729, Nov. 1986.

[13] Z. Zhang and O. Faugeras, "Estimation of displacements from two 3-D frames obtained from stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-14, No. 12, pp. 1141-1156, Dec. 1992.