Construction of multiple views using jointly estimated motion and disparity fields

I. Patras

*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.A. Hendriks

Delft University of Technology,
Dept. of EE, Information Theory Group
Mekelweg 4, 2628 CD Delft, The Netherlands

G. Tziritas*

E-mail: {patras,tziritas}@csd.uch.gr, hendriks@et.tudelft.nl

Keywords: Joint Motion/Disparity estimation, Stereo interpolation

ABSTRACT

This work aims at determining dense motion and disparity fields given a stereoscopic sequence of images for the construction of stereo interpolated images. At each time instant the two dense motion fields, for the left and the right sequences, and the disparity field of the next stereoscopic pair are jointly estimated. The disparity field of the current stereoscopic pair is considered as known. The disparity field of the first stereoscopic pair is estimated separately. For both problems multi-scale iterative relaxation algorithms are used. Stereo occlusions and motion occlusions/disclosures are detected using error confidence measures. For the reconstruction of intermediate views a disparity compensated linear interpolation algorithm is used. Results are given for real stereoscopic data.

1. INTRODUCTION

One of the drawbacks of existing stereoscopic imaging systems is the lack of a look-around capability. The viewer can only observe the recorded scene without distortion from the viewpoint of the cameras. To overcome this problem intermediate views, depending on the position of the viewer, have to be constructed. If a 3D-model of the scene is available the construction of the intermediate views can be done by a perspective transformation. Unfortunately most real world image sequences are too complex to model. Therefore most existing methods which consider the problem of the construction of intermediate views interpolate between subsequent stereoscopic pairs by using the corresponding disparity fields.[9] However in approaches where the disparity fields are independently estimated at subsequent time instances the interpolated images suffer from annoying artifacts due to temporal

inconsistencies. Temporal consistency restrictions have been reported to produce more natural result[3,5] but most approaches found in the bibliography[11] utilize the relations between the motion and the disparity fields in subsequent steps.[2] This work aims at an integrated approach to the problem of dynamic stereoscopic vision, which utilizes the interdependencies of disparity and motion fields in their estimation phase. In this way a temporal consistency restriction is imposed over the disparity fields and thus over the stereo interpolated images. In a similar approach Tamtaoui and Labit[10] jointly estimate the two motion fields under stereoscopic constraints, to solve the problem of motion and disparity estimation.

At any time instant two dense motion fields (for the left and right image sequences) and the dense disparity field for the next stereoscopic pair are computed. The disparity field of the current 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 adaptively 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. For the first stereoscopic pair of the sequence the disparity field is estimated by minimizing a cost function using a similar iterative relaxation algorithm.

The disparity fields which result this way are then used for the construction of intermediate views using a disparity compensated linear interpolation algorithm. The interpolation process includes the detection of stereo occluded areas which is performed using error confidence measures.

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 focal length which is identical 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 this assumption holds, thus $\vec{d}$ is a 1-D vector along the x-axis.

In Section 2 we present a regularization method for obtaining a smooth disparity field from a stereoscopic pair using discontinuity adaptive functions. Also a method for detecting stereo occlusions is proposed. In Section 3 the disparity field of the next stereoscopic pair is partially constructed simultaneously with the estimation of the two motion fields. In Section 4 the detection of motion occlusions/disclosures is described and the construction of the disparity field between the second stereoscopic pair is completed. In Section 5 the stereo interpolation method is presented, and in Section 6 some results are given with real stereoscopic data.

2. DISPARITY FIELD ESTIMATION AND STEREO OCCLUSION DETECTION BETWEEN A STEREOSCOPIC PAIR OF IMAGES

In this section we present a regularization method for the estimation of a dense disparity field from a stereoscopic pair, using discontinuity adaptive functions. The estimated disparity field and error confidence measures are then used for the construction of maps where stereo occluded areas are detected.

2.1. Disparity Field Estimation

The solution to the stereoscopic problem of finding the correspondence between left and right images 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. From the intensity preservation principle, it follows 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[4]

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

where $\mathcal{N}_{\imath\jmath} = \{(\imath - 1, \jmath), (\imath + 1, \jmath), (\imath, \jmath - 1), (\imath, \jmath + 1)\}$ is the 4-point neighborhood of $(\imath, \jmath)$. The dependence of δ on $(\imath, \jmath)$ is omitted for simplifying the notation. $g(.)$ is a discontinuity adaptive function,[7] which, if it is carefully chosen, may regularize the solution and at the same time preserve the discontinuities. In that framework the *adaptive 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:

$$(I_r(\imath + \delta, \jmath) - I_l(\imath, \jmath))\, I_{rx}(\imath + \delta, \jmath) + \lambda\alpha(\delta - \bar{\delta}) = 0 \tag{1}$$

where

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

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

$$I_{rx}(\imath + \delta, \jmath) = I_{rx}(\imath + \bar{\delta}, \jmath) \quad \textit{and} \quad I_r(\imath + \delta, \jmath) = I_r(\imath + \bar{\delta}, \jmath) + (\delta - \bar{\delta}) I_{rx}(\imath + \bar{\delta}, \jmath) \tag{2}$$

Considering the above, the following iterative solution is derived from Eq. (1):

$$\delta^k = \bar{\delta}^{k-1} - \frac{\overline{\Delta I_{rl}} I_{rx}}{\lambda\alpha + (I_{rx})^2} \tag{3}$$

where δ^k is the disparity field estimated at the k iteration and $\overline{\Delta I_{rl}} = I_r(\imath + \bar{\delta}, \jmath) - I_l(\imath, \jmath)$. 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 multi-scale method in a pyramidal form is implemented, where in the upper levels the algorithm is applied to images of sub-multiple dimensions of the original ones.[1] Those images are the result of reduction by low-pass filtering and sub-sampling. 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 sub-sampling, larger values of γ impose a "harder" smoothing, while at finer levels of detail the discontinuities are more carefully preserved.

2.2. Stereo occlusion detection

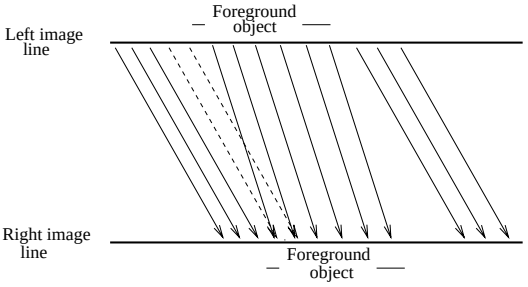


Figure 1: Stereo occlusion detection

For the stereo occlusion detection a 1-D disparity field, assigned to the left image of the stereoscopic pair is assumed. The assumption of only horizontal disparity enables us to detect occlusions in a line-by-line way, that is at each time to take into consideration only the corresponding lines of the left and the right image. Assuming that for areas that can be seen by both cameras the disparity estimation is accurate, Fig. 1 reveals that at points that can be seen only by the left camera multiple matches can occur. To remove the conflict we compare the error confidence measures of the conflicting matches. The measure that we used was a quantity derived from an error analysis of the gradient based methods,[6] extended to the 4-neighborhood of the left point of the match in question.

$$\mathrm{ECM}(i,j) = \sum_{|k|+|l|<2} h(k,l)\frac{|\Delta I(i-k,j-l)|}{\|I_x^l\|} \quad , \quad h(k,l) = \begin{bmatrix} & 1 & \\ 1 & \boxed{2} & 1 \\ & 1 & \end{bmatrix} \tag{4}$$

where $\Delta I(m,n) = I_r(m+\delta(m,n),n) - I_l(m,n)$

The match with the smallest error confidence measure is declared correct and the endpoints at the left image of all the other conflicting matches are declared left occluded areas. Right occluded areas are detected at points in the right image that are not matched.

3. JOINT MOTION/DISPARITY ESTIMATION

Once the stereo occlusion areas in the first stereoscopic pair are detected, a multi-scale iterative relaxation algorithm[8] is applied in order to estimate the two motion fields and (partially) the disparity field for the second stereoscopic pair of images. 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 fields and the second disparity field, 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 $(\imath, \jmath)$ and $(\imath', \jmath')$

$$v_r(\imath', \jmath') = v_l(\imath, \jmath) \quad \text{and} \quad \delta_{t+1}(\imath + u_l, \jmath + v_l) = u_r(\imath', \jmath') - u_l(\imath, \jmath) + \delta_t(\imath, \jmath) \tag{5}$$

Therefore, for points that are not stereo occluded 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}.

The solution with respect to the motion and disparity fields is achieved with the minimization of a cost function which consists of three major parts. One for points that are not stereo occluded, and two for the left and the right stereo occluded areas respectively. For the former the estimation of the left and right motion fields and the estimation of the second disparity field are interconnected, while for the later the motion fields are estimated independently.

For areas that are not stereo occluded the minimization of the squared deviation from the image intensity preservation principle and a smoothness constraint for the estimated fields are considered. The total quantity is:

$$DFD_{ll} + DFD_{rr} + DFD_{lr} + \lambda(G(u_r) + G(u_l) + G(v_l)) \tag{6}$$

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 components of the velocity fields which is achieved by means of the Discontinuity Adaptive Function (DAF)[7] which was defined in Section 2. The term $G(.)$ is defined as the sum of the local cost terms that $g(.)$ imposes, that is (for field u_l for example): $G(u_l) = \sum_{(i,j)} \sum_{p \in \mathcal{N}_{ij}} g(u_l - u_l^p)$. We should note that all the terms refer only to points that are not stereo occluded at the first stereoscopic pair.

For stereo occluded areas the left and the right motion fields are independently estimated in a way similar to the monocular motion analysis.[4] However a DAF is used as a regularization means in this case too. The cost term for such areas (e.g. left occluded areas) is:

$$DFD_{ll} + \lambda(G(u_l) + G(v_l)) \tag{7}$$

where the first term is the mean square DFD between the left images and the later terms refer to the smoothing of the left motion field. For all terms we refer only to points that belong to left occluded areas.

The solution with respect to the requested fields is achieved with the minimization of the total cost function, using an iterative relaxation algorithm which is based on its gradient. Three sets of equations are derived, whose iterative solution provides the estimation for the requested fields. For areas that are not stereo occluded the solution at the k$^{\text{th}}$ iteration is given by the relation:

$$\begin{bmatrix} u_l^k \\ u_r^k \\ v_l^k \end{bmatrix} = \begin{bmatrix} \bar{u}_l^{k-1} \\ \bar{u}_r^{k-1} \\ \bar{v}_l^{k-1} \end{bmatrix} + \begin{bmatrix} \frac{I_{lx}(\overline{\Delta I}_l - \overline{\Delta I}_{rl})}{2I_{lx}^2 + \lambda\alpha_l^u} \\ \frac{I_{rx}(\overline{\Delta I}_r + \overline{\Delta I}_{rl})}{2I_{rx}^2 + \lambda\alpha_r^u} \\ \frac{I_{ly}(\overline{\Delta I}_l - \overline{\Delta I}_{rl}) + I_{ry}(\overline{\Delta I}_r + \overline{\Delta I}_{rl})}{I_{ly}^2 + I_{ry}^2 + (I_{ly} - I_{ry})^2 + \lambda\alpha_l^v} \end{bmatrix} \tag{8}$$

where the notation and the details about the way the relation was derived are explained in Appendix A.

For the stereo occluded areas the resulting set of equations is similar to the one derived at monocular motion analysis. The solution at the k$^{\text{th}}$ iteration for the left motion field is given by the relation:

$$\begin{bmatrix} u_l \\ v_l \end{bmatrix} = \begin{bmatrix} \bar{u}_l \\ \bar{v}_l \end{bmatrix} + \frac{\overline{\Delta I}_l}{\lambda\alpha_l^u\alpha_l^v + \alpha_l^v I_{lx}^2 + \alpha_l^u I_{ly}^2} \begin{bmatrix} \alpha_l^v I_{lx} \\ \alpha_l^u I_{ly} \end{bmatrix} \tag{9}$$

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

A coarse-to-fine multi-scale approach[1] is used in the same way as in Section 2 in order to overcome the limitations of the gradient based algorithm.

4. MOTION OCCLUSION DETECTION AND CONSTRUCTION OF THE DISPARITY MAP

Applying the algorithm described in the previous section we obtain the two dense motion fields, and we can partially construct the disparity field δ^{t+1} for the second stereoscopic pair by using the Eq. (5). However we do not have information about δ^{t+1} at areas that the motion has disclosured, that is they were not visible at the first stereoscopic pair, and at areas that were stereo occluded at the first stereoscopic pair. In order to determine the disparity field at every point of the left image of the second stereoscopic pair, as a first step we detect the motion occluded/disclosured areas in the left image sequence.

The detection of motion occlusions/disclosures is based on an error confidence measures similar to that of Section 2. Via the velocity field (u_l, v_l) assigned to the left image of the first stereoscopic pair pixels at the left image of the second stereoscopic pair are matched with pixels at the left image of the first stereoscopic pair. With a reasoning similar to that of Section 2, occlusions are detected in the image at time instance t whenever multiple matches occur. To remove the conflict we compare the error confidence measures of the conflicting matches. The match with the smallest confidence error is declared right and the endpoints at the left image at time t of all the other conflicting matches are declared motion occluded. Motion disclosures are identified in the image at time instance $t+1$ at the points that are unmatched.

The detection of the motion occlusions/disclosures is followed by the construction of the dense disparity field assigned to the left image of the second stereoscopic pair. To do so, for each point $(\imath, \jmath)$ in the left image of the first stereoscopic pair, which is not motion or stereo occluded we find the corresponding point $(\imath + u^l, \jmath + v^l)$ 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 disclosured or points whose correspondences in the left image of the previous stereoscopic pair are stereo occluded. The disparity values for those areas are estimated using the multi-scale gradient based algorithm described in Section 2.

5. STEREO INTERPOLATION

The disparity fields constructed either by the method described in Section 2 or in Sections 3 and 4 are used for the construction of stereo interpolated images. The objective is to construct an interpolated image at position $s \in [0 \ldots 1]$, where the left and the right image of the pair are considered to be at position 0 and 1 respectively.

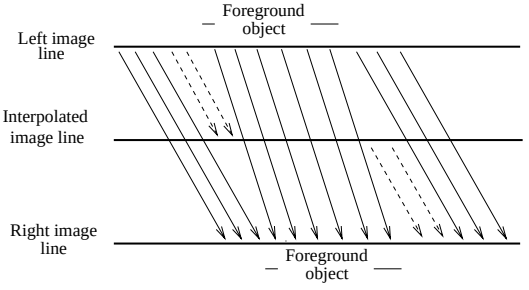


Figure 2: Interpolation

To construct the interpolated image, we project into it the disparity vectors that correspond to pixels in the left image that are not stereo occluded (Fig. 2). Conflicts that still may occur are removed by comparing the error confidence values of the matches that the conflicting disparity vectors imply. Then we assign an intensity

value to each pixel in the interpolated image to which a disparity vector projects. The intensity value assigned is a weighted average of the intensity values of the endpoint pixels of the disparity vector. More specifically if I^l and I^r are the intensity values at the endpoints then the intensity value I^i at the point in the interpolated image is given by:

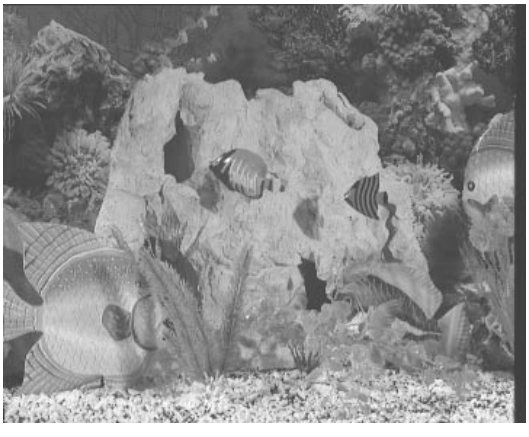
$$I^i = (1-s)I^l + sI^r$$

This procedure leaves a number of holes in the interpolated images at object points that can only be seen by one of the cameras. For those points we identify at which of the two images the corresponding area is visible by comparing the disparity vectors at the edges of the hole. As Fig. 2 reveals, if the vector at the left(right) edge of the hole is larger then the vector at the right(left) edge of the hole, then the correspondences can be found at the left(right) image. The intensity values for the points of the holes are then copied from the image for which the correspondence was found.

6. RESULTS

The algorithms have been tested on real stereoscopic sequences. Intermediate views have been constructed for subsequent frames of the stereoscopic sequences "aqua", "train" and "Claude" with baselines 8.75cm, 8.75cm and 25cm respectively. For all sequences the disparity fields used for the construction of interpolated images for the first stereoscopic pair are estimated by the algorithm of Section 2. The depth map for the 10th stereoscopic pair of the sequence "aqua", which was constructed using the disparity map estimated this way, is presented in Fig. 3(a). The corresponding interpolated image in the middle viewpoint, that is at position $s = 0.5$, is presented in Fig. 3(b). Interpolated images for the first time instances of the sequences "train" and "Claude" are presented in Fig. 4(a) and Fig. 4(b).

(a) Depth map



(b) Interpolated image

Figure 3: Depth map and interpolated image for the 10th stereoscopic pair of "aqua"

The interpolated images for the rest of the sequence are constructed with the use of the disparity fields estimated by the algorithm of Section 3. Interpolated images for subsequent time instances are presented for the sequence "train" which contains the most complex scene of the three (with respect to motion), since the apparent motion includes two independent motions (two trains) and the motion of the scene. The interpolation is done at position $s = 0.5$ and the result is presented in Fig. 5. Finally in Fig. 6 the depth maps at frame 2 and frame 8 are presented.

(a) "train" (b) "Claude"

Figure 4: Interpolated images for the 1st stereoscopic pairs of "train" and "Claude" sequences

7. CONCLUSIONS

A method for the joint estimation of motion and disparity fields for the construction of stereo interpolated images in a stereoscopic image sequence was presented. Results with real data were given in order to exhibit the performance of the proposed method. The evaluation was done qualitatively. Due to the joint motion and disparity estimation, the spatial and the temporal consistency of the disparity fields are good and the intermediate pictures natural.

However, in the joint motion/estimation method proposed in this work the motion occlusion/disclosures and stereo occlusions are detected at steps that follow the estimation of the corresponding fields. In this way the estimation of the disparity and motion fields near these areas is deteriorated. Most visible artifacts in the interpolated images appear in such areas. A method that would incorporate the detection phase into the estimation process might improve further the results. Furthermore the use of a spatial continuity constraint in the detection of occlusions, both for stereo and motion, should be investigated in future work.

8. Acknowledgment

This work has been carried out within the Research Network "Motion Analysis for Advanced Image Communication Systems" (MANADIX), which is supported by the Human Capital and Mobility Programme of the European Community, and within the framework of the European ACTS AC092 PANORAMA project.

(a) frame 2

(b) frame 4

(c) frame 6

(d) frame 8

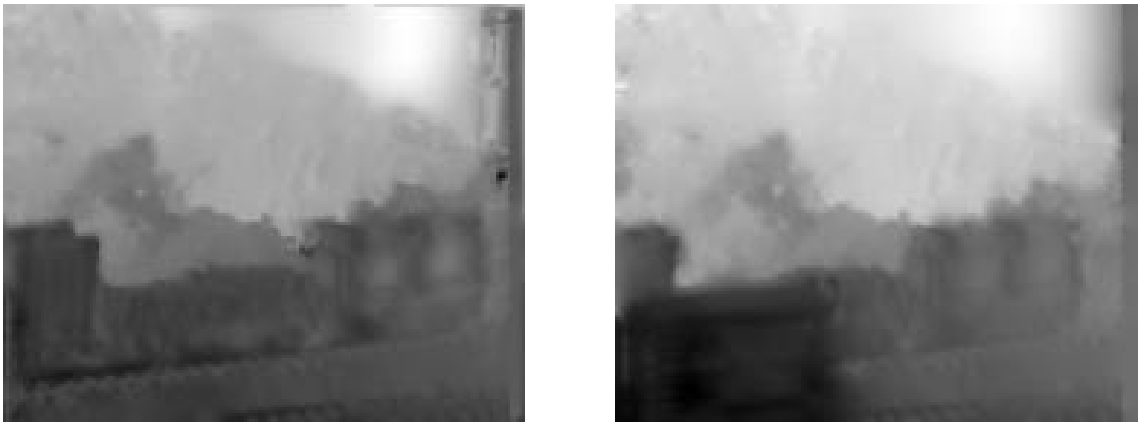
Figure 5: Interpolated images for sequence "train"

A DERIVATION OF THE ITERATIVE SOLUTION FOR JOINT MOTION/DISPARITY FIELD ESTIMATION

The component of the cost function which corresponds to areas that are not stereo occluded at the first stereoscopic pair is:

$$\sum_{(i,j)}((\Delta I_l)^2 + (\Delta I_r)^2 + (\Delta I_{rl})^2 + \lambda \sum_{p\in\mathcal{N}_{(i,j)}} g(u_r - u_r^p)) + \lambda \sum_{(i,j)} \sum_{p\in\mathcal{N}_{(i,j)}} ((g(u_l - u_l^p) + g(v_l - v_l^p)) \qquad (10)$$

where $\Delta I_l = I_l(i,j,t) - I_l(i+u_l, j+v_l, t+1)$, $\Delta I_r = I_r(i',j',t) - I_r(i'+u_r, j'+v_l, t+1)$
and $\Delta I_{rl} = I_r(i'+u_r, j'+v_l, t+1) - I_l(i+u_l, j+v_l, t+1)$
The points (i,j) over which the summation is done, are the points at the first left image that are not stereo occluded. For simplification reasons the dependence of u_l, u_r and v_l on (i,j) is omitted.



(a) frame 2 (b) frame 8

Figure 6: Depth maps for sequence "train"

Let us define an interpolation operation on the u_l field using the adaptive 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 $t+1$

$$I_l(\imath + u_l, \jmath + v_l) \approx I_l(\imath + \bar{u}_l, \jmath + \bar{v}_l) + (u - \bar{u}_l) I_{lx}(\imath + \bar{u}_l, \jmath + \bar{v}_l) + (v - \bar{v}_l) I_{ly}(\imath + \bar{u}_l, \jmath + \bar{v}_l) \tag{11}$$

and the same for $I_r(\imath + u_r, \jmath + v_r)$.

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

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

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.

The solution which is obtained with above assumptions are summarized in the following set of equations:

$$\begin{bmatrix} 2(I_{lx})^2 + \lambda\alpha_l^u & -I_{lx}I_{rx} & I_{lx}(2I_{ly} - I_{ry}) \\ -I_{lx}I_{rx} & 2(I_{rx})^2 + \lambda\alpha_r^u & I_{rx}(2I_{ry} - I_{ly}) \\ I_{lx}(2I_{ly} - I_{ry}) & I_{rx}(2I_{ry} - I_{ly}) & \lambda\alpha_l^v + (I_{ly})^2 + (I_{ry})^2 \\ & & +(I_{ry} - I_{ly})^2 \end{bmatrix} \begin{bmatrix} u_l - \bar{u}_l \\ u_r - \bar{u}_r \\ v_l - \bar{v}_l \end{bmatrix} = - \begin{bmatrix} I_{lx}(\overline{\Delta I}_l - \overline{\Delta I}_{rl}) \\ I_{rx}(\overline{\Delta I}_r + \overline{\Delta I}_{rl}) \\ I_{ly}(\overline{\Delta I}_l - \overline{\Delta I}_{rl}) \\ +I_{ry}(\overline{\Delta I}_r + \overline{\Delta I}_{rl}) \end{bmatrix} \tag{12}$$

The iterative solution of Eq. (8) is obtained by taking into consideration only the diagonal elements of Eq. (12).

9 REFERENCES

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

[2] B. Chupeau. A multiscale approach to the joint computation of motion and disparity. application to the synthesis of intermediate views. In *Proc. 4th European Workshop on 3DTV*, Oct. 1993. Rome, Italy.

[3] E.A Hendriks and Gy. Marosi. Recursive disparity estimation algorithm for real time stereoscopic video applications. In *IEEE International Conference on Image Processing*, pages 891–894, September 1996. Lausanne, Switzerland.

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

[5] E. Izquierdo and M. Ernst. Motion/Disparity analysis for 3D-Video-Conference applications. In *International Workshop on Stereoscopic and Three Dimensional Imaging*, Sept. 1995. Santorini, Greece.

[6] J.K. Kearney, W.B. Thompson, and D.L. Boley. Optical flow estimation: An error analysis of gradient based mathods with local optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2):pp. 229–244, Mar. 1987.

[7] 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.

[8] I. Patras, N. Alvertos, and G. Tziritas. Joint disparity and motion field estimation in stereoscopic image sequences. In *13th International Conference on Pattern Recognition*, pages 359–362, Aug. 1996. Vienna, Austria.

[9] P. Skerjanc and J. Liu. A three camera approach for calculating disparity and synthesizing intermediate pictures. *Signal Processing: Image Communication*, 4:55–64, 1991.

[10] A. Tamtaoui and C. Labit. Constrained disparity and motion estimators for 3dtv 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.

Ioannis K. Patras was born in Thessaloniki, Greece, in 1973. He received his B.S. in Computer Science from the University of Crete in 1994. Since 1994 he has been pursuing his M.Sc. degree, working as a research assistant in Computer Vision and Robotics group of Institute of Computer Science, Foundation of Research and Technology-Hellas. In 1996 he spent four months in the Laboratory of the Information Theory Group of Delft University of Technology as a graduate student. His research interests include Motion Analysis, Active Vision and Parallel and Distributed systems.