MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

3—6

Improved Disparity Estimation by Matching with an Adaptive Window

Ebroul Izquierdo M. *

Department of Electronic Systems Engineering

University of Essex

Abstract

A novel technique for disparity estimation based on block matching with a local adaptive window is introduced in this paper. In the proposed approach the size and shape of the reference window is calculated adaptively according to the degree of reliability of disparities estimated previously and the local variation of the disparity map. The performance of the method has been verified by computer simulations using synthetic data and several natural stereo sequences. Considerable improvements were obtained just in object borders or image areas that become occluded.

1 Introduction

Most of the emerging content-based multimedia technologies are based on efficient methods to solve machine early vision tasks. Among others tasks, pixel correspondence estimation is perhaps the crucial task in multiview image analysis. The solution of this problem is the key technology for the development of most leading-edge interactive telepresence systems. In this paper we present a robust framework for pixel correspondence estimation in video sequences taken simultaneously from different perspectives. An improved concept based on block matching with a local adaptive window is introduced. The size and shape of the reference window is calculated adaptively according to the degree of reliability of disparities estimated previously. This information is obtained from initial disparities, which are estimated by applying a well-known hierarchical block matching strategy [2, 4]. Disparity estimation by block-matching is a basic technique in stereo vision [1, 7]. Although, a great variety of block-matching based methods have been developed in the past, most of them use a correlation window of fixed size and shape. In the known literature only a few approaches can be found which locally change the size or shape of the matching windows. A pioneering work in this direction was introduced by Levine *et al.* [6] in the early seventies. They adapted the window size according to the intensity variation. In contrast to the approach presented in this paper, additional information (like reliability of previously estimated disparities and object contours) is not taken into account in [6]. In a most recent framework presented by Hoff and Ahuja [3], it is asserted that the matching process should integrate contour extraction and 3D surface generation in order to increase the accuracy of the correspondences when a correlation window with fixed shape is used. Although Hoff and Ahuja also argue that the window shape should be chosen according to the local changes in the disparity map, the shape of the correlation window is not directly adapted in their approach. In contrast to this and many other well-known disparity estimators [2, 3, 4] the adaptation of the window size and shape is the central strategy of the technique introduced in this paper.

2 Initial Displacement Estimation

To obtain initial displacement fields the hierarchical-block matching technique presented in [4] is applied. The reliability of the initial estimates is calculated by using a criterion based on the uniqueness constraint [3, 4] together with the analysis of the curvature of the correlation surface generated during the initial hierarchical block-matching step. Next, the initial displacement field is smoothed by applying a iterative filtering procedure. It is well-known that smoothing by applying a standard filter (e.g., a median or a mean-value filter) tends to smooth the displacement fields uniformly over the whole image. This means that discontinuities in the field due to depth or motion differences between different objects are also blurred. For this reason a smoothing process in which discontinuities along the object borders are preserved is more attractive. In this context we apply an iterative schema for vector field regularization, in which the degree of smoothing is weighted according to the reliability of the neighboring vectors. In this schema each vector is updated using the average of the nearest neighbors weighted by their respective reliability.

3 Matching with an Adaptive Window

The improvement of the initial displacement filed is carried out for those sampling positions with unreliable estimates. For any sampling position z we want to choose a window size and shape that provides a new estimate with higher reliability. Let us denote Ω the contour of any reference region. Here we use the term *reference region* instead of reference window to make it clear that by applying the technique described below, an

* Address: Department of Electronic Systems Engineering, University of Essex, Colchester, CO4 3SQ, United Kingdom. E-mail: ebroul@essex.ac.uk

ideal matching area with arbitrary size and shape can be found. The adaptation of Ω should be done taking into account the initial displacement vectors, their reliability and constraints inherent to the correspondence problem. To cope with this task we use a model consisting of an elastic contour with internal energy E_{int}, on which external forces are imposed. Starting with a sufficiently small contour Ω_0 around of z, it is deformed dynamically by the external forces until a *steady state* is reached. Assuming that near object borders the available initial information has a high degree of uncertainty, whereas inside the objects the estimated disparities are more reliable, the external forces should satisfy the following three requirements:

r1. Forces in the direction of sampling positions with reliable disparities should be large and they should be small in the direction of sampling positions with uncertain disparities, i.e., the force field increases proportionally to the disparity reliability.

r2. The external force should tend to expand the reference window in the direction in which the disparity vectors show a smooth course while no expansion force should work in direction of regions of the disparity field with high entropy.

r3. The external force should decrease continuously proportional to the distance between z and Ω.

Expanding Ω_0 continuously, the *steady state* is reached, when the potential energy E^p of the system is zero. A general energy functional for the contour can be given by:

$$E = \alpha E_{\text{int}} + \beta E_{ext}, \tag{1}$$

where E_{int} represents the tension of the elastic contour (internal forces) and E_{ext} is the energy produced by the external forces, which is defined as:

$$E_{ext} = \int_0^{2\pi} \int_0^{\rho(\theta)} F(\theta,\rho) d\rho d\theta . \tag{2}$$

In (2) $F(\theta,\rho)$ is the external force in any point $\tilde{z}$ in the direction given by the angle θ. In other words, if $\tilde{z} = (\theta,\rho)$ (using polar coordinates) is a point lying on Ω, then $F(\theta,\rho)$ describes the force, that expands Ω in the direction of the vector pointing out from z to $\tilde{z}$. This force is define as:

$$F(\theta,\rho) = h_1(\tilde{z},\sigma_\Omega^2) . h_2(\tilde{z},P_d), \tag{3}$$

where σ_Ω^2 measures the degree of smoothness of all disparity vectors in the region surrounded by Ω, $h_1(\tilde{z},\sigma_\Omega^2)$ is a function measuring the amount of support that σ_Ω^2 gives to $F(\theta,\rho)$ and $h_2(\tilde{z},P_d)$ is a function measuring the amount of support that the reliability P_d of the disparity vector d in the position $\tilde{z}$ gives to $F(\theta,\rho)$. The functions $h_1(\tilde{z},\sigma_\Omega^2)$ and $h_2(\tilde{z},P_d)$ should possess the three properties r1, r2 and r3. A possible form for these functions is

$$h_1(\tilde{z},\sigma_\Omega^2) = \begin{cases} K_1 \Big/ \sqrt{\|z-\tilde{z}\|} \cdot (1+\sigma_\Omega^2), & if\ \sigma_\Omega^2 \le B_1 \\ 0 & else \end{cases},$$

and (4)

$$h_2(\tilde{z},P_d) = \begin{cases} K_2 P_d \Big/ \sqrt{\|z-\tilde{z}\|}, & if \quad P_d \ge B_2 \\ 0 & else \end{cases}.$$

Here the constants K_1, K_2 are dampening parameters and B_1, B_2 are thresholds to ensure that external forces work only in the direction of points with smooth and reliable initial disparity vectors.

This model of a deformable elastic contour that is subject to external forces is very similar to the snake model (or active contour model) introduced in [5] in the context of image segmentation. Two main differences between our model and snakes can be pointed out: in the snake model the internal forces consist of two terms modeling tension and rigidity of the elastic contour. The former makes the contour resistant to stretching, whereas the latter makes it resistant to bending. If this term is different from zero, then the contour does not develop corners and it becomes second order continuous. This is an undesirable property in our case, because just rectangular regions (windows) are easy to handle in the matching process. The second difference is that the external forces in usual snakes deform the contour in any direction (contracting or expanding the contour), whereas in our model the contour should only be expanded in the direction normal to the tangents of the contour.

A natural way to cause the contour to expand in order to reach lower potential is to convert potential to kinetic energy. If kinetic energy is dissipated by dampening, then the potential decays until an equilibrium is reached. The dynamics of the model defined by (1) and (2) is very similar to the known snakes. The dissipation of kinetic energy is described by the resulting Euler-Lagrange differential equations from the calculus of variations [5]. Solving a system of partial differential equations in order to find the best reference region for each sampling position on the image is, from the computational point of view, a very expensive approach. We prefer to introduce some simplifications to the model, so that the final method presents a moderate complexity. A different way to convert potential to kinetic energy is to expand iteratively the contour from its initial position until the potential becomes zero. To perform this process, we should define a set of increasing contours

$\Omega_0 \subseteq \Omega_1 \subseteq \ldots \subseteq \Omega_i \subseteq \ldots \subseteq \Omega_n$. The potential in a intermediate step is given by $E^p_{\Omega_i} = E_{\Omega_n} - E_{\Omega_i}$, where Ω_n is the final contour, i.e., the contour for which $E^p_{\Omega_n} = 0$. The energy dissipation in each expansion step can be estimated as the decrement of the potential

$$\Delta E^p_{\Omega_i} = \left| E^p_{\Omega_{i-1}} - E^p_{\Omega_i} \right|, \quad 0 < i \le n. \qquad (5)$$

Under the assumption that the external forces fulfill the property r3, it is clear that $E^p_{\Omega_n} = 0$, if $\Delta E^p_{\Omega_n} = 0$. Consequently we can use (5) in each expansion step in order to prove if the potential is zero.

In the discrete image grid, we can start with a 3x3-window as region of minimal size, and then proceed to expand this window step by step in all possible directions. After each expansion we test if the potential of the system is zero by applying a discrete version of (5). The expansion process is stopped when the potential reaches zero or equivalently when the energy difference in (5) is zero. To perform this procedure a discretization of the equations (1), (2), (3) and (4) is necessary. For the sake of efficiency and to keep the computational cost low, we simplify the model by considering only regions of rectangular form, i.e., we constrain the form of the reference region to being rectangular. Note that a discretization of the continuous model for arbitrarily-shaped contours can be done in a similar manner. This simplification of the model also involves the elimination of the internal forces in (1) assuming lack of tension on the window contour, which extremely reduce the algorithmic complexity. Under this assumption the equilibrium is reached, because due to r3 (and consequently (4)) in each expansion step the external forces decrease. This is equivalent to setting $\alpha = 0$ and $\beta = 1$ in (1). Using rectangular regions, only four expansion directions should be considered: right ($\theta_0 = 0$), top ($\theta_1 = \pi/2$), left ($\theta_2 = \pi$) and bottom ($\theta_3 = 3\pi/2$).

4 Computer Experiments

The performance of the method has been verified using synthetic data and natural stereo sequences. Some selected results obtained for the set of synthetic data are shown in Fig. 1. The image at the top-left of Fig. 1 shows the left image of a box in front of a planar textured background. The right image (displayed at the top right) has been generated by deforming the left image according to the disparity pattern shown in the image in the middle-left of Fig. 1. Here increasing brightness corresponds to increasing horizontal disparity values. The vertical component of the disparity is zero for all vectors. The presented images exactly model stereoscopic images taken by ideal virtual cameras with parallel optical axes and with image planes parallel to the planar background and to the frontal face of the box. Gaussian noise has been added independently to left and right stereo images. The image in the middle-right of Fig. 1 shows the initial disparity field estimated by hierarchical block matching. The corresponding reliability field calculated by applying the method introduced in [4] is displayed in the image at the bottom-left (in this figure increasing brightness corresponds to increasing reliability values). The improved disparity field estimated by matching with adaptive window is shown in the image at the bottom-right.

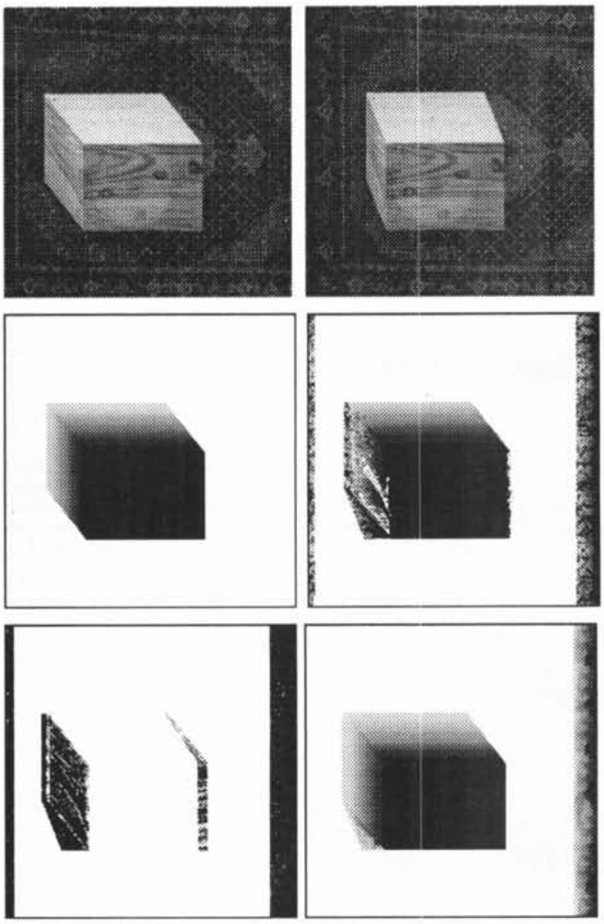

Fig. 1: Results obtained for a set of synthetic data. Stereoscopic scene with known exact disparity map (top), disparity map estimated by applying a hierarchical block-matching approach (middle-right), corresponding reliability values (bottom-left) and disparity map by applying matching with adaptive window (bottom-right).

To examine how the method presented work with real data several experiments have been performed. We report some relevant results obtained for the sequence SAXO. In order to make variations due to depth changes more visible, the vertical component of disparities is neglected in the following representations. Fig. 2 shows results obtained for this sequence. The left first frame of this sequence is shown in the image at the top. The initial disparity field estimated by hierarchical block matching is displayed in the middle. The improved disparities estimated with adaptive window is shown at the bottom. For the sake of clarity only samples along a sparse 8x(d+8) grid are shown (d is the length of the sampled vector).

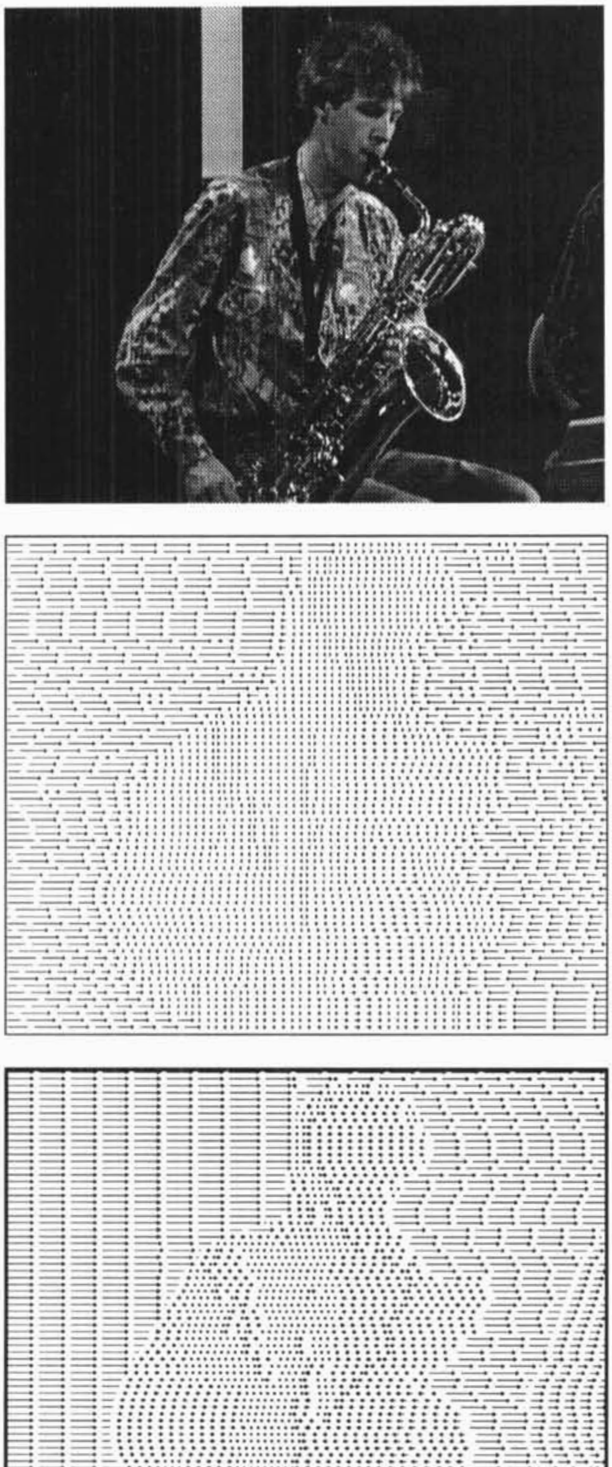

Fig. 2: Results obtained for a natural scene. Original image (top), disparity map estimated by applying a hierarchical block-matching approach (middle), and disparity map by applying matching with adaptive window (bottom).

5 Conclusions

A novel technique for displacement estimation based on block-matching with a local adaptive window is presented in this article. The goal of accurate correspondence estimation preserving discontinuities in the displacement maps due to depth differences or motion between different objects is achieved by applying the proposed method. The model introduced for window adaptation has been designed in order to cope with the problem of obtaining good estimates just within object borders or image areas that become occluded. This approach is based on the principle that sampling positions whose estimated disparity indicates a high reliability have to be more involved in the matching of sampling positions than those whose disparity has a low reliability. The presented technique has been tested using synthetic and natural test data with different degrees of difficulty. The results obtained show that the objective mentioned above can be achieved by applying the introduced method. Although the approaches involve different strategies and two processing steps, the computational complexity of the whole system is relatively low.

References

[1] P. Anandan, "Measuring visual motion from image sequences", *Ph. D. Thesis, University of Massachusetts*, 1987.

[2] M. Bierling, "Displacement estimation by hierarchical blockmatching", *Visual Communications and Image Processing*, SPIE vol. 1001, pp. 942-951, 1988.

[3] W. Hoff and N Ahuja, "Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. PAMI-11, no. 2, pp. 121-136, 1989.

[4] E. Izquierdo, "Stereo matching for enhanced telepresence in 3D-videocommunications", *IEEE Transaction on Circuits and Systems for Video Technology, Special issue on Multimedia Technology, Systems and Applications*, vol. 7, no. 4, pp. 629-643, Aug. 1997.

[5] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models", *Int. J. of Computer Vision*, vol. 1 no. 4, pp. 321-331, Jan. 1998.

[6] M. Levine, D. O'Handley and G. Yagi, "Computer Determination of Depth Maps", *Computer Graphics and Image Processing*, vol. 2, no. 4, pp. 131-150, 1973.

[7] J.-R. Ohm, K. Grüneberg, E. Hendriks, E. Izquierdo M., D. Kalivas, M. Karl, D. Papadimatos, A. Redert, "A realtime hardware system for stereoscopic videoconferencing with viewpoint adaptation", *Image Communication, special issue on 3D technology*, announced for Jan. 1998.