Transactions on Information and Communications Technologies vol 19, © 1997 WIT Press, www.witpress.com, ISSN 1743-3517

Hopfield Network for Stereo Correspondence Using Block-Matching Techniques

Dimitrios Tzovaras and Michael G. Strintzis
Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece
EMail: tzovaras@dion.ee.auth.gr
phone: (+30-31) 996-359, fax: (+30-31) 996-398

Abstract

A neural network based algorithm is presented for solving the stereo vision correspondence problem. The stereo images are divided into blocks and for each block in the left image its corresponding one in the right image is found. The problem is presented as the minimization of a cost function which can be the Lyapunov function of a two-dimensional binary Hopfield neural network. The states of the neurons are updated so as to minimize the cost function. The updating procedure is iterated until the network settles to a stable state. After running the network some of the matched blocks have multiple matches. A post processing procedure is used for finding a single match for every block examined.

1 Introduction

Stereoscopy is useful in, among other applications, the computation of depth information about a scene [1]. The depth information is essential in many applications as robotics, photogrammetry and medical imaging. Computing the displacement or disparity, between two corresponding feature points or

blocks in the left and right images, the three dimensional coordinates of an image point in the scene can be found.
A number of methods for the estimation of displacement vector fields have been proposed in the literature. Both block-based and feature-based techniques are well researched. Block matching techniques may be realized using full (exhaustive) search techniques or by faster, limited search techniques. An efficient block matching algorithm is the hierarchical [2], in which agreement of large blocks is first attained and the block size is subsequently progressively decreased.
The stereo correspondence problem can be solved by :

- Matching every point in the left image with every point in the right image [3].
- Extracting distinct features from each image and try to match them [4].
- Divide each image in blocks and try to match them [5].

The aim of disparity estimation is the matching of corresponding picture elements in simultaneous 2D pictures of the same 3D scene, viewed under different perspective angles. Two of those pictures may be the left and right views of a stereoscopic pair shot by a stereoscopic camera. The disparity vector fields can be used to predict one image of a stereoscopic pair from the other, within a disparity compensated coding scheme. Disparity estimation is crucial in many other applications, such as computation of intermediate views, distance-to-the-camera keyed segmentation for background/foreground mixing, or quality control with depth models.
Sparse disparity fields are used to predict one image of a stereoscopic pair from another, within a coding scheme using disparity compensated prediction [2, 6, 9] or joint motion and disparity compensated prediction [10]. With a multiview display, this allows the observer to watch the scene from varying optical angles. In other applications, the generation of intermediate images is needed even with simple monoscopic displays at the receiver. For example, simulated eye-contact is known to enhance the "telepresence" which is desirable in advanced videoconferencing schemes.
The parallelism and the computational power offered by neural networks has made them an alternative to be effectively used in image processing. In [8] an approach in establishing stereo correspondence between features

of the stereo images using the Hopfield neural network was presented. In the present paper, the Hopfield network is used to establish correspondence between blocks of the stereo images. In the following, the images are divided into $N_B \times N_B$ blocks. Blocks in the left image will be denoted using indices i and j while blocks in the right image using k and l.
The paper is organized as follows. In Section 2 we propose a neural network technique for disparity and depth estimation. In Section 3 the problem of extraction of depth from disparity is examined and a number of techniques are reviewed for solving it. Finally in Section 4 the performance of the proposed scheme is evaluated experimentally and conclusions are drawn in Section 5.

2 Description of the Proposed Technique

The Lyapounov function for a 2D binary (two-state) Hopfield network [7] is given by

$$E = -(1/2)\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ijkl}V_{ik}V_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} I_{ik}V_{jl} \qquad (1)$$

In the case examined, the cost function given below is minimized

$$E = -\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} C_{ijkl}P_{ik}P_{jl} - \sum_{i=1}^{N_l}(1-\sum_{k=1}^{N_r} P_{ik})^2 \sum_{j=1}^{N_l}(1-\sum_{l=1}^{N_r} P_{jl})^2 \qquad (2)$$

In this equation P_{ik} represents a measure of the match between block i in the left image and the block k in the right image. Pik is equal to 1 when a match occurs and to 0 when there is no match between blocks i and k.
The first term in the above equation represents the compatibility of a match between the blocks i and j in the left image and the blocks k and l in the right image. The second and the third term are used to enforce a uniqueness constraint where the probabilities (states of neurons) should add up to 1. Easily,

$$E = -(1/2)\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ijkl} - \delta_{ik} - \delta_{jl})P_{ik}P_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} 2P_{ik} \qquad (3)$$

where δ_{ik} if $i = k$, otherwise 0. The cost function can be Lyapounov function (1) of a Hopfield network with states of the neurons $Vik = P_{ik}$ if the input to each neuron is set to $I_{ik} = 2$. Also, the connection weights between

two neurons must be defined as $T_{ikjl} = (C_{ikjl} - \delta_{ik} - \delta_{jl})$ where C_{ikjl} is the compatibility measure assumed to be sigmoid :

$$C_{ikjl} = \frac{2}{1 + e^{\lambda(X-\theta)}} - 1 \tag{4}$$

$$X = W_1|\Delta d| + W_2|\Delta D| + c|\Delta B_{ik}\Delta B_{jl}| \tag{5}$$

where c is a normalization factor and

$$\Delta B_{ik} = \sum_{r=0}^{N_B-1}\sum_{q=0}^{N_B-1} |L[i_1 + r, j_1 + q] - R[k_1 + p, l_1 + q]| \tag{6}$$

In the above equation L, R are the left and right image respectively and i_1, j_1 and k_1, l_1 are the upper left corner coordinates of blocks i and k, respectively. Δd is the difference in the disparities of the matched blocks i, k and j, l. The other comparison factor ΔD, is the difference between the distance from block i to block j and the distance from block k to block l. The third term represents the L_1 norm of the intensity difference between the corresponding pixels in the left and right image. From the experimental tests the best values for the parameters were $W_1 = 0.2$, $W_2 = 0.2$, $W_3 = 0.6$, $\lambda = 1$ and $\theta = 10$. The largest weight was assigned to W_3 since $\Delta B_{ik}\Delta B_{jl}$ is the most significant term in X for establishing block matching.

The updating rule for the network is given below :

$P_{ik} \rightarrow 0$ if

$$\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ik} - \delta_{jl})P_{jl} + 2 < 0 \tag{7}$$

$P_{ik} \rightarrow 1$ if

$$\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ik} - \delta_{jl})P_{jl} + 2 > 0 \tag{8}$$

No change if

$$\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ik} - \delta_{jl})P_{jl} + 2 = 0 \tag{9}$$

In these equations blocks i are selected randomly from all blocks in the left image. The candidates for correspondence blocks k are selected in a window of dimension 30×2 pixels around the center of block i. All other blocks were assumed to have zero compatibility contributions because they were too far from block i. Blocks j and l are chosen randomly in a region 15×2 pixels around block i in the left and right image, respectively.

3 Depth Estimation from a Pair of Stereoscopic Images

The method for recovering depth information from the dense disparity field that was presented in [6, 10] was used.
It is clear that the depth estimation problem reduces to the disparity estimation problem from a pair of stereoscopic images. In the simple case of parallel stereo camera configuration, the mapping from disparity to depth and reverse, is straightforward, i.e.

$$d = \frac{bf}{z} \tag{10}$$

In this case the epipolar lines are horizontal lines. This observation introduces a constraint to the disparity estimation approach used.
The depth estimation under a more general assumption, of a stereoscopic camera with two converging optical axes is more complicated. Disparity information can be extracted from the depth information using the geometric relationships for a stereo camera with converging optical axes:

$$d_x = f\left(\frac{(x+b/2)*\cos(a/2) - z*\sin(a/2)}{z*\cos(a/2)+(x+b/2)*\sin(a/2)} - \frac{(x-b/2)*\cos(a/2)+z*\sin(a/2)}{z*\cos(a/2)-(x-b/2)*\sin(a/2)}\right), \tag{11}$$

$$d_y = f\left(\frac{y}{z*\cos(a/2)+(x+b/2)*\sin(a/2)} - \frac{y}{z*\cos(a/2)-(x-b/2)*\sin(a/2)}\right), \tag{12}$$

where f is the focal length of the camera, b is the baseline and a is the convergence angle.

The inverse problem, i.e. estimation of depth from disparity, can be solved using least squares techniques as follows.

$$\hat{z} = (AA^T)^{-1}A^T\mathbf{B} = f\frac{a[0]*b[0]+a[1]*b[1]}{a[0]^2+a[1]^2} \tag{13}$$

where

$$\mathbf{A} = \begin{bmatrix} a[0] \\ a[1] \end{bmatrix} = \begin{bmatrix} x_l\left(x_r sin(\theta)+f cos(\theta)\right) - f\left(x_r cos(\theta) - f sin(\theta)\right) \\ y_l\left(x_r sin(\theta)+f cos(\theta)\right) - f y_r \end{bmatrix}. \tag{14}$$

and,

$$\mathbf{B} = \begin{bmatrix} b[0] \\ b[1] \end{bmatrix} = \begin{bmatrix} b\left(f cos(\theta/2) - x_l sin(\theta/2)\right) \\ -b y_l sin(\theta/2) \end{bmatrix}. \tag{15}$$

where f is the focal length of the camera, b is the baseline, and θ is the convergence angle between the coordinate axes of the stereoscopic camera pair. Disparity is used to define the coordinates (x_l, y_l) in the right channel image, using the coordinates of its corresponding point (x_r, y_r) in the left channel image, using

$$d_x = x_l - x_r \qquad d_y = y_l - y_r \tag{16}$$

The accuracy in the depth information is very important in applications such as temporal interpolation and generation of intermediate views. Furthermore, the depth information is very important in object-based coding schemes with 3-D motion estimation and compensation [6].

4 Experimental Results

The stereo images "Sergio" and "Tunnel" [1] were used for the evaluation of the performance of the proposed technique. The original left and right second frames of "Sergio" and "Tunnel" are shown in Figures 1a and 1b and 2a and 2b. Block-based disparity estimation with the proposed technique and with a block size of 8×8 was performed first. The search area for disparity was chosen to be ± 15 and ± 1 pixels for the x and y coordinate respectively. The computed x-component of the block-based estimated disparity field is shown in Fig. 1c and 2c.

The network was initialized setting $P_{ik} = 1$ for all i and k blocks, and usually settled down after less than 20-30 iterations. The network settles down when it is in its minimum energy. However, local minima can not always be avoided. The updating procedure is iterated until the network reaches a stable state. The Hopfield network presented in [8] for feature matching needed more than 1000 iterations to settle down. The present network converges much faster. The drawback of using the Hopfield network is that after running it, some of the matched blocks still have multiple matches. To find a single match the techniques presented in [8] for exploiting the noisy y disparity were used. The disparity vector field computed using the present network, approximate the one provided by the common full search method. The depth map can also be computed using the techniques described in Section . The computed depth is computed from a dense disparity field

[1] The image sequence "Tunnel" was shot by CCETT for the purposes of the RACE DISTIMA and the ACTS PANORAMA projects.

and has the same resolution with the original image. Figures 1d and 2d show the depth map corresponding to "Sergio" and "Tunnel", respectively, quantized into 256 levels. These depth maps were computed from a dense disparity field estimated using the Hopfield network described above and a a block consisting of one pixel.

4 Conclusions

A neural network based algorithm was presented for solving the stereo vision correspondence problem. The stereo images were divided into blocks and for each block in the left image its corresponding one in the right image was found. The problem was presented as the minimization of a cost function which can be the Lyapunov function of a two-dimensional binary Hopfield neural network. The states of the neurons were updated so as to minimize the cost function. The updating procedure was iterated until the network settles to a stable state. After running the network some of the matched blocks had multiple matches. A post processing procedure is used for finding a single match for every block examined.

5 Acknowledgement

This work is based on work supported by the ACTS PANORAMA project.

Keywords : Disparity estimation, Hopfield neural network, block matching

References

[1] U. R. Dhond and J. k. Aggarwal, "Structure From Stereo - A Review", *IEEE Trans. on Systems, Man and Cybernetics,* Vol. 19, No. 6, pp. 1489-1510, Nov. 1989.

[2] D. Tzovaras, M. G. Strintzis, and H. Sahinoglou, "Evaluation of Multiresolution Techniques for Motion and Disparity Estimation," *Signal Processing : Image Communication*, vol. 6, pp. 59–67, Mar. 1994.

[3] D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity", *Science,* Vol. 194, pp. 283-287, 1976.

[4] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm", *IEEE Trans. on PAMI,* Vol. PAMI-7, pp. 17-34, 1985.

[5] H. G. Mussman, P. Pirsch and H. -J. Grallert, "Advances in Picture Coding," *Proc of the IEEE,* Vol. 73, pp. 523-548, Apr. 1985.

[6] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-Based Coding of Stereo Image Sequences using Joint 3-D Motion/Disparity Compensation," *IEEE Trans. on Circuits and Systems for Video Technology,* Apr. 1996.

[7] J. Hopfield, "Neural Networks and Physical Systems With Emergent Collective Computational Abilities", *Proc. Nat. Acad. Science,* Vol. 79, pp. 2554-2558, May 1984.

[8] N. M. Nasrabadi and C. Y. Choo, "Hopfield Network for Stereo Correspondence", *IEEE Trans. on Neural Networks,* Vol. 3, No. 1, Jan. 1992.

[9] M. Ziegler, "Digital Stereoscopic Imaging and Application, A Way Towards New Dimensions, The RACE II project DISTIMA," in *IEE Colloq. on Stereoscopic Television*, (London), 1992.

[10] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Disparity Field and Depth Map Coding for Multiview 3D Image Generation," *Signal Processing (Image Communication)*, accepted for publication.

[11] M. G. Strintzis, D. Tzovaras, and N. Grammalidis, "Depth Map and Disparity Field Coding for the Communication of Multiview Images," in *Proc. 35th Int'l Conf. on Digital Signal Processing '95*, (Limassol, Cyprus), June 1995.

[12] N. Grammalidis, S. Malassiotis, D. Tzovaras, and M. G. Strintzis, "Stereo image sequence coding based on 3D motion estimation and compensation," *Signal Processing : Image Communication*, vol. 7, No. 2, pp. 129–145, Aug. 1995.

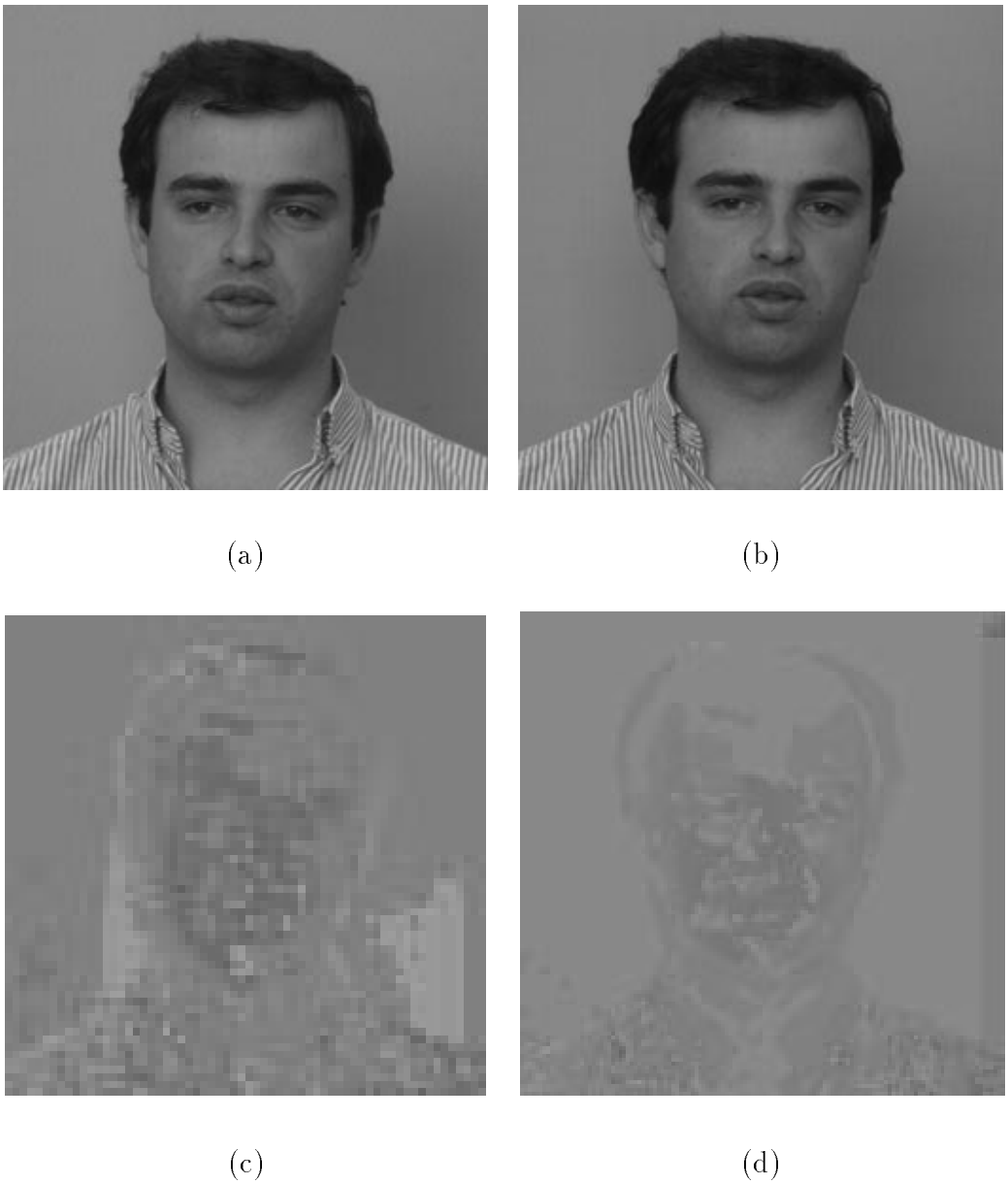

Figure 1: (a) Original left channel image "Sergio" (frame 2). (b) Original right channel image "Sergio" (frame 2). (c) Block-based estimate of disparity. (d) Pixel-based estimate of depth.

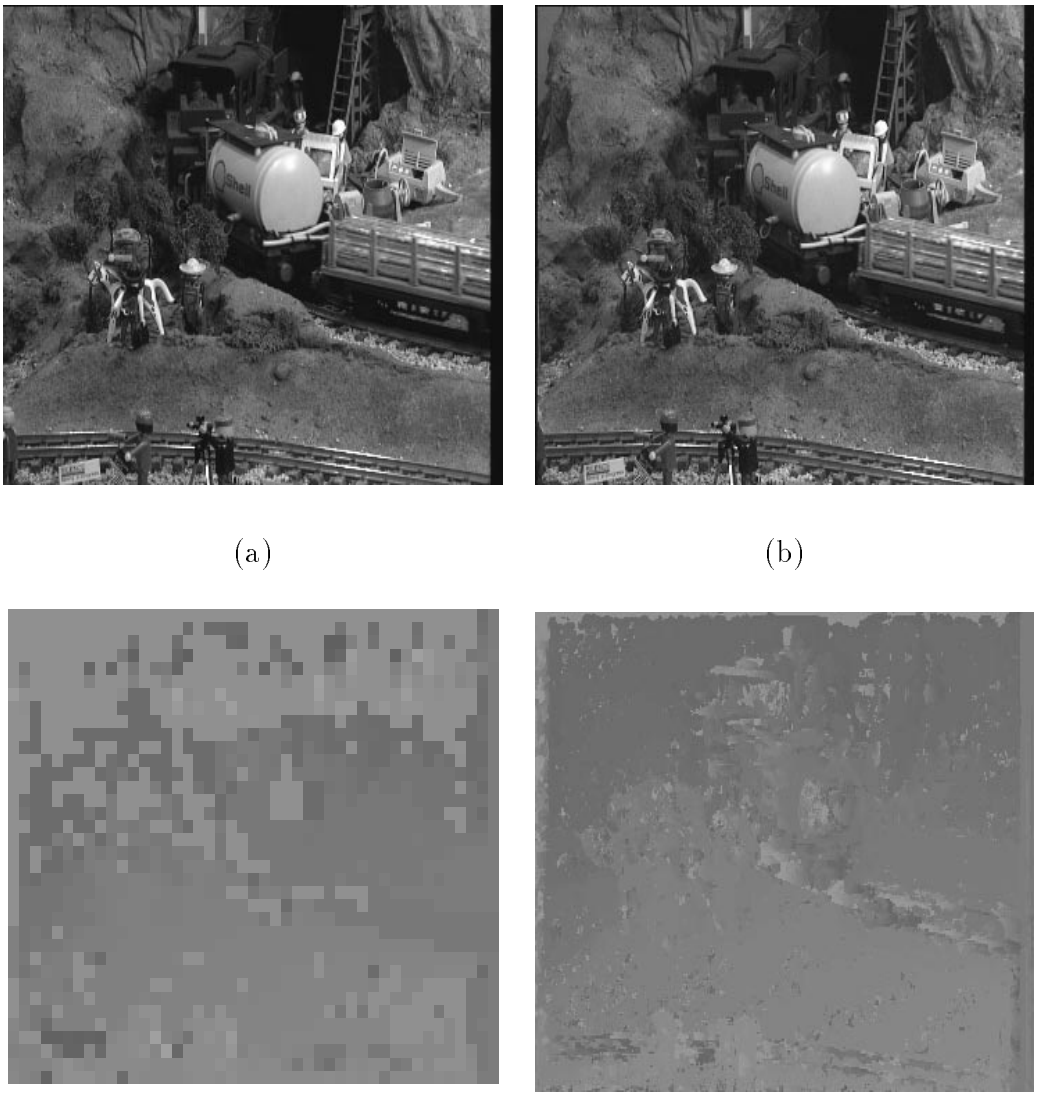

Figure 2: (a) Original left channel image "Tunnel" (frame 2). (b) Original right channel image "Tunnel" (frame 2). (c) Block-based estimate of disparity. (d) Pixel-based estimate of depth.