Motion Correspondence Using a Neural Network

G. H. Sarigianidis[1] and D. Pycock
School of Electrical and Electronic Engineering
The University of Birmingham
E-mail:sarigiag@eee.bham.ac.uk

Abstract

Identifying corresponding features in an image sequence is an important issue in motion analysis. We present a solution based on the assumption of smooth motion using point features. Local constraints are used to make the method robust against occlusion and imperfect feature extraction. A global cost function is defined which is minimised by a mapping of feature points onto a 2-D Hopfield neural network. Three variants of the Hopfield network model are considered. Results obtained using synthetic and natural image data show that this method is robust against occlusion and poor feature detection.

1 Introduction

Time-varying images may result from camera motion, the motion of objects in the scene and variations in scene illumination. The changes produced provide a rich set of cues for understanding a scene, its structure, and the motion of the objects in the image. The goals of dynamic scene analysis are motion estimation and object structure recovery. Applications include the detection, recognition and tracking of moving objects, robotic vision and motion-compensated coding.

There are two principal approaches to dynamic scene analysis: pixel-based and feature-based methods. In a pixel-based method the displacement of each pixel in the image is estimated from local image intensity changes, as in *optical flow*[7]. This method is usually used when the displacement field is small. Feature-based methods involve three steps. First, a set of relatively sparse, but highly descriptive, features are located in each frame. Next, corresponding features in two successive frames are identified (*correspondence problem*). Finally, the motion parameters and object structure are derived from the correspondences found. In this paper we present a method for solving the correspondence problem.

The purpose of a motion correspondence method[1][2][12][14] is to match a set of identifiable physical features (such as points, edges or regions) over a frame sequence. In the most popular methods groupings of image features are used with domain and feature related constraints (e.g. rigidity, smoothness of motion, or common motion constraints) to guide matching. These approaches can be divided into those based on matching isolated features and those based on finding a global match for all the features. The first approach uses similarity measures such as intensity statistics[1] and average computed speed[2] to find a match for each

[1] The authors acknowledge the support of SERC and SHELL U.K.

BMVC 1993 doi:10.5244/C.7.65

feature. Its major disadvantage is that a feature in one frame may match to several in the next. The second approach finds the optimal global match between features in two consecutive frames. All possible combinations of potential matches are constructed. A cost function is defined based on the smoothness of motion[12][14] or the rigidity of the feature sets under motion[15]. The objective of the method is to find the minimum-cost combination of feature matches.

This paper describes an extension of global feature matching to incorporate smoothness of motion constraints in the form of a maximum change in the position and velocity of feature points between frames. The robustness of the matching is enhanced by a concept of hypothetical feature points. The correspondence problem is formulated as the minimisation of a cost function by a Hopfield-Tank network model[4].

The Hopfield neural network model

The Hopfield network is a recurrent, one-layer network constructed by inter-connecting a large number of neurons. Each neuron is described by its current state u_i and output V_i; u_i and V_i are usually related through a monotonic increasing output function $V_i=g(u_i)$. A non-linear $g(u_i)$ limits possible values of V_i to the range -1 to +1 or 0 to 1. It is frequently a step or a sigmoid function. The output of each neuron i is fed to the input of every other neuron j with a connection strength T_{ij}. Each neuron has an offset bias i_i^b associated with each input. The state u_i of neuron i is updated as a function of the total input to that neuron.

In the discrete version[5] $g(u_i)$ is a step function and the state of a neuron at time $t+1$ is related to output of the other neurons at time t by:

$$u_i^{t+1} = \sum_{j \neq i} T_{ij} V_j^t + i_i^b \qquad (1)$$

The stable states of the network[5] are associated with the local minima of the Liapunov function:

$$E = -\frac{1}{2} \sum_i \sum_{j \neq i} T_{ij} V_i V_j - \sum_i i_i^b V_i \qquad (2)$$

In the continuous version the dynamics of the network are given by:

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j \neq i} T_{ij} V_j + i_i^b \qquad 3(a) \qquad g(u_i) = \frac{1}{1+\exp\left(u_i / \lambda\right)} \qquad 3(b)$$

The continuous network has stable states[4][6] which are local minima of the Liapunov function:

$$E = -\frac{1}{2} \sum_i \sum_{j \neq i} T_{ij} V_i V_j - \sum_i i_i^b V_i + \frac{1}{\tau} \int_0^{V_i} g^{-1}(V) dV \qquad (4)$$

In equation (3b) large values of λ increase the ability of the network to reach a global minimum but produce a poorly differentiated output. Low values of λ produce a well differentiated output but the performance becomes similar to that of the discrete version. Gain annealing is an adaptation of the continuous Hopfield model in which λ is decreased in successive iterations. Therefore the network has

both an enhanced ability to find a global minimum and provides a well differentiated output. A logarithmic cooling scheme guarantees convergence of a network[4].

The Hopfield network has been proposed as an optimisation machine for solving problems expressible as constrained minimisation of a cost function[4]:

$$E^{cost} = -\frac{1}{2}\sum_i \sum_{j \neq i} T_{ij}^{cost} V_i V_j - \sum_i i_i^{cost} V_i \quad (V_i \in \{0,1\}) \tag{5}$$

This cost function is related to the Liapunov function of a network (as in either (2) or (4) with $\tau \to \infty$). The Hopfield network has been successful in many applications[4][9][11][16]. Zhu et al. [16] present a Hopfield network to find corresponding points in two 3-D synthetic images with the constraint of rigidity and points that are all from one object.

A new application of the Hopfield network for establishing motion correspondence in 2-D projections from 3-D scenes is presented here. Multiple objects are admitted in the analysis with the constraint of smooth motion.

2 Motion Correspondence

A motion correspondence method selects and extracts a set of image features and the establishes feature correspondence in successive frames. Corners are good features for motion correspondence because they are small and located in areas of high image intensity variance. Most existing corner detectors fail to produce a consistent set of features[13] because of occlusion, poor lighting and object motion. Rather than trying to identify a perfect corner detector the simple corner detector proposed by Harris and Stephens[3] has been used. We have sought to accommodate its limitations in the design of a matching procedure.

Feature correspondence

The smoothness of motion assumption first introduced by Jenkin [9] for motion stereo is used by Rangarajan and Shah [12] and Sethi and Jain[14] for monocular images. Jenkin argued that the 3-D location of a given point and its velocity vector from frame to frame remain relatively unchanged. Thus it is reasonable to assume that any physical object (rigid and non-rigid) follows a smooth trajectory and covers a small distance in the time between frames.

In a sequence of m frames, $f^1, f^2, \cdots, f^m$ (where f^t consists of a set of feature points), the ith point of frame t is represented by P_i^t, the vector of its two-dimensional co-ordinates. It is assumed that the trajectory of any point in the stationary image plane is smooth[13]. Therefore, considering three frames at a time, correspondences can be found by minimising the displacement and velocity change that a mapping of potential matches produces. This can be achieved using a function whose value is related to the change in velocity and displacement of a point in two successive frames.

In practice the number of feature points can change from frame to frame due to occlusion or poor feature extraction. Even when the same number of points are identified in each frame they do not necessarily represent the same set of features.

Consequently it is important to allow for the possibility of obtaining incomplete trajectories. Constraints which limit local change in velocity and displacement prevent inappropriate correspondences being identified and help to maintain the smoothness of any trajectory found. A more realistic formulation of the problem is to seek the maximal set of complete or partially complete trajectories that minimises the sum of local velocity changes and displacements. The constraints applied are that the local velocity change for a point does not exceed some value $a_{\max}$ and the displacement of any point between two successive frames is less than some value $d_{\max}$. Local constraints limit the acceptable location of a feature point, given its location in the two previous frames.

Hypothetical points are introduced in the establishment of correspondence to accommodate incomplete trajectories. Frame to frame correspondences are computed using the displacement function:

$$d\left(P_i^t, P_{\phi(i)}^{t+1}\right) = \begin{cases} \left\| \overline{P_i^t P_{\phi(i)}^{t+1}} \right\| & \text{if both points are true feature points} \\ d_{\max} & \text{otherwise} \end{cases} \tag{6}$$

where $\overline{X_i^k X_j^{k+1}}$ represents a vector from point i in frame k to point j in frame k+1, $\|X\|$ denotes the magnitude of vector X and $\phi(i)$ is the point to which i is mapped in the frame k+1 under the mapping ϕ. This displacement function constrains hypothetical points to displacement $d_{\max}$. Changes in local velocity are computed using:

$$\mathrm{a}\left(P_i^{t-1}, P_{\Phi^{t-1}(i)}^{t}, P_{\phi(i)}^{t+1}\right) = \begin{cases} \left\| \overline{P_i^{t-1} P_{\Phi^{t-1}(i)}^{t}} - \overline{P_i^{t} P_{\phi(i)}^{t+1}} \right\| & \text{if all points are true feature points} \\ a_{\max} & \text{otherwise} \end{cases} \tag{7}$$

where $\Phi^{t-1}(i)$ is the point in frame t corresponding to feature point i in frame t-1. This function constrains hypothetical points to velocity change $a_{\max}$. Thus the introduction of hypothetical points in the preceding or subsequent frame is discouraged. The domain of functions $d()$, $a()$is $[0,+\infty)$. Two functions $g1()$ and $g2()$ are defined to map $d()$ and $a()$ into the domain (-1,1). This is achieved using functions of the form:

$$f(x) = \frac{2}{1+\exp\left(\lambda(x-\theta)\right)} - 1 \tag{8}$$

where x is $d()$ or $a()$ and θ is $d_{\max}$ or $a_{\max}$ for $g1()$ and $g2()$ respectively.

The following net cost function, $g()$, is proposed:

$$\mathrm{g}\left(P_i^{t-1}, P_{\Phi^{t-1}(i)}^{t}, P_{\phi(i)}^{t+1}\right) = \mathrm{g1}\left(P_{\Phi^{t-1}(i)}^{t}, P_{\phi(i)}^{t+1}\right) + \mathrm{g2}\left(P_i^{t-1}, P_{\Phi^{t-1}(i)}^{t}, P_{\phi(i)}^{t+1}\right) \tag{9}$$

The likelihood for each isomorphic mapping ϕ between feature points in frames f^t and f^{t+1} is:

$$\Pi\left(\phi \mid f^{t-1}, f^t, f^{t+1}\right) = -\sum_{d=1}^{N} g\left(P_d^{t-1}, P_{\Phi^{t-1}(d)}^{t}, P_{\phi(d)}^{t+1}\right) \tag{10}$$

Given three frames f^{t-1}, f^t and f^{t+1} with known correspondence Φ^{t-1}, where the number of feature points in frames f^t and f^{t+1} is N_1 and N_2 respectively, the procedure to establish motion correspondence is defined in Figure 1.

Step 1: If $N_1 > N_2$ add to frame f^{t+1} N_2-N_1 hypothetical points.
If $N_1 < N_2$ add to frames f^{t-1} and f^t N_1 -N_2 hypothetical points.

Step 2: Determine the correspondence Φ^t using :

$$\Pi\left(\Phi^t \mid f^{t-1}, f^t, f^{t+1}\right) = \min_{\phi \in \Phi} \Pi\left(\phi \mid f^{t-1}, f^t, f^{t+1}\right) \tag{11}$$

where Φ is the space of all isomorphic mappings.

Step 3: Verify that the correspondences established by Φ^t comply with the displacement and velocity change constraints, using the rule:

$$d\left(P_i^t, P_{\phi(i)}^{t+1}\right) \le d_{\max} \text{ and } a\left(P_i^{t-1}, P_{\Phi^{t-1}(i)}^t, P_{\phi(i)}^{t+1}\right) \le a_{\max} \rightarrow \begin{cases} \text{True} & \text{correct match} \\ \text{False} & \text{incorrect match} \end{cases} \tag{12}$$

Figure 1. Algorithm for motion correspondence

Feature points having incorrect correspondences together with those that match with hypothetical points are declared as feature points without correspondences. In the first two frames of a sequence correspondences are established by using displacement information only and setting velocity changes to the maximum value a_{max} for all feature points.

Hopfield neural network solution

To find the correspondences, Φ^t in Step 2, a global cost function incorporating each constraint is formed, such that non-isomorphic mappings are penalised, and mappings which minimise (10) are favoured.

Denote by p_{ik} a match between the ith feature point in frame f^t and the kth feature point in the frame f^{t+1}. Let p_{ik} be 1 if the two features are matched under a mapping ϕ and 0 otherwise. Motion correspondence can be determined by minimising the cost function:

$$E = A \sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{l\neq k}^{N} p_{ik}\,p_{il} + B \sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{j\neq i}^{N} p_{ik}\,p_{jk} + C\left[\sum_{i=1}^{N}\left(1-\sum_{k=1}^{N} p_{ik}\right)^2 + \sum_{k=1}^{N}\left(1-\sum_{i=1}^{N} p_{ik}\right)^2\right] - D\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} c_{ikjl}\,p_{ik}\,p_{jl} \tag{13}$$

The terms associated with *A*, *B* and *C* in the above equation penalise mappings between feature points in the two frames that are not isomorphic. The term associated with *A* is zero if a feature point of f^t matches with only one feature point of f^{t+1}, otherwise a penalty is imposed in *E*. Similarly, the term associated with *B* is zero if a feature point of f^{t+1} matches with only one feature point of f^t. The term associated with C reinforces the uniqueness constraint that each feature point in one frame can match with only one feature point in the other. Finally, the term associated with *D* represents the degree of smoothness of the trajectories established with a match between a pair of points (i,j) in f^t and a pair of points (k,l) in f^{t+1}. The smoothness measure, c_{ikjl}, takes account of both the displacement and velocity changes:

$$c_{ikjl} = \frac{1}{2}\left[g\left(P_i^{t-1}, P_{\Phi^{t-1}(i)}^t, P_k^{t+1}\right) + g\left(P_j^{t-1}, P_{\Phi^{t-1}(j)}^t, P_l^{t+1}\right)\right] \tag{14}$$

The domain of the smoothness measure is $(-1, g_{max}]$. The value of this measure varies smoothly between $+g_{max}$, which indicates correct matches for the two feature point pairs *(i,k)* and *(j,l)*, to asymptotically approach -1 when both feature point pairs do not match.

A two-dimensional Hopfield network minimises the cost function (13) and thereby finds the correspondence between the feature points in frames f^t and f^{t+1}. The network is an NxN array of neurons, where N is the number of feature points in each frame. The rows of the network represent the feature points in f^t and the columns the feature points in f^{t+1}. The 'on' state of each neuron represents a match between a feature point in f^t and one in f^{t+1}. The cost function is equivalent to the Lyapunov function of a Hopfield network (2)[13] with unit states of $V_{ik} = p_{ik}$ and $V_{jl} = p_{jl}$, and an input to each unit, $I_{ik} = 2C$. The connection weight between two units is defined as:

$$T_{ikjl} = \left[D\, c_{ikjl} - A\, \delta_{ij}(1-\delta_{kl}) - B\delta_{kl}(1-\delta_{ij}) - C(\delta_{ij}+\delta_{kl}) \right] \qquad (15)$$

where δ_{ij} is the Kronecker delta. The connections T_{ikjl} are symmetric that is; $T_{ikjl} = T_{jlik}$ and the self-feedback T_{ikik} to each unit is zero. The output V_{ik},of each unit, represents the degree of match between the feature point *i* of frame f^t and the feature point *k* of frame f^{t+1}.

3 Results

Results are presented for experiments on synthetic and natural images. Assorted motions and multiple frames are considered. Two criteria are used as measures of efficiency. The first, Incorrect Correspondence Ratio(ICR), is defined as:

$$\text{ICR} = \frac{\text{total number of incorrect correspondence}}{\text{total number of correspondence}} \text{x } 100\ \% \qquad (16)$$

The second is the distortion measure proposed in[12]. The trajectory set established by various methods over the same data may have a similar cost and ICR without being identical. The distortion measure indicates the amount of deviation of a trajectory set from the true trajectory. It is defined as the Euclidean distance between the point and that of the true trajectory. The distortion measure for a trajectory set is the sum of the distortion measure of each trajectory point.

Three variants of the Hopfield network, the discrete, the continuous, and the continuous with gain annealing, have been considered with respect to synthetic image data and their performance compared with that of the Rangarajan and Shah method[12]. For natural images, results were obtained using the gain annealing Hopfield network. The cost function parameters in (13) were A=B=5, C=1.5 and D=0.8. The local maximum displacement and change in velocity parameters were d_{max}= 15 and a_{max}=10.

Synthetic image data

In the first experiment ten frames of synthetic objects with 5,10,15,25, and 30 corners were generated. In each case the objects in the first frame sequence were translated only while in the second frame sequence they were translated, rotated

and scaled. In all experiments the corners were identified manually and there was no occlusion. Each frame of a sequence contains the same number of feature points. To make the sequence more realistic the co-ordinates of each feature point were randomly perturbed by ±5 %. The results of applying each model and the Rangarajan and Shah method are shown in Figure 2 where the problem size represents the number of corners in each frame. The gain annealing outperforms both the other implementations and the Rangarajan and Shah method. This results from the enhanced optimisation achieved by the use of an annealing schedule. The performance of the continuous model is almost the same as that of the Rangarajan and Shah method while the performance of the discrete model is the poorest.

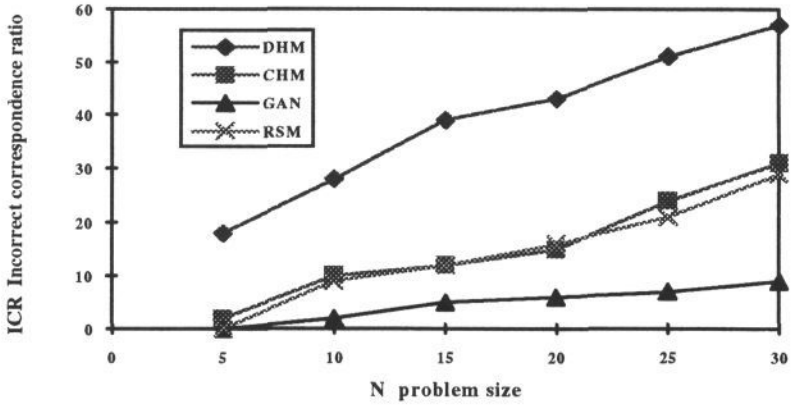


Figure 2. The performance of the proposed method using Discrete (DHM), Continuous (CHM), and Gain annealing (GAN) Hopfield Model and the performance of the Rangarajan and Shah method (RSM).

To investigate the behaviour of the method described in this paper further a comparison was made using the synthetic data of four points in five frames shown in Figure 3[12].

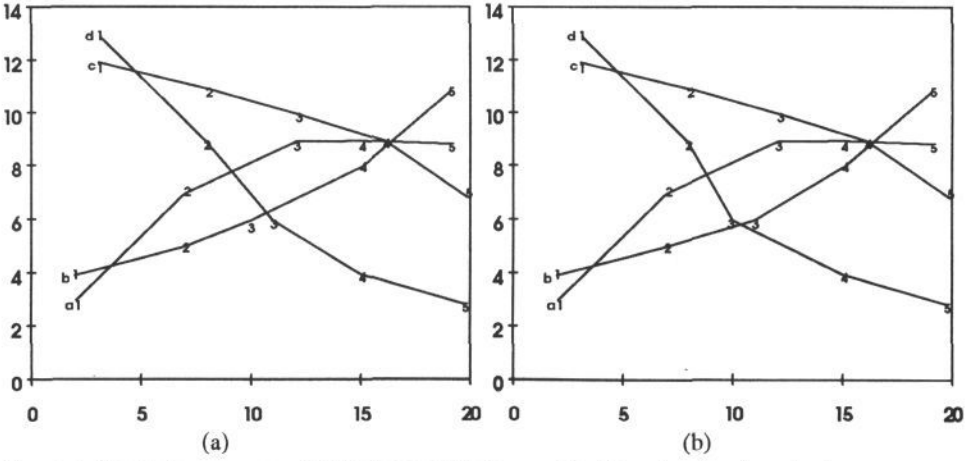
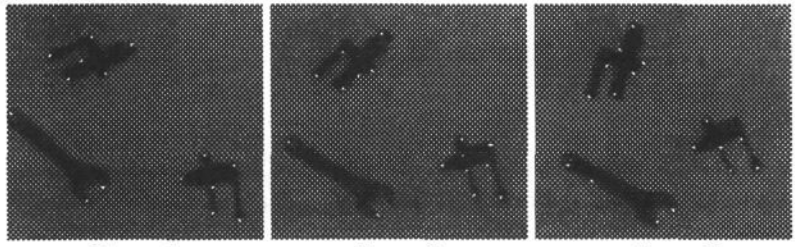


Figure 3. Synthetic sequence: (a) Minimum trajectory set for Neural network method, (b) Minimum trajectory set for the Rangarajan and Shah method.

Each point in Figure 3 is marked with a number corresponding to the frame in which this point appears. The trajectories are shown by lines in Figures 3(a) and (b). Figures 3(a) and (b) show the trajectory set found by the proposed method and the Rangarajan and Shah method respectively. The proposed method finds the optimum trajectory set (the same found by an exhaustive search) while Rangarajan and Shah method finds a trajectory set in which the wrong corresponding points are assigned to trajectories *b* and *d* in the third frame. The distortion measure of the optimum trajectory set found by the proposed method is 0, for the optimal trajectory set found by the Rangarajan and Shah method it is 2 and for the optimal trajectory set found by Setchi and Jain method it is 12 [12].

Natural image data

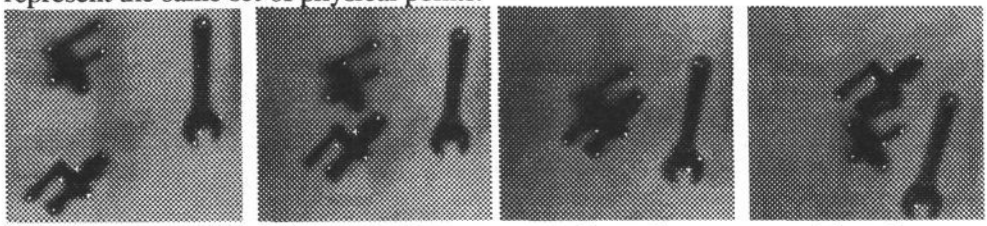
In this section the results of applying the method in two frame sequences are reported. In the *tools 1 sequence,* shown in Figure 4, the motion of three engineering tools in a six frame sequence is considered. The tools are a tap, a wrench and a clamp. In this sequence the tap is rotated while the two other tools are translated. Although there is no occlusion in the sequence the number of feature points detected in the six frames is 15, 14, 15, 13, 17, and 12. This is an effect of non-ideal illumination.



Frame 1 Frame 3 Frame 5

Figure 4. *Tools 1 sequence*. The identified corners in each frame are superimposed.

The second example is an eight frame sequence of the engineering tools where all the tools are translated in different directions as illustrated in Figure 5. The tap and the engineering clamp approach each other, cause occlusion (in frames 4,5,6,7) then move apart. Some feature points appear in some frames and disappear in subsequent frames due to poor feature detection. As a result the number of corners in each of the eight frames is 15, 12, 12, 11, 11, 11 and 15. Although in some frames the number of corners does not change, the points detected do not always represent the same set of physical points.



Frame 1 Frame 3 Frame 5 Frame 7

Figure 5. *Tools 2 sequence*. The identified corners in each frame are superimposed.

Figure 6(a) shows the trajectories obtained for the *Tools 1 sequence.* Nine corner points are present throughout the sequence yielding nine complete trajectories. Seven incomplete trajectories have been established which are the result of poor feature point detection. Finally, there are six points that appear only once in the frame sequence (one in frame 1, one in frame 4 and four in frame 5).

Figure 6(b) shows the established trajectories for the *Tools 2 sequence.* There are two complete trajectories for two points that appear in all the frames of this sequence and twenty one incomplete trajectories, due to either occlusion or poor feature detection. Fifteen points appear in one frame only. The results of the above experiments were verified visually and the ICR of 3.8 and 5.7 were found for the *Tools 1* and the *Tools 2* sequences respectively. The overall performance as measured by ICR is high even for the *Tools 2 sequence* affected heavily by occlusion.

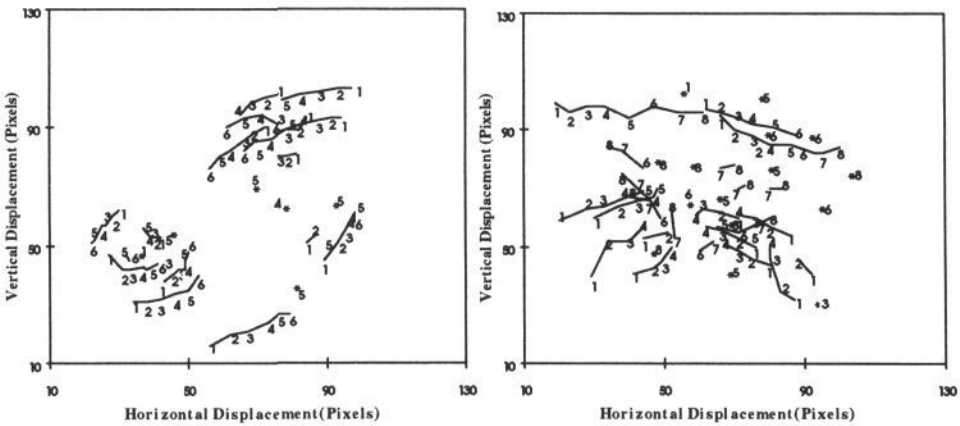


Figure 6. (a) Trajectories obtained for the *Tools 1 sequence.* (b)Trajectories established for the *Tools 2 sequence.* Points without correspondences are marked by *.

4 Conclusions

In this paper, motion correspondence has been formulated as optimisation of a global cost function using a smoothness of motion constraint to establish correspondence. Local limitations on the allowable displacement and change in velocity make the method robust against occlusion and poor feature detection. The cost function is mapped to a Hopfield neural network for minimisation. Three variants of the Hopfield network have been implemented and compared; the discrete, the continuous and the gain annealing versions. The gain annealing version has been demonstrated to perform better than the other two and the continuous version has been shown to perform better than the discrete version. The advantages of using the Hopfield model for motion correspondence are high accuracy, ease of use and potential for parallel implementation.

Future work should investigate the performance of this approach in situations which do not comply with the smoothness of motion assumption. In such cases it may be advantageous to include information about the spatial arrangement of

feature points in different frames. Such information will give rise to a non-quadratic cost function and the need to use higher-order Hopfield network models.

5 References

1. **J K Aggarwal, L S Davis and W N Martin,** Correspondence processes in dynamic scene analysis, *Proc. IEEE,* **69**,1981, 562-571.
2. **S T Barnard and W B Thompson,** Disparity analysis of images, *IEEE Trans. Pattern Anal. Mach. Intelligence,* **PAMI-2**, 1980, 333-340.
3. **C Harris and M Stephens,** A combined corner and edge detector, *Proc. Alvey Vision Conf.*, 1988, 147-151.
4. **J J Hopfield and D W Tank,** Neural computation of decisions in optimization problems, *Biological Cybernetics,* **52**, 1985, 141-152.
5. **J J Hopfield,** Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Academy of Sciences USA*, **79**, 1982, 2554-2558.
6. **J J Hopfield,** Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Academy of Sciences USA,* **81**, 1984, 3088-3092.
7. **B K P Horn and B G Schunk,** Determining optical flow, *Artificial Intelligence*, **17**, 1981, 185-203.
8. **M Jenkin,** Tracking three dimensional moving light displays, *Proceedings Workshop Motion: Representation Contr. Toronto,* 1983, 66-70.
9. **S Z Li** Toward 3D vision from range images: An optimization framework and parallel networks, *CVGIP: Image Understanding,* **55**, 1992, 231-260.
10. **H Moravec,** *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover,* Tech Report CMU-RI-TR-3, Garnegie-Mellon University, Robotics Institute, 1980.
11. **N M Nasrababi and C Y Choo,** Hopfield network for stereo correspondence, *IEEE Trans. on Neural Networks,* **3**, 1, 1992, 5-13.
12. **K Rangarajan and M Shah,** Establishing motion correspondence, *CVGIP: Image Understanding,* **54**, 1, 1991, 56-73.
13. **G H Sarigianidis,** *Motion analysis using neural networks,* M.Phil. dissertation, University of Birmingham, Birmingham U.K., 1992.
14. **I K Sethi, and R Jain,** Finding trajectories of feature points in a monocular image sequence, *IEEE Trans. Pattern Anal. Mach. Intelligence,* **PAMI-9**, 1987, 56-73.
15. **S Ullman,** *The Interpretation of Visual Motion,* Cambridge, MIT Press, 1979.
16. **P Y Zhu, T Kasvand and A Krzyzak,** Motion estimation based on point correspondences using Neural Network, *IJCNN IEEE Int. Conf. on Neural Networks,* Vol II, 1990, 869-874.