ELSEVIER

Pattern Recognition Letters 23 (2002) 253–260

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Robust evidence-based object tracking

Pelopidas Lappas, John N. Carter *, Robert I. Damper

Image, Speech and Intelligent Systems (ISIS) Group, Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

Received 9 January 2001; received in revised form 22 May 2001

Abstract

We extend the velocity Hough transform (VHT) for tracking objects with arbitrary velocity by finding an optimal, smooth trajectory that maximises its associated energy. Optimisation is achieved by temporal dynamic programming (DP). Tracking in noise is much superior to the standard Hough transform (SHT). © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Hough transform; Motion tracking; Dynamic programming

1. Introduction

Motion tracking is an important task in computer vision. A new technique for tracking of parametric objects is described that extends the velocity Hough transform (VHT) to cater for arbitrary motion. Like the VHT, the new technique processes the whole image sequence, gathering global evidence of motion and structure. However, we do not assume constant linear velocity but rather allow arbitrary velocity. The method tries to find a smooth trajectory in the parameter space with maximum energy, where the latter incorporates both the structure of the moving object and the smoothness of motion. Optimisation is effected using a time-delay dynamic programming (DP) algorithm.

The two main approaches for motion estimation are optical flow and feature-based techniques. The latter is advantageous in cases of illuminance change or when the optical flow is large. The choice of features for a vision system is very important. No such system can well work unless good features can be identified and tracked from frame to frame. According to our choice, we can have either too many or too few features. Another problem is that existing methods cannot easily distinguish moving and static features. A common assumption in many approaches is that the tracking or correspondence problem is solved: accordingly, only the selection of features and/or the representation of the object are considered. Selecting appropriate features and computing the correspondence between points in a sequence of frames are both proven to be difficult problems.

The standard Hough transform (SHT) is known for its robust performance in noisy environments

* Correspondiong author. Tel.: +44-2380592405; fax: +44-2380594498.

E-mail addresses: pl99r@ecs.soton.ac.uk (P. Lappas), jnc@ecs.soton.ac.uk (J.N. Carter), rid@ecs.soton.ac.uk (R.I. Damper).

PII: S0167-8655(01)00109-X

and in situations of occlusion (Illingworth and Kittler, 1988). The VHT proposed by Nash et al. (1997) takes advantage of temporal and structural information simultaneously, by incorporating motion in the evidence-gathering process of the Hough transform, enabling global analysis of the temporal image sequence. In its simplest form it requires constant linear velocity, which limits its application.

In the remainder of this paper, Section 2 presents in detail the new tracking algorithm. Performance under noise is investigated in Section 3, including a comparison with other evidence-based methods, and Section 4 concludes.

2. Object tracking algorithm

Here, we explain the proposed algorithm for object detection in a long image sequence. Initially, we gather evidence of structure on a frame-by-frame basis and, incorporating the constraint of smoothness of motion, we then try to find an optimal, smooth trajectory that is supported by the data.

2.1. Arbitrary motion and assumptions

The following assumptions are made for the simplification of the problem. The image sequence is captured from a single camera, far enough from the moving object that we do not need to take scaling factors into account. The object is rigid undergoing arbitrary but smooth translational and rotational motion. We demonstrate our algorithm with respect to a moving circle, but the method can easily be adapted for any parametric and arbitrary nonparametric shape (Aguado et al., 1998). Most object-tracking algorithms consider information within a single image frame: relatively little work focuses on global analysis of a temporal image sequence, as used here. Broida and Chellappa (1986) have applied the iterated Kalman filter to motion estimation with a long sequence of noisy images. Shariat and Price (1990) were among the first to exploit the time flow information from three or more frames in a sequence. Using long sequences, a motion–estimation system tolerates noise and distortions better because the global analysis exploits both temporal and spatial information. Hence, the analysis is tolerant of instances where a feature is missing or corrupted in some frames. The main disadvantage is that use of a long image sequence increases complexity and computation cost.

2.2. Global structure evidence gathering

Considering the whole image sequence as a three-dimensional process $I(x,y,t)$ with spatial variables (x,y) and where t represents the time index of a frame, we can transform the data sequence into a parameter space $P(u,v,a_1,a_2,\ldots,t)$:

$$I(x,y,t) \overset{\text{SHT}}{\rightarrow} P(u,v,a_1,a_2,\ldots,t),$$

where (u,v) is the position of the object described by $a_1,a_2,\ldots,$ at time t. In the case of a circle, a_1 is its radius.

2.3. Constrained search

Each Hough image, $P(u,v,a_1,a_2,\ldots,t)$, consists of a set of weighted feature points. Motion tracking involves finding the correspondence of these features between frames. The correspondence problem is combinatorially explosive: considering all possible trajectories will not be feasible even for a moderate number of frames and feature points. Even if we know all possible trajectories, how do we determine which is the correct one? To cope with the complexity of this problem, we utilise constraints of maximum and minimum velocity (Rangarajan and Shah, 1991). If an upper bound on the velocity is known a priori, then, given a position in one frame, we can limit the search for possible matches in the next frame to a small neighbourhood of the position in the present frame. Similarly, if the object is moving, it must change position by some minimal amount between frames. These constraints enable us to perform a limited search in a smaller temporal neighbourhood, so reducing complexity. We also constrain the size and the shape of the objects to be fixed along each possible trajectory.

2.4. Smoothness of motion

Sethi and Jain (1987) proposed an approach for establishing correspondence in a monocular image sequence using the smoothness of motion. The argument is that the speed and direction of a given point will be relatively unchanged from one frame to the next for all moving objects, rigid and nonrigid, provided the sampling rate is high enough.

Our cost function uses the following criteria: the motion must be smooth in velocity and direction, and the trajectory must pass through the points of the parameter space with the maximum peak value. The velocity and direction between frames $t-1$ and t are denoted V_{t-1} and ϕ_{t-1}, respectively (Fig. 1). So, to assess the fitness of any trajectory, we consider three constraints.

The first constraint adds the peak values of the accumulator space through which the trajectory passes:

$$E_1 = \sum_{t=1}^{N} \mathrm{Peak}_t, \tag{1}$$

which ensures that the trajectory will pass through the points of the parameter space with the maximum structure evidence. The second constraint expresses the smoothness in direction between two consecutive frames as

$$E_2 = \sum_{t=2}^{N-1} |\phi_{t-1} - \phi_t|. \tag{2}$$

The third constraint penalises points in the parameter space which correspond to large changes in velocity:

$$E_3 = \sum_{t=2}^{N-1} |V_{t-1} - V_t|. \tag{3}$$

Combining Eqs. (1)–(3) gives the cost function to be maximised:

$$E = w_1 E_1 - w_2 E_2 - w_3 E_3, \tag{4}$$

where w_1, w_2 and w_3 are weights that can be adjusted to vary the relative contribution of each term. To find the optimal trajectory maximising E, we apply a DP scheme. Despite the hard constraints introduced, the computation cost of the direct enumeration is still high, and increases exponentially as the number of frames increases. DP overcomes all these difficulties, always yielding the absolute maximum and allowing hard constraints to be enforced.

With DP, we can identify a global optimum in one multi-stage procedure based on the principle that *an optimal decision for each of the remaining states must not depend on previously reached states or previously chosen decisions* (Bellman and Dreyfus, 1962). We divide the optimisation problem into stages, corresponding to frames, with a policy decision required at each, namely to maximise the cost function. Each stage has a number of associated state variables. In our case, these are the weighted features (i.e., the peaks of each accumulator array). Our network is not fully connected

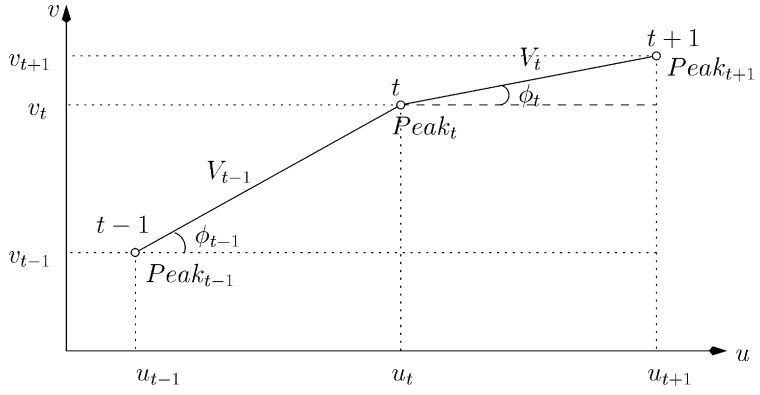

Fig. 1. Smoothness of motion is defined relative to peak accumulator values for adjacent frames.

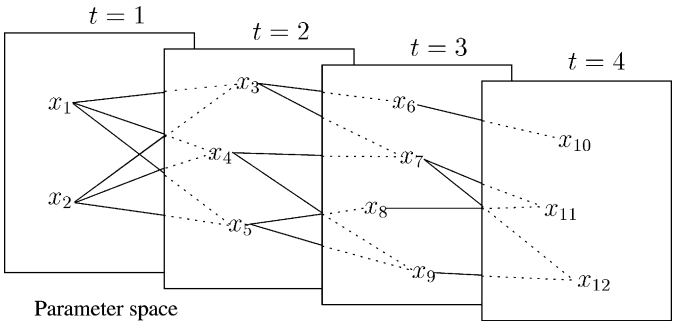

Fig. 2. Schematic illustration of the temporal dynamic programming algorithm.

because we perform a constrained search in a small temporal neighbourhood, see Section 2.3. A schematic representation of a simplified temporal DP algorithm, for 4 stages and 12 state variables, is depicted in Fig. 2. For every trajectory, we define an energy function of the form:

$$E = E(x_1, x_2, \ldots, x_t),$$

where $x_1, x_2, \ldots, x_t$ are the state variables, or the points in the parameter space, and t is the stage index.

Because of the smoothness of motion constraints, Eqs. (2) and (3), we introduce a delay, or time lag, in our system so we cannot apply the standard form of DP. This means that the policy decision in the current state depends upon that of the previous state: therefore, the principle of optimality is not applicable. To overcome this difficulty, we implement a temporal, or time-delayed, DP algorithm in which the two-element vector of state variables, (x_t, x_{t+1}), is fixed. This idea is depicted, for the previous example, in Fig. 3.

So the energy function can be written in the form:

$$E(x_1, x_2, \ldots, x_t) = E_1(x_1, x_2, x_3) + \cdots + E_{t-2}(x_{t-2}, x_{t-1}, x_t),$$

where

$$E_{t-1}(x_{t-1}, x_t, x_{t+1}) = w_1 \text{ Peak}_t + w_2|\phi_{t-1} - \phi_t| + w_3|v_{t-1} - v_t|.$$

In this case, the optimal value is a function of two temporal peaks in the motion trajectory of the form:

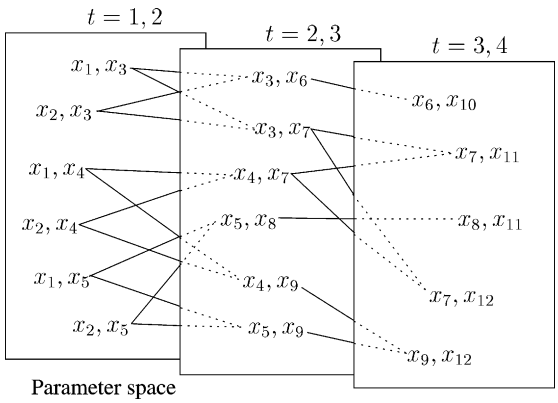

Fig. 3. Temporal (time-delayed) dynamic programming algorithm.

$$S_t(x_t, x_{t+1}) = \min_{(x_{t-1}, x_t)} [S_{t-1}(x_{t-1}, x_t) + E_{t-1}(x_{t-1}, x_t, x_{t+1})].$$

We can now apply the standard form of DP.

2.5. Implementation issues

In our algorithm, the important features in each frame are the centroids of candidate objects, in this example circles. On a frame-by-frame basis, evidence for the existence of all possible centroids is carried out using a Hough-based technique. The resulting accumulators are filtered using nonmaximal suppression to find all local maxima. DP is then applied to all the local peaks, without any threshold or other selection.

Partial occlusion, where the accumulator value for the centroid of the object is not a global

maximum (either by physical partial occlusion or some noise process), is automatically built into the algorithm, as selecting centroid positions with smaller accumulator values along a valid trajectory maximizes the overall energy in Eq. (4).

To cater for full occlusion, where there is no evidence for the object in one or more frames, an additional state is added to each stage to the DP algorithm. This allows the DP algorithm to ignore noise or other false peaks when doing so results in a greater total energy. After the optimum trajectory is found, a second stage is used to interpolate missing centroids with respect to the smoothness constraint. This modification has been successfully implemented and tested using sequences of 9 frames. The algorithm performs correctly even when up to 4 frames are missing.

In this technique we have not attempted to find the optimum weights for Eq. (4), and in the following experiments we have used fixed and equal weight values. We note, in common with snake curves (Davison et al., 2000), that the second constraint in our cost function is a rigidity term and the third is an elastic term. Accordingly, we can choose weights to favour large or small changes in velocity or direction.

3. Experimental results

The algorithm was tested on both synthetic and real images. Following Nash et al. (1997), we have chosen a moving circle of fixed radius to evaluate our algorithm. With synthetic images, we performed two trials with increasing levels of noise: one for constant linear motion and one for curvilinear motion. In both, the sequence consisted of 10 frames, each of which is a binary image. The added noise had a uniform probability density function; affected pixels had their polarity inverted.

The first experiment quantified the noise performance of the new tracking algorithm compared with the SHT and the VHT. Each image of the 10-frame sequence consisted of 120×120 pixels. The circle was of known radius, $r = 10$, moving with constant linear velocity in both x and y directions. For a given noise level, 60 sequences were generated, with the level of noise increasing from 0% to 50% in 2% increments. Fig. 4 depicts the tracking performance as a function of noise for SHT, VHT and the new evidence-based tracking algorithm. The error measure employed here is the Euclidean distance between the centre of the detected circles and ground truth, averaged over 60 different sequences. As shown in Fig. 4, the new method offers superior performance over the SHT and gives comparable results to the VHT: until about 40% noise, the new method and the VHT give comparable (essentially perfect) robustness to noise. For greater levels of noise, the new algorithm is inferior to the VHT, but still gives better results than the SHT. However, it avoids the strong assumption of constant linear velocity.

In the second noise experiment, the circle was moving on a parabolic trajectory (Fig. 5), generated for a range of curvature angles. The smaller the curvature angle, the tighter the curve. Curvature angles range from 180° (a straight line) to 20°, where the circle reverses direction in a few pixels. Again, our sequence consists of 10 frames, each 120×120 pixels. At each combination of curvature angle and noise level, we generated 30 images. Results of the simulation are depicted in Fig. 6 for the new tracking algorithm and in Fig. 7 for the SHT. Despite some small curvature angles, implying abrupt changes in motion, the new tracking algorithm is very robust, as seen by the large region of perfect performance.

We also evaluated the performance of the new tracking algorithm on one real image sequence: the well-known table tennis sequence (City University-IPL Image Database). A ball bounces on a table-tennis bat, reaches a maximum elevation and then falls under gravity. The sequence consists of 10 frames and each frame has resolution 360×243 pixels. The background (texture) of the sequence is quite complex, as seen from the edge-detected image (10th frame) on the left of Fig. 8, containing many potential but false circle features. The right image of the same figure illustrates the same frame of the sequence, where the ball is blurred by motion and the SHT fails to track it. Results with our new method are depicted with a circle and those of the SHT with a rectangle. The results for the whole

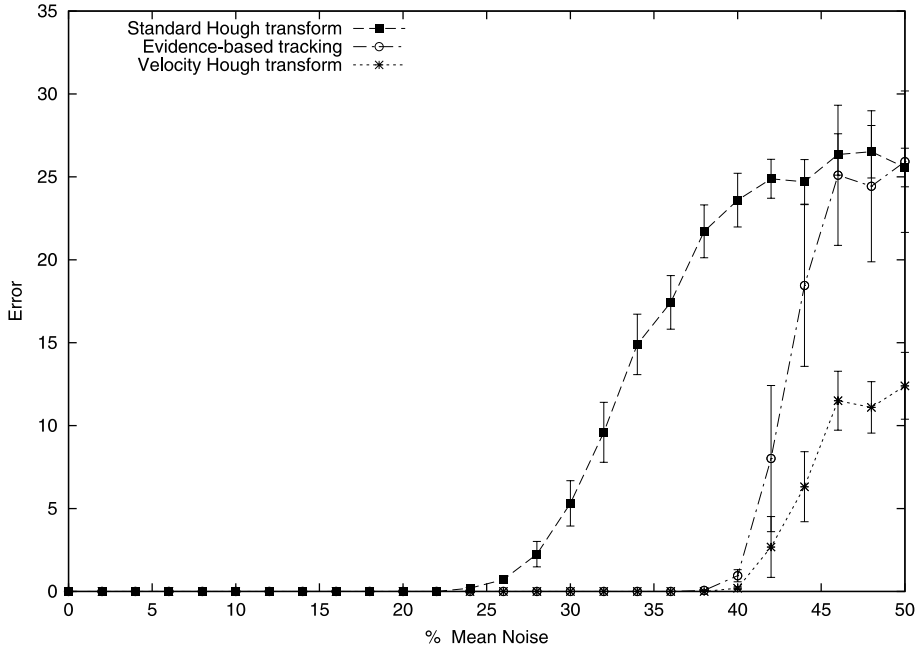

Fig. 4. Comparison of the noise performance of SHT, VHT and the evidence-based algorithm with constant linear velocity (synthetic image).

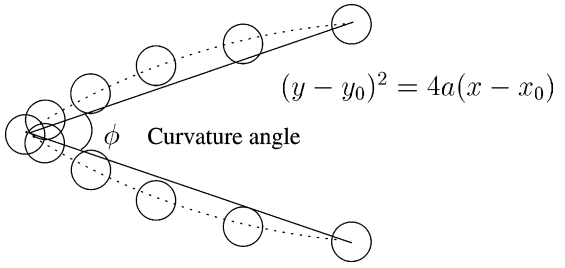

Fig. 5. Defining curvature angle.

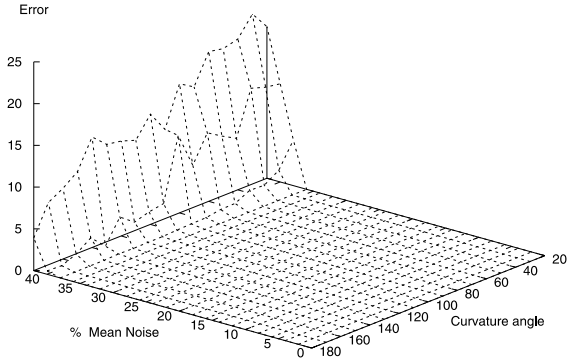

Fig. 6. Noise performance of new tracking algorithm for different curvature angles.

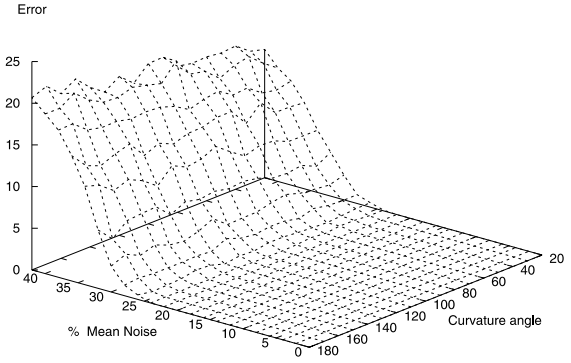

Fig. 7. Noise performance of SHT for different curvature angles.

sequence in a restricted region around the ball (marked with a white rectangle in the right of Fig. 8) are depicted in Fig. 9.

4. Conclusions and future work

This work has adopted the concepts of smoothness of motion and the global evidence-

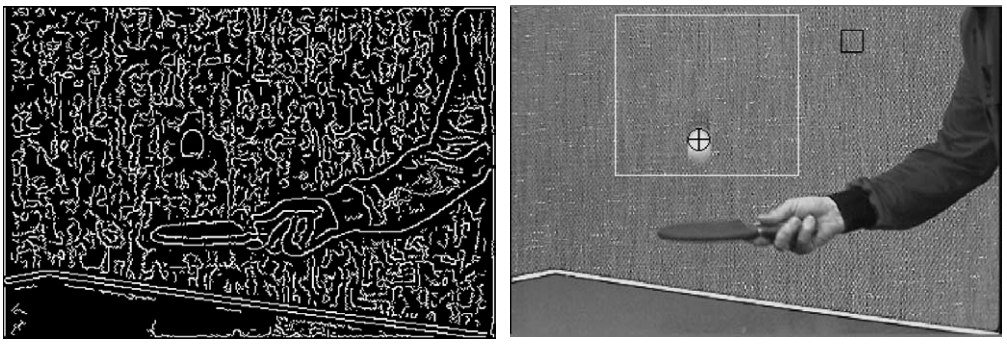

Fig. 8. 10th frame of table-tennis sequence. Left: edge-detected image. Right: ball position from new method (circle) and SHT (rectangle).

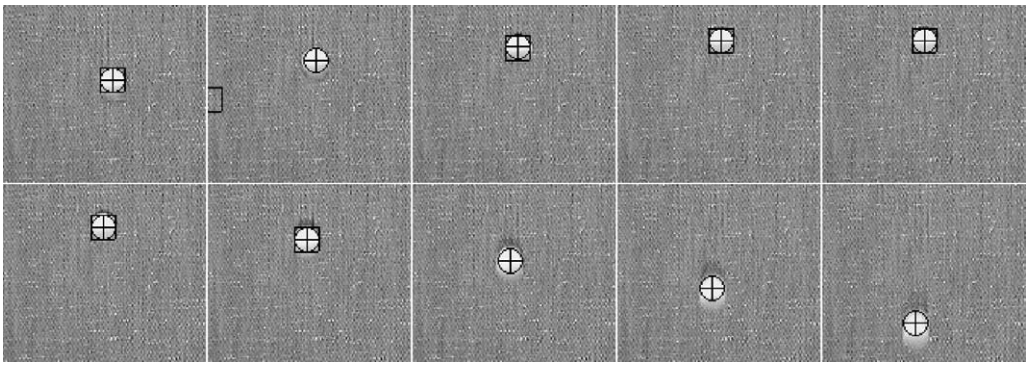

Fig. 9. Ball position obtained by new method (circle) and SHT (rectangle where present) for complete table-tennis sequence.

gathering of structure from earlier work. There are several advantages of our approach relative to other trajectory-based algorithms using the notion of smoothness of motion. First, our algorithm requires no initialisation, since we use a Hough-based approach for evidence gathering. Second, the whole image sequence is processed globally and optimally using a temporal (time-delay) DP algorithm. Third, a weighting scheme ensures that we use only 'good' features. Finally, the method copes with arbitrary smooth motion: no assumptions of constant or linear velocity are imposed. Nonetheless, we can still handle relatively abrupt changes as shown in the results for high curvature (i.e., small curvature angle) motion. These advantages are reflected in excellent tracking performance in high levels of noise – considerably above the performance of the SHT.

The method is also suitable for extension to tracking nonparametric and parametric objects other than circles. In future work the GHT (Ballard, 1981) will be utilised to generate accumulators for the centroid, scale and rotation. The cost function for the DP algorithm will be modified to impose smoothness constraints on the changes in size and orientation from frame to frame. Alternatively, a scale/rotation invariant version of the GHT (Kassim et al., 1999) may be utilised to track only centroids. Finally the algorithm will be extended to track multiple models and evaluated in more realistic environments.

References

Aguado, A.S., Nixon, M.S., Montiel, E.S., 1998. Parameterizing arbitrary shapes via Fourier descriptors for evidence-gathering extraction. Comput. Vision Image Understanding 69 (2), 202–221.

Ballard, D.H., 1981. Generalising the Hough transform to detect arbitrary shapes. Pattern Recognition 13 (2), 111–122.

Bellman, R.E., Dreyfus, S.E., 1962. Applied Dynamic Programming, second ed. Princeton University Press, Princeton, NJ.

Broida, T.J., Chellappa, R., 1986. Estimation of motion parameters from noisy images. IEEE Trans. Pattern Anal. Machine Intell. 8 (1), 90–99.

City University-IPL Image Database. Table tennis image sequence. http://www.image.cityu.edu.hk/imagedb.

Davison, N.E., Eviatar, H., Somorjai, R.L., 2000. Snakes simplified. Pattern Recognition 33 (10), 1651–1664.

Illingworth, J., Kittler, J., 1988. A survey of the Hough transform. Comput. Vision Graphics Image Process. 44 (1), 87–116.

Kassim, A.A, Tan, T., Tan, K.H., 1999. A comparative study of efficient generalised Hough transform techniques. Image Vision Comput. 17 (10), 737–748.

Nash, J.M., Carter, J.N., Nixon, M.S., 1997. Dynamic feature extraction via the velocity Hough transform. Pattern Recognition Lett. 18 (10), 1035–1047.

Rangarajan, K., Shah, M., 1991. Establishing motion correspondence. Comput. Vision Graphics Image Process. 54 (1), 56–73.

Sethi, I.K., Jain, R.C., 1987. Finding trajectories of feature points in a monocular image sequence. IEEE Trans. Pattern Anal. Machine Intell. 9 (1), 56–73.

Shariat, H., Price, K.E., 1990. Motion estimation with more than two frames. IEEE Trans. Pattern Anal. Machine Intell. 12 (5), 417–434.