

OBJECT TRACKING VIA THE DYNAMIC VELOCITY HOUGH TRANSFORM

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

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

ABSTRACT

Motion tracking is an important task in computer vision. A new technique, the dynamic velocity Hough transform (DVHT), 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 an optimal, 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. The constrained optimisation problem is solved using a temporal (time-delay) dynamic programming algorithm. Tracking in noise is much superior to the standard Hough transform.

1. MOTIVATION AND APPROACH

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 and in situations of occlusion [1]. The VHT proposed by [2] 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 our approach, 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 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 [3]. 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. 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. OBJECT TRACKING ALGORITHM

Here, we explain the proposed algorithm for object detection in a long image sequence. Considering the whole image sequence as a three-dimensional process $I(x, y, t)$ with spatial variables (x, y) , where t represents the time index of a frame, we can transform the data sequence into a parameter space $P(u, v, a_1, a_2, \dots, t)$:

$$I(x, y, t) \xrightarrow{SHT} P(u, v, a_1, a_2, \dots, t)$$

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

Each Hough image, $P(u, v, a_1, a_2, \dots, t)$, 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

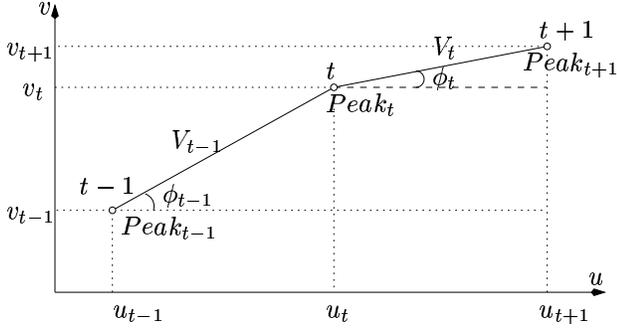


Fig. 1. Smoothness of motion is defined relative to peak accumulator values for adjacent frames.

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. 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.

In [4], an approach is proposed 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. This approach is used here.

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 accumulator 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 (Figure 1). So, to assess the fitness of any trajectory, we consider three constraints. The first adds the peak values of the accumulator space through which the trajectory passes:

$$E_1 = \sum_{t=1}^N Peak_t \quad (1)$$

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:

$$E_2 = \sum_{t=2}^{N-1} |\phi_{t-1} - \phi_t| \quad (2)$$

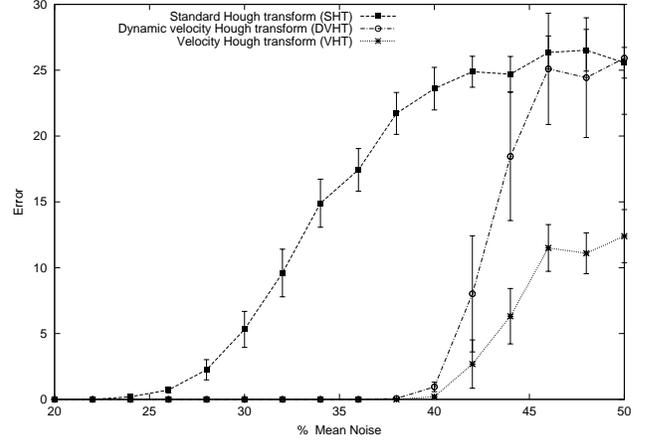


Fig. 2. Comparison of the noise performance of SHT, VHT and DVHT with constant linear velocity (synthetic image).

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

$$E_3 = \sum_{t=2}^{N-1} |V_{t-1} - V_t| \quad (3)$$

Combining equations (1), (2) and (3) gives the cost function to be maximised:

$$E = w_1 E_1 - w_2 E_2 - w_3 E_3$$

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 dynamic programming (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. DP divides 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 because we perform a constrained search in a small temporal neighbourhood. Because of the smoothness of motion constraints, equations (2) and (3), we introduce a delay, or time lag, in our system so we cannot apply the standard form of dynamic programming. 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 [5]. To overcome this difficulty, we implement a temporal, or time-delayed, DP algorithm in which the two-element vector of the current and the next state variables, (x_t, x_{t+1}) , is fixed.

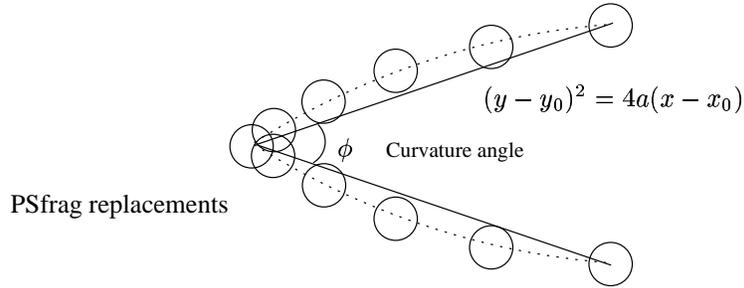


Fig. 3. Defining curvature angle.

3. EXPERIMENTS

The algorithm was tested on both synthetic and real images. Following [2], 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 standard Hough transform (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. Figure 2 depicts the tracking performance as a function of noise for SHT, VHT and the DVHT. 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 Figure 2, the DVHT offers superior performance over the SHT and gives comparable results to the VHT: until about 40% noise, the DVHT and the VHT give comparable (essentially perfect) robustness to noise. For greater levels of noise, the DVHT 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 (Figure 3), 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 generate 30 images. Results of the simulation are depicted in Figure 4 for the DVHT and in Figure 5 for the SHT. Despite some small curvature angles, implying abrupt changes in motion, the DVHT tracking algorithm is very robust, as seen by the large area of perfect

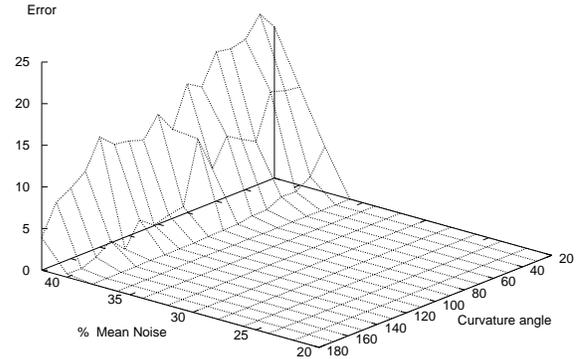


Fig. 4. Noise performance of DVHT for curvilinear motion for different curvature angles.

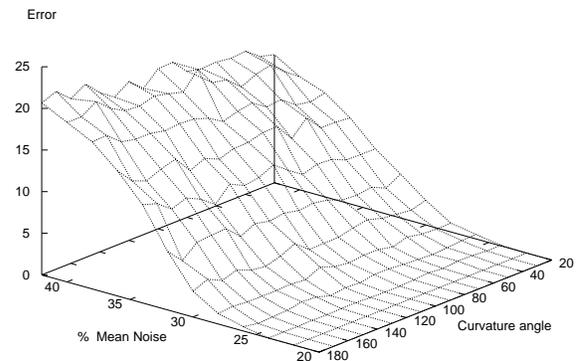


Fig. 5. Noise performance of SHT for curvilinear motion for different curvature angles.

performance.

We also evaluated the performance of the DVHT tracking algorithm on one real image sequence: the well-known table tennis sequence [6]. 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 Figure 6, 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 the DVHT are depicted with a circle and those of the SHT with a rectangle. The results for the whole sequence in a restricted region around the ball

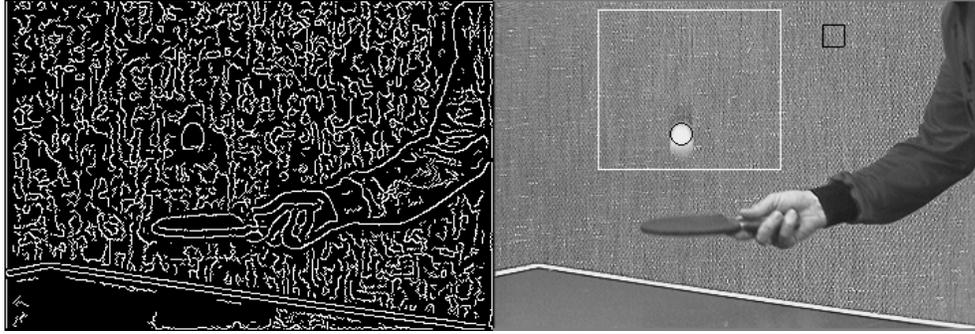


Fig. 6. 10th frame of table-tennis sequence. Left: edge-detected image. Right: ball position from DVHT (circle) and SHT (rectangle).

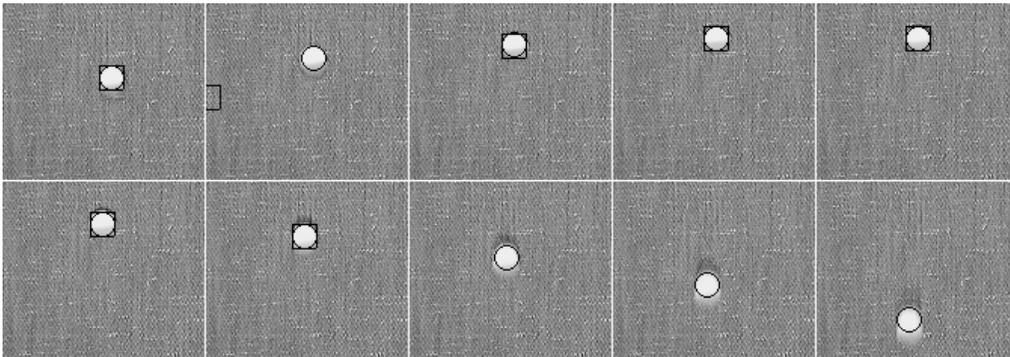


Fig. 7. Ball position obtained by DVHT (circle) and SHT (rectangle where present) for complete table-tennis sequence.

(marked with a white rectangle in the right of Fig. 6) are depicted in Figure 7.

4. DISCUSSION AND CONCLUSION

This work has adopted the concepts of smoothness of motion and the global evidence-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) dynamic programming 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 standard Hough transform. The method is also suitable for tracking parametric and nonparametric objects other than circles.

5. REFERENCES

- [1] J. Illingworth and J. Kittler, “A survey of the Hough transform,” *CVGIP*, vol. 44, no. 1, pp. 87–116, 1988.
- [2] J. M. Nash, J. N. Carter, and M. S. Nixon, “Dynamic feature extraction via the velocity Hough transform,” *Pattern Recogn. Lett.*, vol. 18, no. 10, pp. 1035–1047, 1997.
- [3] A. S. Aguado, M. S. Nixon, and E. M. Montiel, “Parameterizing arbitrary shapes via Fourier descriptors for evidence-gathering extraction,” *CVGIP*, vol. 69, no. 2, pp. 202–221, 1998.
- [4] I. K. Sethi and R. Jain, “Finding trajectories of feature points in a monocular image sequence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 56–73, 1987.
- [5] R. Bellman and S. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, Princeton, NJ, 1962.
- [6] CityU-IPL Image Database, “Table tennis sequence,” <http://www.image.cityu.edu.hk/imagedb>.