# EVIDENCE-BASED OBJECT TRACKING VIA GLOBAL ENERGY MAXIMIZATION

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

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

## ABSTRACT

This paper describes a robust algorithm for arbitrary object tracking in long image sequences. This technique extends the dynamic Hough transform proposed in our earlier work to detect arbitrary shapes undergoing affine motion. The proposed tracking algorithm processes the whole image sequence globally. First, the object boundary is represented in lookup-table form, and we then perform an operation that estimates the energy of the motion trajectory in the parameter space. We assign an extra term in our cost function to incorporate smoothness of deformation. The object is actually rigid, so by 'deformation' we mean changes due to rotation or scaling of the object. There is no need for training or initialization, and an efficient implementation can be achieved with coarse-to-fine dynamic programming and pruning. The method, because of its evidence-based nature, is robust under noise and occlusion.

## 1. INTRODUCTION

Most motion tracking techniques consider information in the current frame plus a small number of previous frames to predict the motion and structure parameters for the next frame. In cases of fast moving objects, noise, and clutter, the wrong decision at the current frame will compromise tracking of the object for subsequent frames; only a global processing technique can give the optimal result. Apart from the velocity Hough transform and the dynamic Hough transform there are no efficient algorithms in the literature to find the optimal trajectory [1, 2]. Our approach provides a method of integrating shape extraction within an energy maximization framework. It defines the object to be tracked in terms of maxima of a motion trajectory with an associated energy function. The choice of this energy function is a compromise between evidence from image data and integration of motion and deformation constraints. An energy maximization method is required to extract the desired motion trajectory and determine the structure parameters of the object. Our method seeks a global optimum.

## 2. PROBLEM STATEMENT

Let us consider an image sequence as a three-dimensional space $(x, y, t)$ comprising two spatial dimensions $(x, y)$ corresponding to every image plane and one temporal dimension $t$. Then in this space, a moving object generates a trajectory. Hence the determination of the moving object amounts to the determination of the motion trajectory by processing globally the image sequence. The motion trajectory tries to link points—potential centroids of an object—according to local measures of continuity and smoothness and specifically continuity in direction, displacement, and deformation. Such quantities tend to be locally smooth, but can change dramatically from the first to the last frame. They should be consistent with the observed data. Such problems can be naturally formulated in terms of energy maximization.

Maximization methods attempt to model global image properties, i.e., characteristics of the moving object that cannot be captured by local correlation techniques or with parametric motion models. We consider a very general definition of smoothness which can accommodate not only irregular sampling but also missing data. Furthermore, pairwise interactions of adjacent points on the trajectory contribute to a global nature of the smoothness.

As a shape coding method, we use the generalized Hough transform (GHT) [3, 4]. The image sequence is pre-processed, first applying an edge detection algorithm to each frame, and then transforming into a 5-dimensional parameter space $P(u, v, \theta, s, t)$ where $(u, v)$ is the position of the centroid, $\theta$ is the orientation, $s$ is the scale factor of the object, and $t$ is the time index for each frame, $1 \leq t \leq N$. The motion trajectory is represented by a set of discrete points in each frame, where we can consider the speed and direction of a point at frame $t$ to be:

$$V_t = \sqrt{(x_{t-1} - x_t)^2 + (y_{t-1} - y_t)^2} \qquad (1)$$

$$\phi_t = \arctan\left[\frac{(y_{t-1} - y_t)}{(x_{t-1} - x_t)}\right] \qquad (2)$$

where $(x_{t-1}, y_{t-1})$ and $(x_t, y_t)$ are the locations of the point in frames $(t-1)$ and $t$. The problem is to determine the trajectory which satisfies some appropriate energy criterion in this space.

## 3. ENERGY REPRESENTATION

An energy function to assess the fitness of any trajectory can be considered to have the following three terms: (a) Hough energy representing the points of the parameter space with maximum structure evidence; (b) motion energy determined by the smoothness of velocity and direction of the motion trajectory [5]; (c) deformation energy representing the variations of the object over time, scale and orientation. We consider the latter two terms to have equal weight in the energy formulation, so that:

$$E_{\text{traj}} = w_1 E_{\text{Hough}} - w_2(E_{\text{motion}} + E_{\text{def}}) \tag{3}$$

where $w_1$ and $w_2$ are weights that can be adjusted to vary the relative importance of each term.

The first term forces the trajectory to pass through the points in the parameter space with maximum structure evidence, using the specific form:

$$E_{\text{Hough}} = \sum_{t=1}^{N} p_t$$

which simply adds the peak values, $p_t$, of the accumulator space through which the trajectory passes. The motion energy represents the elasticity and rigidity of the trajectory, and has the form:

$$E_{motion} = \sum_{t=2}^{N-1} |V_{t-1} - V_t| + \sum_{t=2}^{N-1} |\phi_{t-1} - \phi_t|$$

where $V_t$ and $\phi_t$ are as in equations (1) and (2). The first term penalizes the points in the parameter space which correspond to large changes in speed, and the second term penalizes large changes in direction.

The deformation energy expresses the smoothness of deformation, which means that the object will deform in size and orientation gradually during time. This energy term favors small changes in orientation and scale and penalizes abrupt changes. It is given by:

$$E_{\text{def}} = \sum_{t=2}^{N-1} |s_{t-1} - 2s_t + s_{t+1}| + \sum_{t=2}^{N-1} |\theta_{t-1} - 2\theta_t + \theta_{t+1}|$$

where $s_t$ and $\theta_t$ are the scaling factor and orientation of the object at frame $t$. To find the optimal trajectory that maximizes the cost function (3), we apply a dynamic programming (DP) scheme [6].

## 4. OPTIMIZATION

Following our previous work [2], the optimization problem to find the parameters of $P(u, v, \theta, s, t)$ is efficiently solved using dynamic programming. DP allows the introduction of constraints that cannot be violated, called hard constraints, as well as second-order continuity constraints, which are inherent in the energy formulation. These latter are known as soft constraints because they are not satisfied absolutely, only to a certain degree. We divide the optimization problem into stages, corresponding to frames, with a policy decision required at each, namely to maximize the energy function. Each stage has a number of associated state variables. In our case, these are the weighted features, points in the parameter space (i.e., peaks of each accumulator array). For each trajectory, we associate an energy function:

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

where $x_1, x_2, \ldots, x_N$ are the state variables, or the points in the parameter space. Because we wish to represent the smoothness of motion and deformation, we introduce a time lag, or delay in our system; therefore, the principle of optimality is not applicable. Hence, to overcome this difficulty, we implement a time-delayed DP algorithm, in which the two-element vector of state variables, $(x_t, x_{t+1})$, is fixed. So the energy function can be written:

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

The recursion that relates the cost or reward earned during previous stages is a function of two temporal state variables of the form:

$$R_t(x_t, x_{t+1}) = \max_{(x_{t-1}, x_t)} \begin{bmatrix} R_{t-1}(x_{t-1}, x_t) + \\ E_{t-1}(x_{t-1}, x_t, x_{t+1}) \end{bmatrix}$$

## 5. IMPLEMENTATION ISSUES

The global evidence-based search technique considers all possible peaks in the Hough space, even those with zero value, to find the optimal smooth trajectory. Use of Hough techniques avoids the need for initialization, which can contribute to major error in other approaches. The motion trajectory problem involves finding the possible correspondence of features among frames. This correspondence problem is combinatorially explosive even with a DP scheme. To cope with the complexity of this problem, we need to perform a constrained search. Fortunately, the local connectivity of the motion trajectory can be exploited to reduce the computation time dramatically. A search window (a hard constraint set by knowledge of the maximum and minimum allowable speeds) determines the extent to which the motion trajectory is allowed to stretch or bend at that point. These constraints are employed to prevent impossible motion trajectories, and are both qualitatively and computationally beneficial [7].

Besides maximum and minimum velocity, further constraints should be introduced to control the amount of deformation. This means that the object scaling and orientation cannot exceed some maximum predefined values. The principle of smoothness of deformation exploits the fact that, because of inertia, the size and rotation of the object cannot change instantaneously. This assumption will be valid for all moving objects. Provided the sampling rate is high enough, the changes in scale and orientation will be gradual. Thus, we introduce some further constraints that enable us to perform a limited search in a smaller temporal neighborhood of the parameter space, so reducing the complexity of the problem.

The time complexity of the DP scheme can be reduced further by employing an absolute pruning technique. We perform a two-step search. Initially, we calculate the best trajectory that passes though the points of the parameter (Hough) space considered alone. By a backtracking procedure, we then prune all points in the parameter space lying on trajectories with an energy smaller than a threshold value (equal to 0.8 of the maximum energy in this work). Hence, we reduce dramatically the solution space, and subsequently we can perform a more extensive search considering the motion and deformation terms as well. This is done with $w_1 = 0.8$ and $w_2 = 0.2$ in equation (3). Further time and memory reductions can be achieved using the coarse-to-fine DP algorithm [8] whereby we form a series of coarse approximations by aggregating states into superstates. For each coarse approximation, the optimal trajectory is found using DP. The superstates along this optimal trajectory are noted and the process is iterated until the optimal path is found. This idea can be simply adapted to our 5D optimization problem. In each frame, we merge state variables to form 5D hyperstate variables.

## 6. SIMULATIONS AND RESULTS

Two simulations were carried out: one to test performance in the presence of noise and the other to test robustness to object occlusion with a small amount (10%) of noise. Image sequences were synthesized so that for each frame of the sequence, the quantity of noise present or occlusion bar width is known. In both cases, the generated sequence is binary. The error measure employed is the root mean square error of the estimated parameters relative to ground truth, averaged over 50 trials (i.e., different seed for the pseudorandom noise generator).

The first simulation was designed to quantify the noise performance of the new tracking algorithm compared with

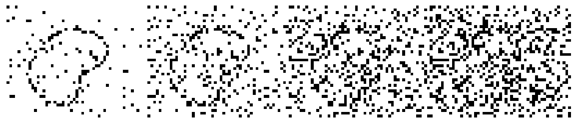

**Fig. 1**. Typical frames from image sequences showing the arbitrary object with 4%, 12%, 20% and 28% added noise.

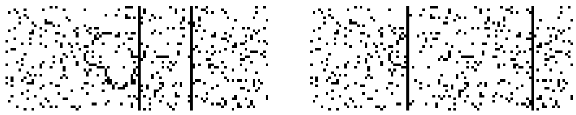

**Fig. 2**. Typical frames showing the arbitrary object occluded with bars of width 20 pixels and 48 pixels (with 10% noise).

the GHT. Each image of a 32-frame sequence consisted of 320 × 280 pixels. The object to be tracked has arbitrary shape and moves with constant linear velocity in the $x$ direction, and is rotated and scaled at a constant rate through the image sequence. The added noise had a uniform probability density function; affected pixels had their polarity inverted. The level of noise varied from 0% to 30% in 2% increments. Figure 1 shows typical frames for a representative range of added noise. In the occlusion simulation, the object was moving with the same parameters as before, but we added an occlusion bar in the middle of each frame. The width of the occlusion bar varied from 0 to 60 pixels in 4 pixel increments. The object dimension is 26 × 25 pixels, and it moves by 10 pixels per frame. The 11-frame sequence consisted of 320 × 280 pixels and the object in some frames is partially or totally occluded by the bar (Figure 2).

As shown in Figures 3 and 4, our method offers superior performance over the GHT for simulations with added noise and occlusion, especially for the more demanding conditions. Thus, despite total occlusion over several frames, we can still track the occluded object with low error (Fig. 4).

## 7. CONCLUSIONS

Robust tracking of objects in noise is an outstandingly important problem in computer vision. We have considered the tracking of a moving object as an energy maximization problem. That is, the motion trajectory is represented by an energy function with an image-dependent term, a term penalizing large changes in velocity (speed and direction) and a second-order smoothness term. This is then maximized over the image sequence using time-delay dynamic programming to exploit the temporal correlation between adjacent points in the motion trajectory and so determine the global optimum. Efficiency can be improved using coarse-to-fine DP with pruning of points on trajectories below some energy threshold in Hough space. The method

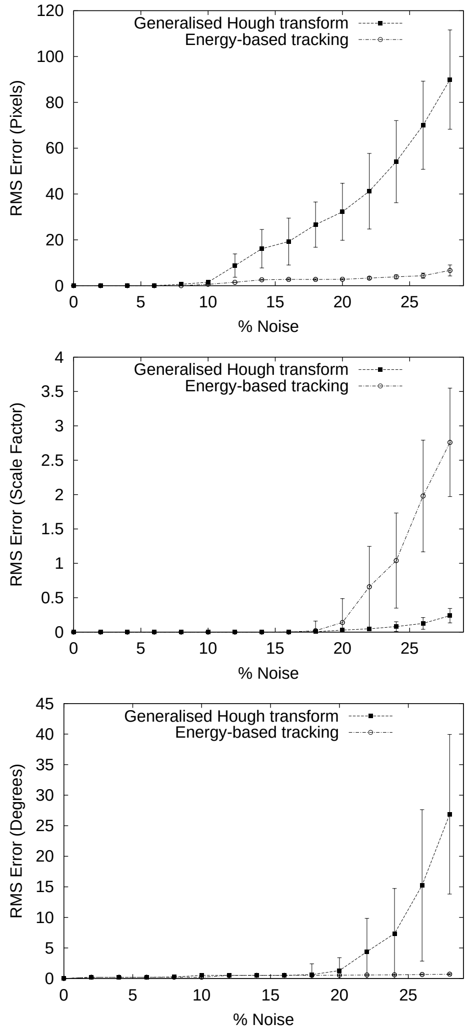


**Fig. 3**. Comparison of the noise performance of GHT and energy-based tracking. Top: rotation error; Middle: scale error; Bottom: translation error. Error bars are standard deviations over 50 trials.

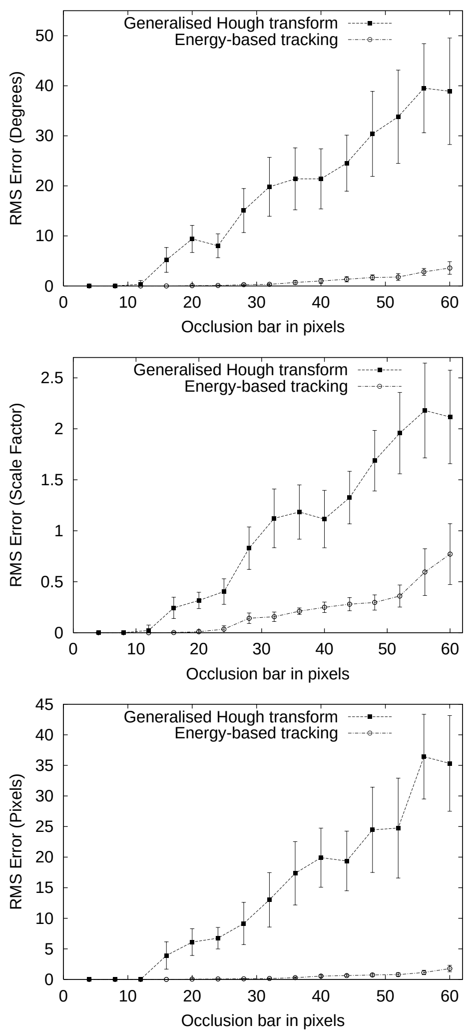


**Fig. 4**. Comparison of the occlusion performance of GHT and energy-based tracking. Top: rotation error; Middle: scale error; Bottom: translation error. Error bars are standard deviations over 50 trials.

gives superior results compared to the standard generalized Hough transform and proved to be especially robust under noise and occlusion.

## 8. REFERENCES

[1] J. M. Nash, J. N. Carter, and M. S. Nixon, "Dynamic feature extraction via the velocity Hough transform," *Pattern Recognition Letters*, vol. 18, no. 10, pp. 1035–1047, 1997.

[2] P. Lappas, J. N. Carter, and R. I. Damper, "Robust evidence-based object tracking," *Pattern Recognition Letters*, vol. 23, no. 1–2, pp. 253–260, 2002.

[3] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[4] A. A. Kassim, T. Tan, and K. H. Tan, "A comparative study of efficient generalised Hough transform techniques," *Image and Vision Computing*, vol. 17, no. 10, pp. 737–748, 1999.

[5] K. Rangarajan and M. Shah, "Establishing motion correspondence," *Computer Vision, Graphics and Image Processing – Image Understanding*, vol. 54, no. 1, pp. 56–73, 1991.

[6] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.

[7] I. K. Sethi and R. C. Jain, "Finding trajectories of feature points in a monocular image sequence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 56–73, 1987.

[8] C. Raphael, "Coarse-to-fine dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1379–1390, 2001.