Pergamon

Pattern Recognition, Vol. 27, No. 12, pp. 1591–1603, 1994
Elsevier Science Ltd

0031-3203/94 $7.00 + .00

0031-3203(94)00079-4

CYCLIC MOTION DETECTION FOR MOTION BASED RECOGNITION†

PING-SING TSAI,‡ MUBARAK SHAH,‡ KATHARINE KEITER§ and TAKIS KASPARIS§
‡ Computer Science Department, University of Central Florida, Orlando, FL 32816, U.S.A.
§ Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, U.S.A.

(*Received 22 September* 1993; *in revised form* 3 *June* 1994; *received for publication* 28 *June* 1994)

Abstract—The motion of a walking person is analyzed by examining cycles in the movement. Cycles are detected using autocorrelation and Fourier transform techniques of the smoothed spatio-temporal curvature function of trajectories created by specific points on the object as it performs cyclic motion. A large impulse in the Fourier magnitude plot indicates the frequency at which cycles are occurring. Both synthetically generated and real walking sequences are analyzed for cyclic motion. The real sequences are then used in a motion based recognition application in which one complete cycle is stored as a model, and a matching process is performed using one cycle of an input trajectory.

Cyclic motion   Spatio-temporal curvature   Motion-based recognition

1. INTRODUCTION

Humans are very good at analyzing motion. Experiments in psychology have revealed that people are able to perceive the motion of objects from Moving Light Displays (MLD). A MLD is simply a two-dimensional movie of a collection of bright dots attached to a moving object. Upon viewing MLDs, people can recognize different types of motion undertaken by a person, such as walking forward, backward and jumping. Recognition of complicated motions, such as couple-dancing, and sophisticated judgements such as the gender of a subject and the gait of a familiar person, have also been reported. Only the dots are seen in the display (not the whole object) and there is no structure present since none of the dots are connected. Even though all parts of the object are not seen and no structure explicitly exists, humans are able to derive in their minds the three-dimensional structure of the object from the motion information. From this structure, they can recognize specifically what the moving object is and how it is moving. This is one of the theories about how humans interpret MLD type stimulus. According to this theory humans use motion information in the MLD to recover the three-dimensional structure, and subsequently use structure for recognition. There has been significant interest over the last decade in the Computer Vision community in the structure from motion theory. In this work, three-dimensional coordinates of points on the moving objects and their three-dimensional motion is recovered from a sequence of frames. This problem is formulated in terms of systems of non-linear equations given two-dimensional positions of moving points among frames. Interesting theoretical work related to the number of points required for a solution, the uniqueness of such a solution, and the effect of noise on the solution has been studied. In these approaches, it is assumed that the recovered three-dimensional structure will subsequently be used for recognition.

Another theory, which has received much less attention in the Computer Vision community, is the theory that the *motion information in the MLD* is directly used for recognition. By recognition, in this context, we mean the recognition of action through motion. For instance, the distinction between walking and running using the motion of several points on a human body is one form of motion recognition. The distinction of the different gaits of two persons using motion is in a general sense one form of object recognition. This motion based recognition is in contrast to commonly known object recognition which employs explicit three-dimensional or two-dimensional shape. Both forms of motion based recognition have been strongly demonstrated by Goddard in his recent Ph.D. thesis.[1]

A strong case for the theory that motion information is directly used for recognition is made by Johansson[2] in his paper on visual perception of biological motion. In this paper he studies motion patterns without the interference of the form aspect of the object (a human being in this case). He represents the motion of a body using bright spots to describe the motions of the main joints. He maintains that the pendulum-like motions of the body's extremities are highly specific for different types of motions, and that it is the mathematical spatio-temporal relations in the patterns created by the moving bright dots that determine perceptual response. The question of how points moving together on a screen

† This research was supported by the National Science Foundation under grant number CDA 9122006, CDA 9200369, and IRI 9220768.

can give such a definite impression of a walking person is explained by maintaining that recognition is dependent upon general principles for grouping in visual perception. Experiments were performed using a vector analysis type model to show that common motion vector components among the moving bright points are separated from the rest of the motion pattern, and are seen as reference frames for the pendulum motion. The recognition of walking from the motion of the bright spots is shown to be independent of the course of the common component. When a common component was subtracted from the element motions in the walking pattern, all observers still immediately recognized a walking pattern. Similarly, all observers immediately identified a walking pattern when an extra component was added to the primary motion of each element. These results show that the motion patterns carry all the essential information needed for immediate visual identification of such human motions as walking. It was also demonstrated that people immediately report seeing a walking human being when they are only presented with a little less than half a step cycle (one "step").

We are interested in pursuing the second theory of movement analysis, which deals with the direct use of motion information for recognition. It is our belief that visual interpretation is a highly complex task. A single source of information, for instance the structure of an object, is not sufficient for robust and accurate recognition. We need to employ a combination of multiple cues such as motion, specularities, texture, etc., and exploit information in each cue using several alternate ways. The structure from motion methods compute intrinsic surface properties, such as depth values. As pointed out by Witkin and Tenenbaum,[3] depth maps and other maps of 2.5-dimensional sketches are still basically just images. They still must be segmented and interpreted before they can be used for a more sophisticated task.

In this paper, cyclic motion is detected in the motion paths of joint elements during human walking. Cyclic motion can be defined as the motion undertaken by an object that follows a repeating path over time. Examples include a person walking, running, skipping, riding a bike, a pendulum swinging, a ball bouncing, wings flapping and piston moving. An application of cyclic motion detection is the detection of gait problems in an injured person by comparing the path followed by specific points on the walking body of an injured person to the path created by the same points on a healthy person. Similarly, athletic performance can be improved when an expert examines the paths created by points on an athletes body during training. The detection of cycles is also useful in recognition problems, since specific types of motion may be recognized according to the cycles a moving object makes.

We use correlation and Fourier transform techniques to detect cycles in two-dimensional trajectories created by points on a moving object. We consider the trajectory as a spatio-temporal curve in (x, y, t) space. Cyclic motion is detected by finding cycles in the curvature of this spatio-temporal curve. We assume that the oscillatory direction of an object is perpendicular to the viewing axis of the camera, and the input two-dimensional trajectories contain cycles. A very good example for this assumption is the metronome. The repeated pattern of a metronome can be observed well in front of the metronome, while it cannot be observed well from the side. We also assume the orthographic projection of image formation, so we do not need to consider the distortion of trajectories. The detected cycles are then applied to a method proposed by Rangarajan et al.[4] for matching pairs of single trajectories. Instead of storing all the trajectories with different cycles as models in order to find the correct match for an input trajectory, we store the trajectory with one complete cycle as our model, and do the matching with one cycle of the input trajectory.

In a first experiment the cycles in a sequence of points generated by a program that simulates the movement of a walking person are detected and extracted. Then, a real walking sequence is tested for cycles. In all cases, the correct frequency of the cycles is detected from the Fourier transform of pre-processed curvature functions of the trajectories.

2. RELATED WORK

A great deal of work has been done in the field of psychology to show that people can recognize objects from their trajectories.[2,5] It has been theorized that humans can recognize an object based on the motion of several points on that object by inferring the three-dimensional structure of the object from the transformations the two-dimensional image undergoes. Cutting[6] gives examples of six different types of motion: *rolling wheels, walking people, swaying trees, aging faces, the rotating night sky* and *expanding flow fields.* Todd[5] is interested in distinguishing between rigid and several types of non-rigid motion such as bending, stretching, twisting and flowing. By displaying the trajectories of either rigid or non-rigid objects. Todd shows that human observers are able to distinguish between the two. Goddard[7] has proposed a computational model for visual motion recognition in the moving light displays. He believes that the visual system continuously computes invariants used to represent objects and movement. These invariants are used to index into two-dimensional memory models. Having identified the most likely candidate, the viewpoint is computed and a verification stage operating in three-dimensional confirms or denies the hypothesis. Another possible method would use motion information to reconstruct various static qualities, and use those static qualities to index into memory and recognize the object. However, Goddard has argued for a recognition process operating directly on motion information. Engel and Rubin[8] describe an implementation of an algorithm for detecting motion boundaries given discrete position input. Motion boundaries comprise starts, stops, pauses and force impulses. Their algorithm represents image motion velocity in polar coordinates. Force impulses are asserted when the slope of zero-crossing of the second derivative of speed or direction exceeds a threshold.

In Computer Vision the work related to detection of motion before recognition has been reported. Allmen and Dyer[9] detect cyclic motion by tracking curvature extrema in spatio-temporal images. Repeating patterns are detected using a scale-space representation. In their approach, three-dimensional spatio-temporal volumes are formed by stacking a dense sequence of image frames, and when an edge operator is applied, this ST-volume contains surfaces and volumes which represent object motion swept out through time. ST-curves are detected on the ST-surfaces by connecting edge points into contours, and the curvature extrema are then found. The curvature extrema are used as tokens which are connected from one frame to the next, forming ST-curves. The ST-curves recover the cyclic behavior of the ST-surfaces. Repeating patterns in the ST-curves are then detected by matching the scale-space features of every curve. Both fine and coarse cyclic motion can be observed since curvature scale-space represents curvature over many scales.

Koller *et al.*[10] characterize vehicle trajectories by motion verbs. They exploit internal representation of about ninety German motion verbs to automatically characterize trajectory segments. The English translation of their German verbs are: to reach, to come nearer, to move away, to accompany, to go beside, etc.

Hogg[11] addresses the problem of finding a known object in an image using a generate and test strategy. He models humans with generalized cylinders of varying sizes. From the model, the occluding edges are predicted, and the hypothesis is verified by the number of edge points lying near the predicted edges. When dealing with a sequence of images, a difference picture is used to identify the approximate position of moving objects in the first frame. He also uses kinematic constraints to reduce the search space in identifying the object in subsequent frames.

Tsotsos *et al.*[12] present a framework for the abstraction of motion concepts from sequence of images. The framework includes: representation of knowledge for motion concepts that is based on semantic networks; and associated algorithms for recognizing these motions concepts.

Polana and Nelson[13] used similar techniques to judge the degree of periodicity; this differs with our main concern which is to extract one cycle from an input trajectory with unknown cycles, and subsequently uses it for matching. They considered an image sequence as a spatio-temporal solid with two spatial dimensions and one time dimension, and detected periodicity using the Fourier transform. They compute reference curve (which is essentially a trajectory) by tracking the centroid of moving region in several frames. They use reference curve to align the frames, and then compute gray-level signals at every pixel in the image frame. The gray-level signals are used to detect periodicity. The gray-level signals used by Polana and Nelson[13] are different from the curvature signals, generated from the trajectories, used in our approach.

3. DETECTION OF CYCLES USING THE FOURIER TRANSFORM

A trajectory is defined as a sequence of points $((x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_t, y_t))$, ordered by an implicit time dimension. We can represent a two-dimensional trajectory as two one-dimensional trajectories, $x(t)$ and $y(t)$, or two one-dimensional time functions, namely speed and direction. Once coordinates of points that make up a trajectory are acquired, this one-dimensional information can be Fourier transformed to detect cycles. However, when a two-dimensional trajectory is represented by two one-dimensional signals, different frequencies from the two signals may be detected, and a problem is how to combine two different frequencies to get the correct frequency for the trajectory.

To avoid these problems we will instead consider a trajectory as spatio-temporal curve $([x(1), y(1), 1], [x(2), y(2), 2], [x(3), y(3), 3], \ldots, [x(t), y(t), t])$. We compute the curvature of this curve which is a function of time by using a one-dimensional version of the quadratic surface fitting procedure described by Besl and Jain.[14] The curvature, κ, is defined as follows:

$$\kappa = \frac{\sqrt{A^2 + B^2 + C^2}}{((x')^2 + (y')^2 + (t')^2)^{3/2}} \tag{1}$$

where

$$A = \begin{vmatrix} y' & t' \\ y'' & t'' \end{vmatrix}, \quad B = \begin{vmatrix} t' & x' \\ t'' & x'' \end{vmatrix}, \quad \text{and} \quad C = \begin{vmatrix} x' & y' \\ x'' & y'' \end{vmatrix}.$$

The notation $|\cdot|$ denotes the determinant. We use the discrete approximation to compute the derivatives, for example, $x'(t) = x(t) - x(t-1)$ and $x''(t) = x'(t) - x'(t-1)$. Since we assume Δt to be constant, t' will equal 1, and t'' will be 0.

A number of pre-processing steps can be used to improve the detection of cycles. The curvature function exhibits large and narrow impulses at points of sudden changes on the trajectory. These impulses contain large high frequency components that may interfere with the detection of cycles, and it would be beneficial to be suppressed. A median filter is particularly suitable filter for this task since it can suppress narrow impulses while preserving smoother regions of the curvature. In our work the first step is to suppress narrow impulses using a conditional median filter[15] which can better preserve the shape of the curvature function while suppressing the large and narrow impulses. This filter performs median filtering only on samples where the absolute value of the difference between the sample and the corresponding median exceeds a threshold. With this strategy smooth signal regions remain intact, but sufficiently narrow and large impulses are suppressed. The second pre-processing step is to remove the DC component of the curvature in order to avoid the zero frequency impulse. We subtract the average value of the curvature function from the original curvature function before we perform the Fourier transform. The third step is to compute the autocorrelation of the curvature. If the motion is cyclic there will be some self-similarity within the curvature function which becomes more evident in the autocorrelation function. Finally the Fourier transform of the autocorrelation is used to detect the presence of cycles and the period of the cyclic motion. A large impulse will occur on the frequency axis of the Fourier magnitude plot at the fundamental frequency of the cycles that are present. Smaller impulses may also be present (harmonics) at integer multiples of the fundamental.

This approach for detecting cycles is simpler than one that uses curvature scale space, because scale space approach essentially matches portions of scale space to find repeated patterns of curvature for periodicity, which is time consuming. Also, our approach can detect periodicity not evident in the spatial domain because of the presence of uncorrelated noise. It is also computationally efficient because the Fourier transform can be computed via the FFT (Fast Fourier Transform) algorithm.

4. EXPERIMENTS

In our experiments we used the FFT algorithm to compute the Fourier transform. To achieve sufficient

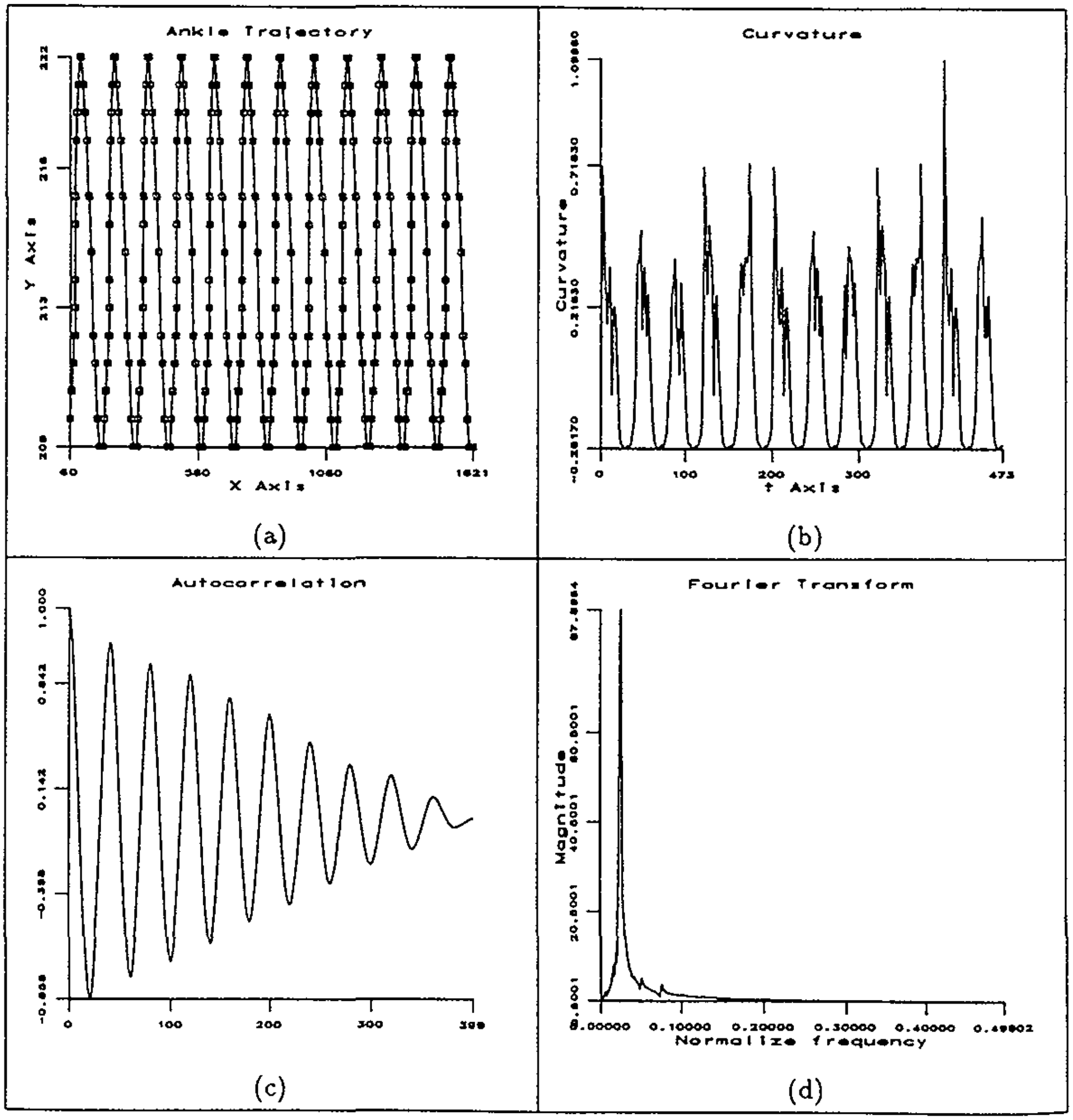

Fig. 1. Results for the synthetic walking person trajectory (obtained from a program by Cutting). (a) The trajectory of the right ankle point (480 frames). (b) The curvature function of (a). (c) The autocorrelation function of (b). (d) The magnitude of the Fourier Transform of the autocorrelation function.

frequency resolution the data array to be transformed was padded with zeros to become of length 2048 samples and a 2048-point FFT was used. It should be noted that computation of the autocorrelation function prior to transformation is not necessary because by the Wiener-Khintchine theorem the Fourier transform of the autocorrelation of a signal is the same as it's energy spectral density (Fourier magnitude squared). However, in order to demonstrate the effect of each pre-processing step, the autocorrelation function was still computed. In general on a Sun-4 Sparc workstation, for a sequence of 512 frames, computing the curvature function takes about 0.1 s, each pre-processing step takes less than 0.1 s, and a 2048-point FFT takes 0.4 s. Overall, it takes less than 0.8 s to process a sequence with 512 frames.

4.1. *Synthetic data*

The first experiment was performed using synthetic data obtained from a program by Cutting,[6] which generates files containing the coordinates of certain points on the body of a simulated walking person. Values are input to the program to determine factors such as hip swing and shoulder excursion, and the program uses laws of physics to determine the x and y coordinates of each point as the person walks. Feature points are at the following locations: ankle, wrist, elbow, knee (right and left), hip, shoulder (right) and head. For each cycle there are 40 instances at which coordinates are calculated, and the program outputs coordinates for 12 cycles, giving a total of 480 frames. Figure 1

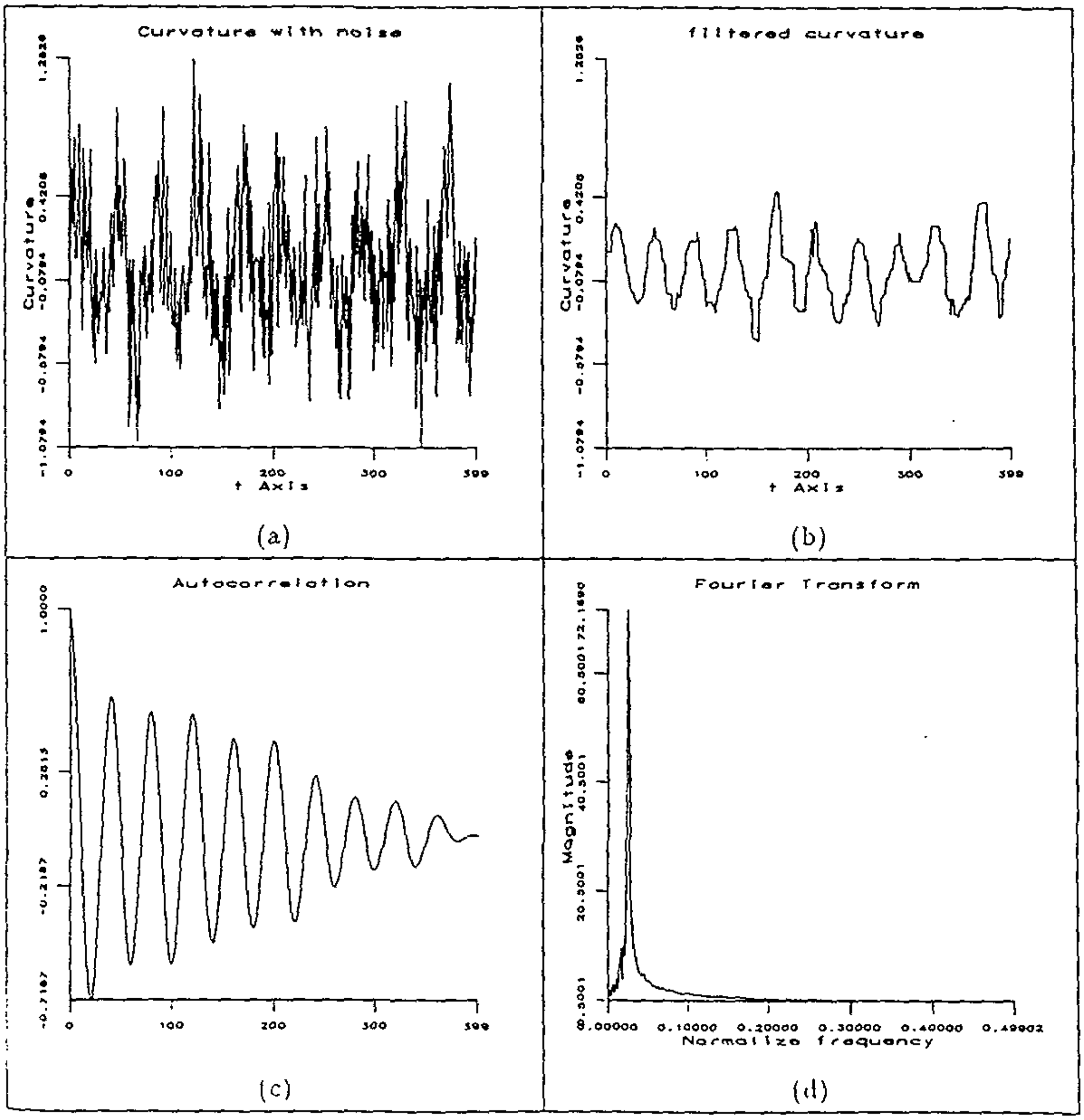

Fig. 2. Results for the noisy example. (a) The curvature function of Fig. 1(b) with noise. (b) Result after filtering through the conditional median filter. (c) The autocorrelation function of (b). (d) The magnitude of the Fourier Transform of the autocorrelation function.

shows the results for motion of the right ankle point. The 12 cycles that were created by the Cutting program are shown in Fig. 1(a), which shows the trajectory created by the x and y coordinates of the right ankle point. The curvature function is shown in Fig. 1(b). The result of the autocorrelation is shown in Fig. 1(c), and the magnitude of the Fourier transform of the autocorrelation is shown in Fig. 1(d). We can clearly see that a large impulse occurs on the frequency axis of the Fourier magnitude plot.

In order to illustrate that the proposed method can deal with the presence of uncorrelated noise, we added Gaussian noise with variance 0.1 to the curvature function of Fig. 1(b), and the resultant noisy curvature is shown in Fig. 2(a). The result after filtering through the conditional median filter is shown in Fig. 2(b). The autocorrelation of the filtered curvature is shown in Fig. 2(c), and the magnitude of the Fourier transform is shown in Fig. 2(d). We can see that the autocorrelation of the noisy filtered curvature is still very similar to the autocorrelation of the noise-free curvature [as shown in Fig. 1(c)], and a large impulse clearly occurs on the frequency axis of the Fourier magnitude plot.

Figure 3 demonstrates that the proposed method can detect and extract one cycle from a trajectory with unknown number of cycles. Figure 3(a) and (c) are the curvature functions of Fig. 1(b) with different length

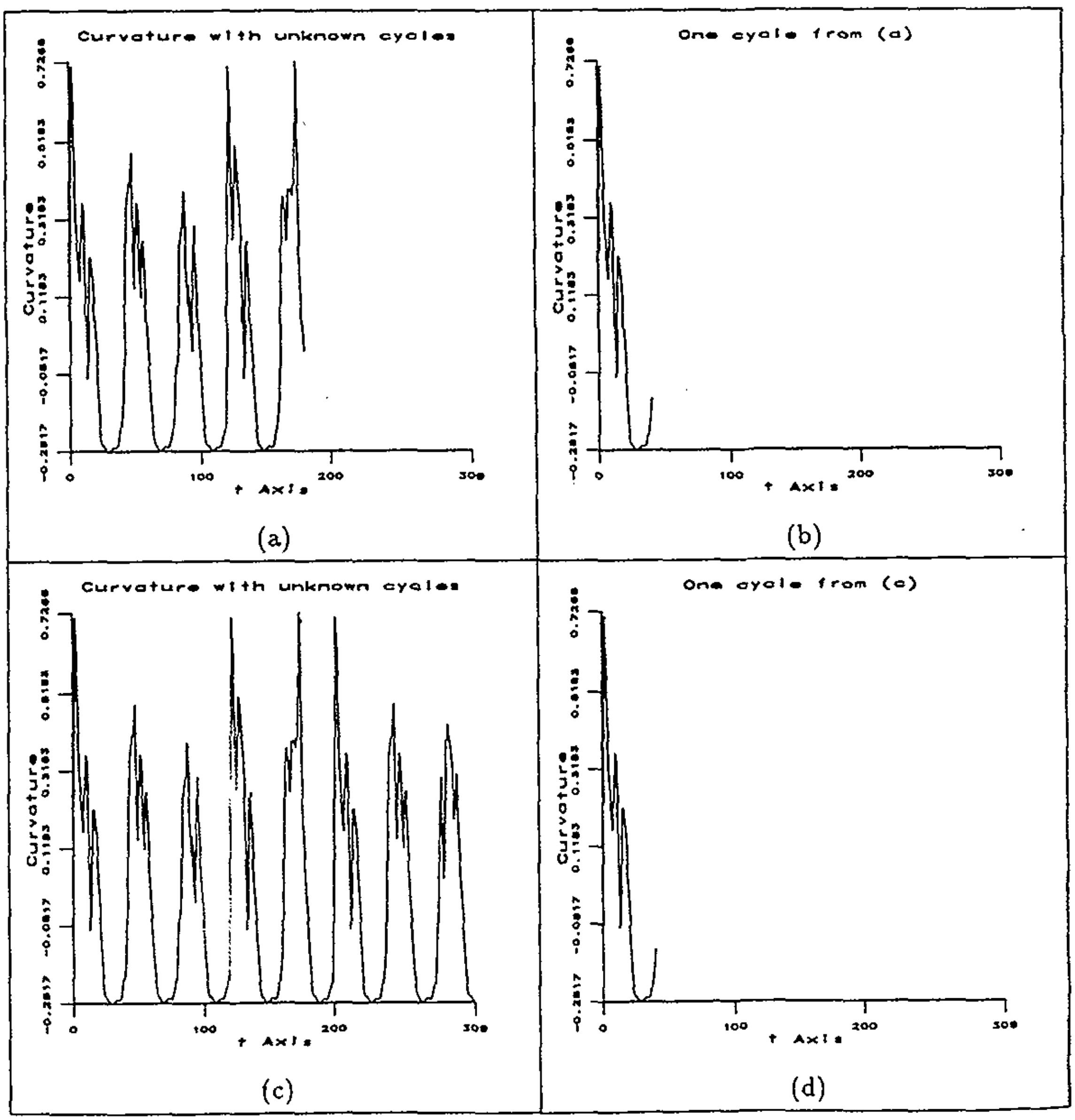

Fig. 3. Results for cycle detection and extraction. (a) The curvature function of Fig. 1(b) with unknown cycles (180 frames). (b) One cycle extracted from (a) using the proposed method. (c) The curvature function of Fig. 1(b) with unknown cycles (310 frames). (d) One cycle extracted from (c) using the proposed method.

(180 frames and 310 frames). The proposed method successfully detected and extracted the same cycle [as shown in Fig. 3(b) and (d)] for both cases.

4.2. *Real data*

The proposed method was also tested on a real walking sequence that was obtained from Goddard[11] at University of Rochester. He used the WATSMART image processing system, which is designed specifically to gather data on human gait. It consists of two cameras arranged to give a stereo view, a set of light-emitting diodes (LEDs), a calibration frame, and an IBM RT with software for processing the camera signals. The LEDs are taped to the actor, and wires string from the LEDs to the computer. As the actor moves, the system gathers data by sequentially flashing each LED and recording the location of the LED in the camera frames. The system can gather data from 8 LEDs at up to 400 frames s^{-1}. Software computes the three-dimensional location of each LED in each frame by examining the two-dimensional frame data from the two cameras. Goddard operated the system at 100 frames s^{-1}, with LEDs attached to the six proximal joints (shoulder, elbow, wrist, hip, knee, ankle) and the two most visible distal joints (wrist and ankle). Actor motion was roughly perpendicular to the calibration axis (the average of two camera axes). He then converted raw data files

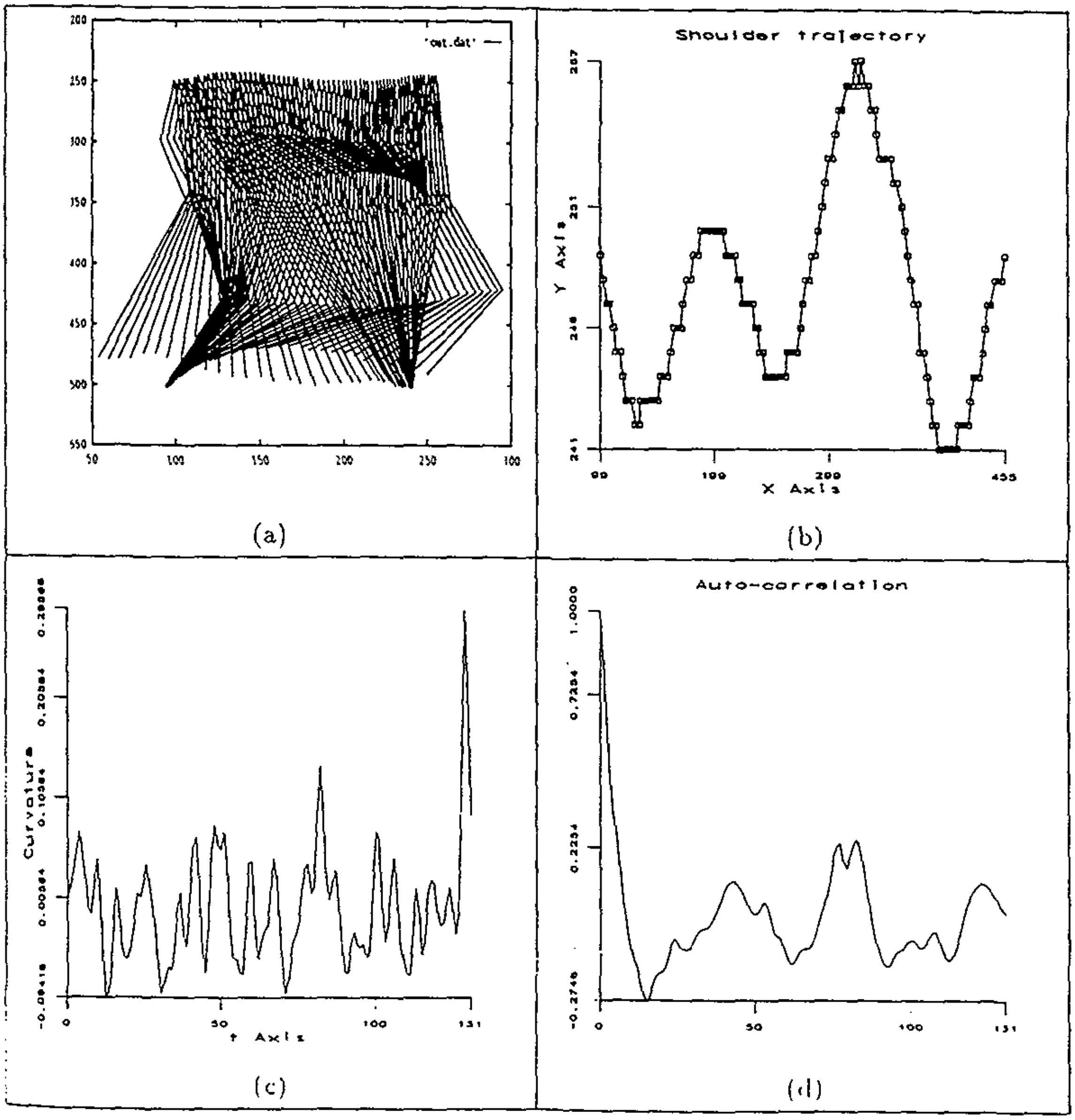

Fig. 4. Results for the real walking person. (a) The stick figure of the first to frames. (b) The trajectory of the shoulder point (132 frames). (c) The curvature function of (a). (d) The autocorrelation function of the pre-processed curvature (c). (e) The magnitude of the Fourier Transform of the autocorrelation function. (f) One cycle extracted from (c) using the proposed method. (g) The correspondent cycle of the trajectory (b).

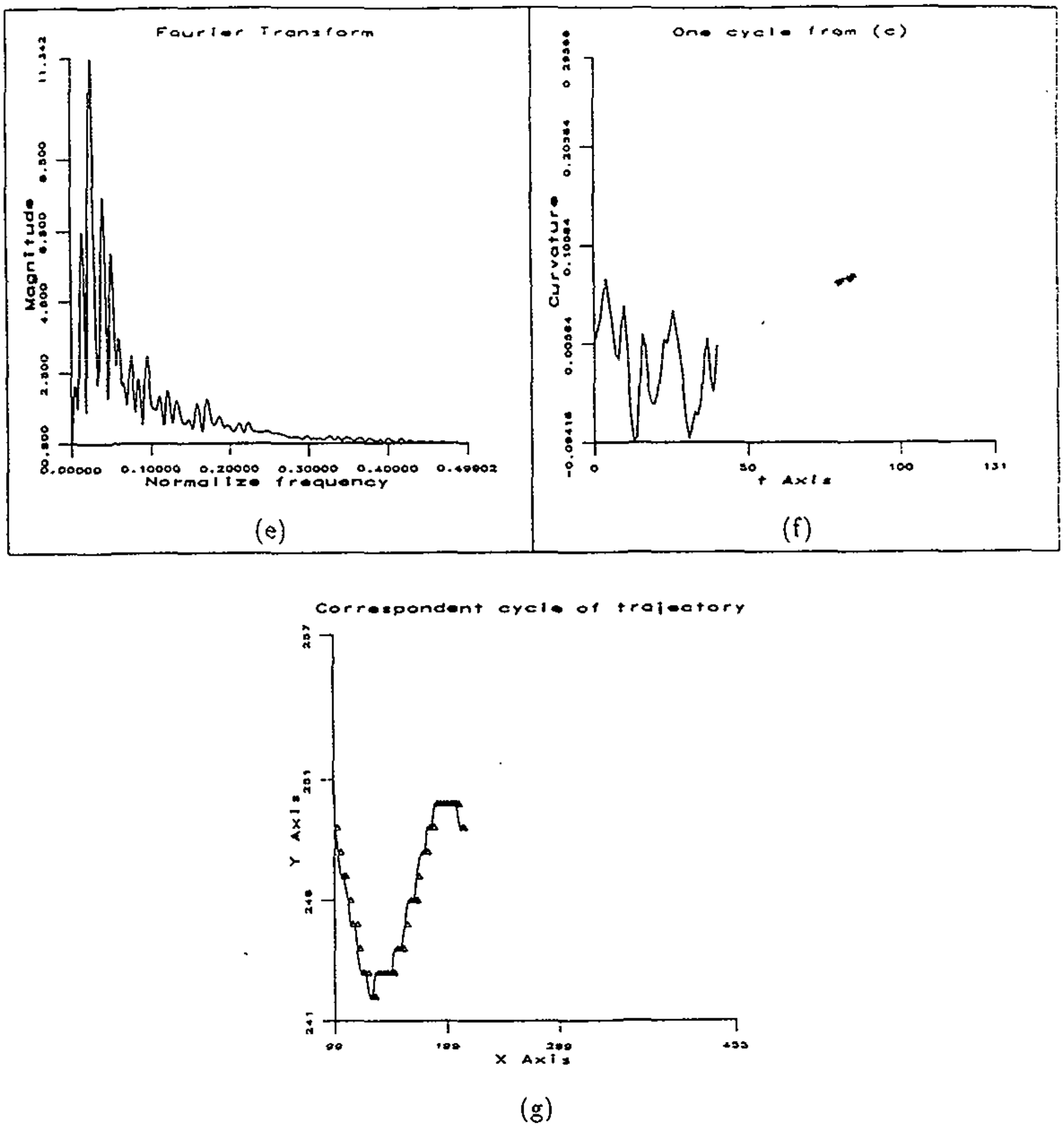

Fig. 4. (*Continued.*)

containing three-dimensional LED locations to two-dimensional by omitting the depth coordinate, which did not change much during the recordings. The raw data were converted into orientation angles for each of the limb segments and software was written to display and edit the data in order to smooth it and ensure that the starting and ending point of a cycle were identical. Finally, he resampled the data to produce 60 frames per cycle for each set of data in order to normalize the cycle time. A walking person sequence obtained from Goddard with 132 frames was tested. The stick figure of the first 60 frames is shown in Fig. 4(a). Eight points on the person's body are shown in the stick figure. The trajectory of the shoulder point is shown in Fig. 4(b). The curvature function is shown in Fig. 4(c). The auto-correlation of the pre-processed curvature function is shown in Fig. 4(d). The magnitude of the Fourier Transform of the autocorrelation function is shown in Fig. 4(e). A large impulse is clearly shown on the frequency axis of the Fourier magnitude plot. Figure 4(f) shows one cycle which is extracted using the proposed method. The correspondent cycle of the trajectory is shown in Fig. 4(g).

5. APPLICATION: MOTION-BASED RECOGNITION

One important application for cyclic motion detection is in motion-based recognition. In many cases, where an object has a fixed and predefined motion, the trajectories of several points on the object may seem

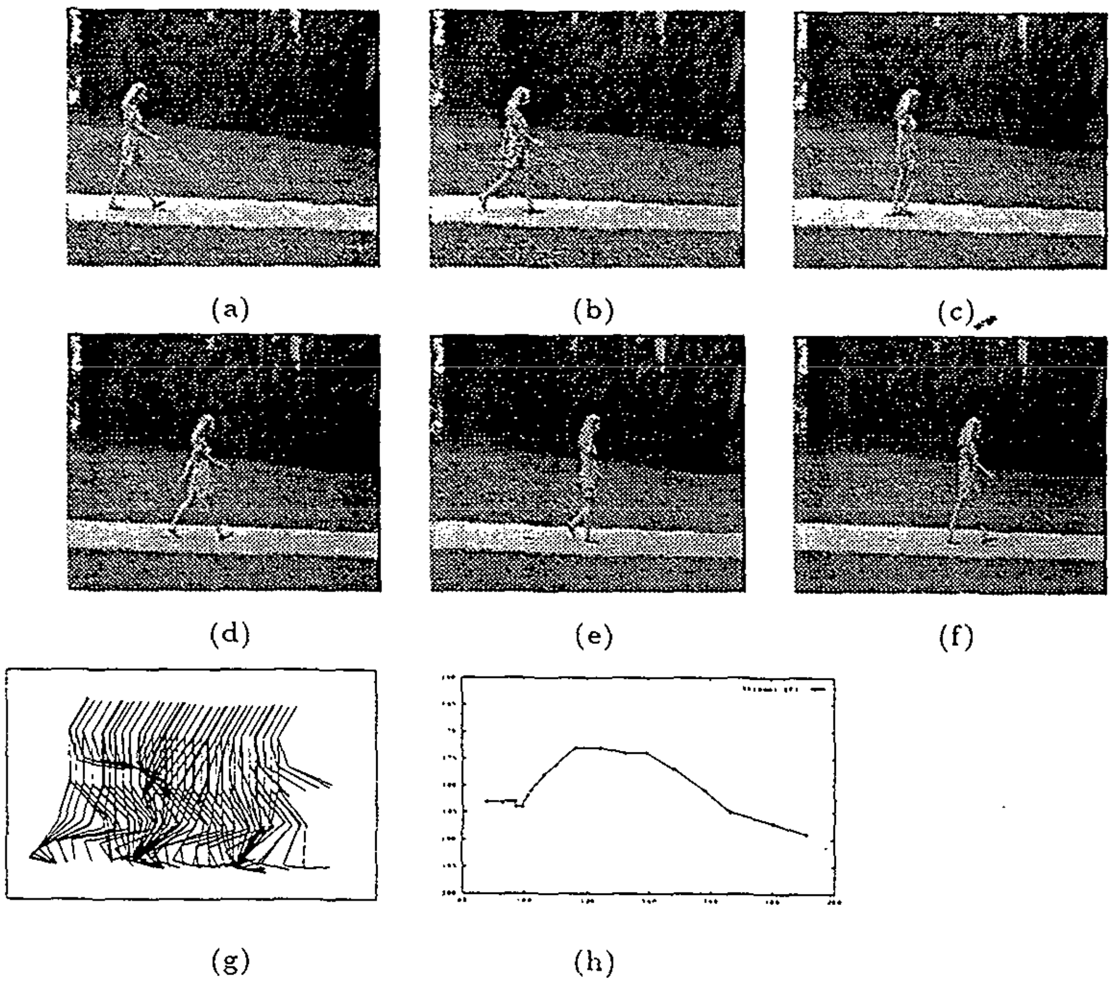

Fig. 5. Image Sequence $K1$ in which person K is walking. There are 32 frames in this sequence. (a)–(f) Frames 1, 6, 12, 18, 24 and 30 of the sequence $K1$. (g) The stick figure drawings of 9 body points. (h) Trajectory of $K1_{\text{heel}}$.

to uniquely identify the object. Therefore, it should be possible to recognize certain objects based on motion information obtained from the trajectories of representative points. Rangarajan *et al.*[4] proposed a method for matching pairs of single trajectories utilizing a scale-space representation as the basis for matching. They represent a two-dimensional trajectory as two one-dimensional functions, namely speed and direction, and convolve the one-dimensional speed and direction signals with the second derivative of Gaussian over a range of σ values to produce the two-dimensional scale-space image. They then determine the strength and polarity by applying the first derivative of the Gaussian at each zero-crossing point in the scale-space image. The strength and polarity of each zero-crossing is referred to as the *zero-crossing potential*. Match scores between the two trajectories are determined by computing the difference between their smoothed zero-crossing potentials.

The Rangarajan method assumes that the correspondence between the model trajectory and the input trajectory is known. For an object with cyclic movement, they need to store all the trajectories with difference cycles as models in order to find the correct match for an input trajectory. Since we can detect cycles in an input trajectory (assuming the object has cyclic motion), we only need to store and do the matching with one complete cycle as our model. In order to minimize the computation and problems due to the noise sensitivity of the direction function, we use one cycle of the filtered curvature for the matching algorithm, instead of speed-direction as Rangarajan did. The modified matching algorithm is summarized as following:

(1) Compute the curvature signal, $\kappa[t]$, from one complete cycle of the input trajectory [using equation (1)], and filter it through a conditional median filter.

(2) Generate the curvature scale-space image by convolving the filtered curvature signal with the second derivative of the Gaussian over a range of σ values, and locate the curvature zero-crossings by scanning the

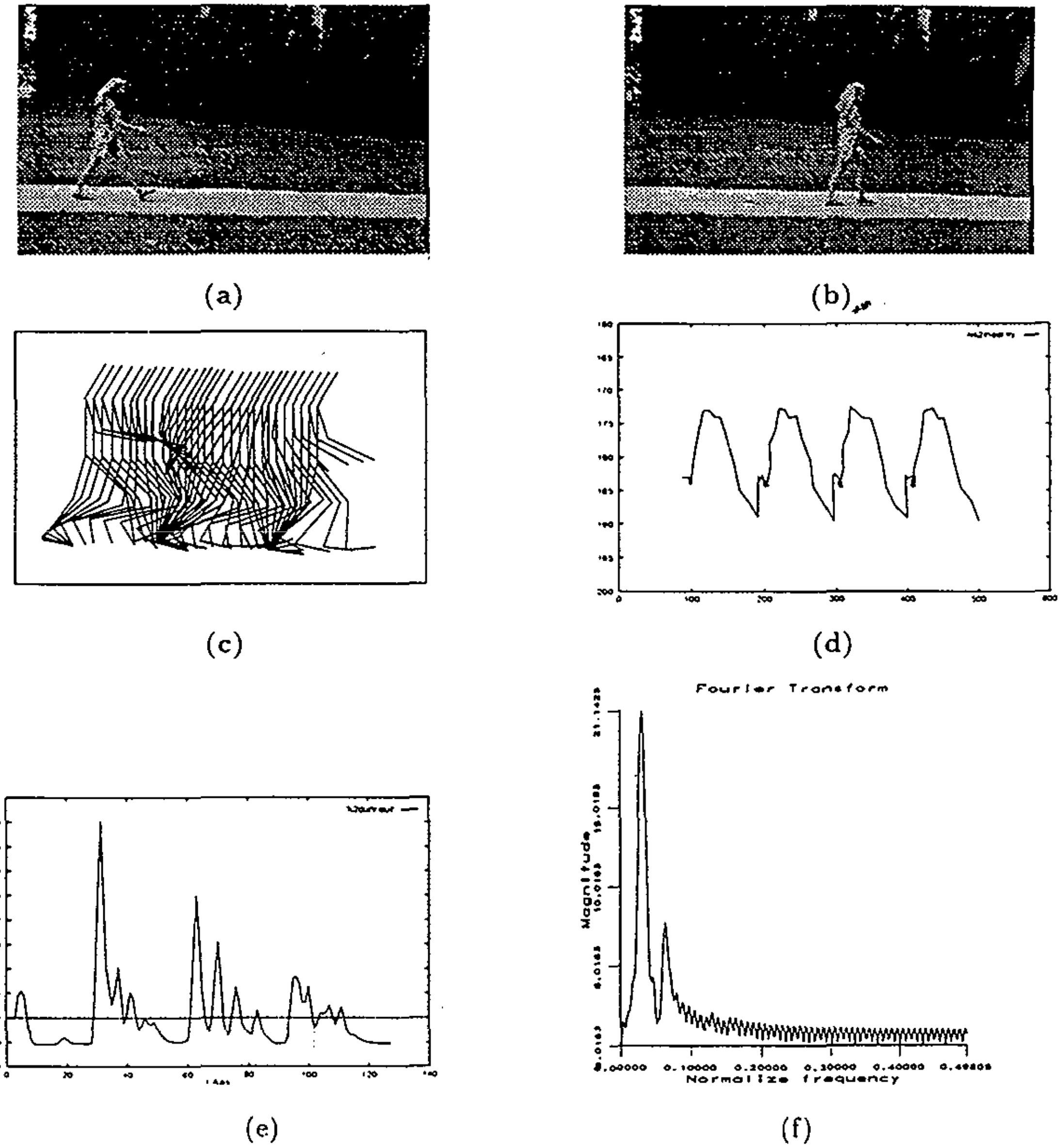

Fig. 6. Image Sequence K^2 in which person K is walking. (a), (b) First and last frames of the sequence $K2$. (c) The stick figure drawings of 9 body points. (d) Trajectory of K^2_{heel}. There are 128 frames which are obtained by repeating the original sequence four times. (e) The curvature function. (f) The magnitude plot of the Fourier Transform of the pre-processed curvature function. The proposed method extracted 32 frames as one cycle.

scale-space image and testing the values in a neighborhood around each point.

(3) Determine the strength and polarity by applying the first derivative of the Gaussian to the curvature $\kappa[t]$ (i.e. $\kappa[t]*(-t/\pi\sigma^2)e^{-t^2/2\pi\sigma^2}$) at each zero-crossing point in the scale-space image. The strength $(|\kappa[t]*(-t/\pi\sigma^2)e^{-t^2/2\pi\sigma^2}|)$ and polarity (sign) of each zero-crossing is referred to as the *zero-crossing potential*. This step produces a two-dimensional array, $\beta_\kappa[t,\sigma]$, containing the zero-crossing potentials at each point. In this array, points which are not zero-crossings will hold a zero value.

(4) Diffuse the zero-crossing potentials β_κ using a two-dimensional Gaussian mask with sigma equal to one, and store the result in array γ_κ.

(5) Scale the value in γ_κ by the scaling factors $\Sigma\Sigma\alpha_\kappa[t,\sigma]/\Sigma\Sigma\gamma_\kappa[t,\sigma]$, where $\alpha_\kappa[t,\sigma]$ is the diffused zero-crossing potentials for one complete cycle of the model trajectory.

(6) Perform an element by element subtraction of the α and γ arrays, and store the result in array ε_κ.

(7) Compute the match score as $1-[\Sigma\Sigma|\varepsilon_\kappa(t,\sigma)|/2*|\Sigma\Sigma\alpha_\kappa(t,\sigma)|]$.

A perfect match between trajectories will produce a match score of 1.

To demonstrate this application, the walking sequences of persons K and W shown in Figs 5, 6 and 7 were used. We videotaped a person, K walking at two different times, and generated two distinct image sequences, $K1$ and $K2$. We also videotaped another person, W, and generated a single image sequence, $W1$. From each image sequence, we produced a set of trajectories by manually tracking nine body points. There are 32 frames in each sequence. Figure 5(a) through (f) show frames 1, 6, 12, 18, 24 and 30 of sequence $K1$. Figure 5(g) shows the stick figure of the nine body points. Figure 5(h) is the trajectory of the *left heel* point, which is used as the model.

In order to get longer sequences with different number of cycles, we generated the sequence K^2 with 128 frames by repeating the distorted version of the original $K2$ sequence four times, and the sequence W^1 with 256 frames by repeating the distorted version of the original $W1$ sequence eight times. The distortion was introduced by adding some random noise after the first cycle of each sequence to avoid the precise replication.

Figure 6(a) and (b) show the first and last frames of sequence $K2$. Figure 6(c) shows the stick figure. Figure 6(d) is the trajectory of the *left heel* point. The curvature of the trajectory is shown in Fig. 6(e). The magnitude plot of the Fourier Transform of the pre-

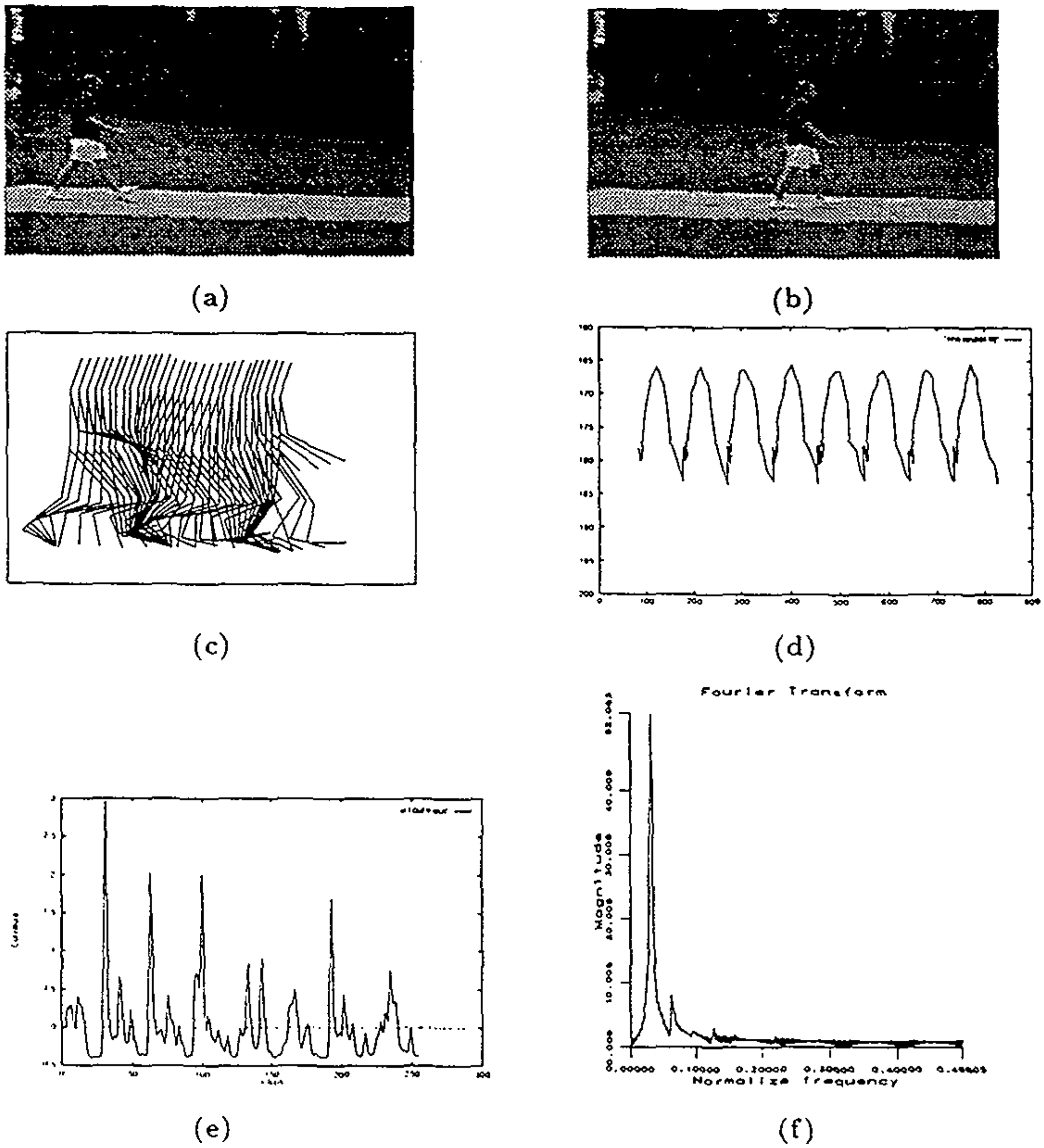

Fig. 7. Image Sequence W^1 in which person W is walking. (a), (b) First and last frames of the sequence $W1$. (c) The stick figure drawings of nine body points. (d) Trajectory of W^1_{heel}. There are 256 frames which are obtained by repeating the original sequence eight times. (e) The curvature function. (f) The magnitudue plot of the Fourier Transform of the pre-processed curvature function. The proposed method extracted 32 frames as one cycle.

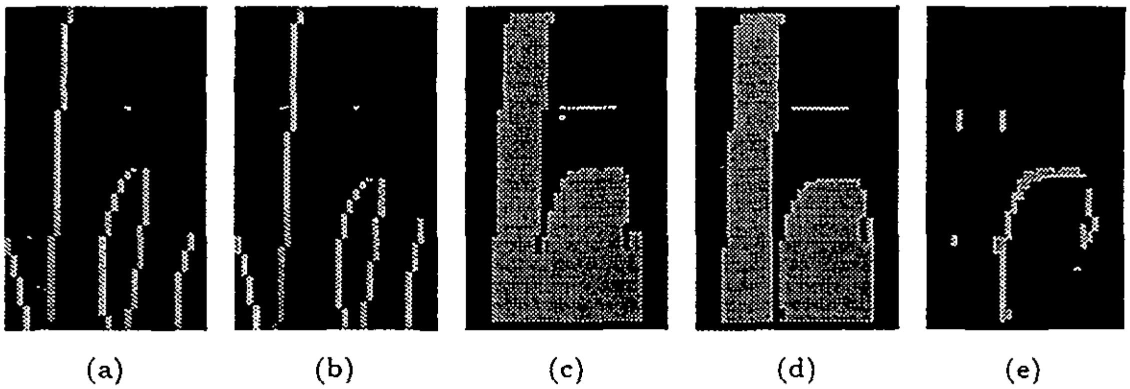

Fig. 8. The matching results for $K1$ and $K2$. (a) The zero-crossing potential of the curvature scale-space of trajectory $K1$. (b) The zero-crossing potential of the curvature scale-space of trajectory $K2$. (c) The diffused version of (a). (d) The diffused version of (b). (e) The difference picture between (c) and (d). The match score between $K1$ and $K2$ is 0.836.

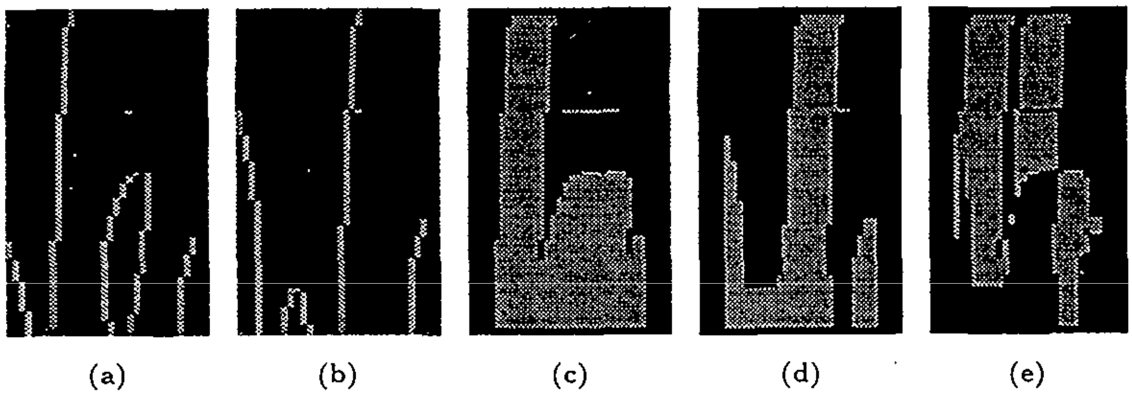

Fig. 9. The matching results for $K1$ and $W1$. (a) The zero-crossing potential of the curvature scale-space of trajectory $K1$. (b) The zero-crossing potential of the curvature scale-space of trajectory $W1$. (c) The diffused version of (a). (d) The diffused version of (b). (e) The difference picture between (c) and (d). The match score between $K1$ and $W1$ is 0.137

processed curvature function is shown in Fig. 6(f). The proposed method extracted 32 frames as one cycle, which is detected correctly.

Figure 7(a) and (b) show the first and last frames of sequence $W1$. Figure 7(c) shows the stick figure. Figure 7(d) is the trajectory of the *left heel* point. The curvature of the trajectory is shown in Fig. 7(e). The magnitude plot of the Fourier Transform of the pre-processed curvature function is shown in Fig. 7(f). The proposed method also extracted 32 frames as one cycle.

The matching results for sequences $K1$ and $K2$ using one complete cycle are shown in Fig. 8. Figure 8 (a) and (b) are zero-crossing potentials of the curvature scale-space of trajectories $K1$ and $K2$, and Fig. 8(c) and (d) are the diffused version of the zero-crossing potential. The difference picture between Fig. 8(c) and (d) is shown in Fig. 8(e). The match score between $K1$ and $K2$ is 0.836. (For a perfect matching the match score should be 1.)

The matching results for sequences $K1$ and $W1$ using one complete cycle are shown in Fig. 9. Figure 9(a) and (b) are zero-crossing potentials of the curvature scale-space of trajectories $K1$ and $W1$, and Fig. 9(c) and (d) are the diffused version of the zero-crossing potential. The difference picture between Fig. 9(c) and (d) is shown in Fig. 9(e). The match score between $K1$ and $W1$ is 0.137, which is low enough to declare a mismatch.

It is clear that the cyclic motion detection is helpful in reducing the overhead of the motion-based recognition.

6. CONCLUSIONS

In this paper, we presented a method for cyclic motion detection using autocorrelation and Fourier Transform techniques. We represent a two-dimensional trajectory as a one-dimensional signal: curvature, which is a function of time. Cycles are detected successfully in the frequency domain by using the Fourier Transform of the pre-processed curvature signal of the trajectory. The proposed method was tested on some synthetic data and real data of walking person. We also demonstrated an application, motion-based recognition, for the cycle detection method.

REFERENCES

1. N. H. Goddard, The perception of articulated motion: Recognizing moving light displays. Ph.D. thesis, University of Rochester (1992).
2. G. Johansson, Visual perception of biological motion, *Scientific Am.* 232, 76–89 (June 1975).
3. A. P. Witkin and J. M. Tenenbaum, On perceptual organization, in (A. P. Pentland, ed.) *From Pixels to Predicates: Recent Advances in Computational and Robot Vision* pp. 149–160. Norwood, NJ (1986).
4. K. Rangarajan, B. Allen, and M. Shah, Matching motion trajectories, *Pattern Recognition* 26, 595–610 (July 1993).
5. J. T. Todd, Visual information about rigid and nonrigid motion: a geometric analysis, *J. Experimental Psychology: Human Perception and Performance* 8, 238–252 (1982).
6. J. E. Cutting, A program to generate synthetic walkers as dynamic point-light displays, *Behaviour Research Methods and Instrumentation* 10(1), 91–94 (1977).
7. N. H. Goddard, The interpretation of visual motion: recognizing moving light displays, in *Proc. Workshop on Visual Motion*, (March 20–22), Irvine CA, pp. 212–220 (1989).
8. S. A. Engel and J. M. Rubin, Detecting visual motion boundaries, in *Workshop on Motion: Representation and Analysis* (May 7–9), Charleston, SC. *IEEE Publ.* 86CH2322-6 (1986).
9. M. Allmen and C. Dyer, Cyclic motion detection using spatiotemporal surfaces and curves, in *Proc. 10th Int. Conf. on Pattern Recognition*, pp. 365–370. IEEE Computer Society (June 1990).
10. D. Koller, N. Heinze and H. Nagel, Algorithmic characterization of vehicle trajectories from image sequences by motion verbs, in *Conference on Computer Vision and Pattern Recognition*, pp. 90–95 (June 1991).
11. D. Hogg, *Parallel Architecture and Computer Vision*, Finding a known object using a generate and test strategy, pp. 119–132. Oxford Science Publications (1988).
12. T. Tsotsos, J. Mylopous, D. Convey and S. Zucker, A framework of visual motion understanding, *IEEE PAMI*, pp. 563–573 (November 1980).
13. R. Polana and R. Nelson, Detecting activities, in *Proc. of Computer Vision and Pattern Recognition*, pp. 2–7 (June 1993).
14. P. J. Besl and R. C. Jain, Invariant surface characteristics for 3D object recognition in range images, *CVGIP* **33**, 33–80 (1986).
15. T. Kasparis, N. Tzannes and Q. Chen, Detail preserving adaptive median filters, *J. Electronic Imaging* 1(4), 358–364 (1992).

About the Author—PING-SING TSAI was born in Taipei, Taiwan, R.O.C. He received the B.S. degree in information and computer engineering from Chung-Yuan Christian University in 1985. He is pursuing a Ph.D. in computer science at the University of Central Florida.

About the Author—MUBARAK SHAH received his B.E. degree in 1979 in electronics from Dawood College of Engineering and Technology, Karachi, Pakistan with the highest grades in the whole University, and was awarded a five year Quad-e-Azam (Father of Nation) scholarship for his Ph.D. He spent 1980 at Philips International Institute of Technology, Eindhoven, The Netherlands, where he completed, a E.D.E. diploma. Dr Shah received his M.S. and Ph.D. degrees both in Computer Engineering from Wayne State University, Detroit, Michigan, respectively in 1982 and 1986. Since 1986 he has been with the University of Central Florida, where he is currently as Associate Professor of Computer Science. Professor Shah has published his research in motion, edge and contour detection, multisensor fusion, and shape from shading and stereo. He is an associate editor of *Pattern Recognition* journal. He has served on the program committees of the *Machine Vision and Robotics Conference*, and the *ACM Computer Science Conference*. He has chaired sessions at several conferences, including the *International Conference on Pattern Recognition*, and the *SPIE Applications of AI Conference*.

About the Author—KATHARINE KEITER was born in Chapel Hill, NC. She received a Bachelors degree in computer engineering from the University of Central Florida in 1993 and is presently working as an engineer for Motorola's Government and Systems Technology Group in Scottsdale, AZ.

About the Author—TAKIS KASPARIS received the Diploma of electrical engineering from the National Technical University of Athens, Greece, in 1980, and the MEEE and Ph.D. degrees in electrical engineering from the City College of New York in 1982 and 1988. From 1985 until 1989 he was also an electronics consultant. In 1989 he joined the Electrical and Computer Engineering Department of the University of Central Florida, Orlando, where he is presently an assistant professor. His research interests are in non-linear signal and image processing, pattern recognition, DSP applications and electronic circuits.