Describing Motion for Recognition

Jim Little Jeffrey Boyd

Technical Report 95-??

May 1995

Laboratory for Computational Intelligence
Department of Computer Science
University of British Columbia
Vancouver, British Columbia, CANADA V6T 1Z4
email: little@cs.ubc.ca

Abstract

Our goal is to describe motion of a moving human figure in order to recognize individuals by variation in the characteristics of the motion description. We begin with a short sequence of images of a moving figure, taken by a static camera, and derive dense optical flow data for the sequence. We determine a range of scale-independent features of each flow image as a whole, ranging from the motion of the centroid of the moving points (assuming a static background), to the integral of the torque relative to the centroid.

We then analyze the periodic structure of these sequences. All elements are multiples of the fundamental period of the gait, but they differ in phase. The phase is time-invariant, since it is independent of the sampling period. We show that there are several regularities in the phase differences of the signals. Moreover, some scalar measures of the signals may be useful in recognition.

The representation is model-free, and therefore could be used to characterize the motion of other non-rigid bodies. [1]

[1]This research was supported by the Natural Sciences and Engineering Research Council of Canada and the Networks of Centres of Excellence Institute for Robotics and Intelligent Systems, Project A-1.

1 Introduction

The pattern of motion in human gait has been studied in kinesiology, as well as in computer vision, as images of moving light displays. Kinesiology describes the forward propulsion of the torso by the legs, the ballistic motion of swinging arms and legs and the connections among these motions[LW82, PB81]. The literature on moving light displays provides an introduction to modeling moving figures[CS94]. There are two main theories about recovery of gait. The first is model-based: the 3-D structure of the model is recovered from the moving lights and then interpreted. The second theory emphasizes determining features of the motion fields without structural reconstruction. Recent theoretical work demonstrates the recoverability of affine motion characteristics from sequences[SD94].

There have recently been several attempts to recover characteristics of gait from image sequences, without the aid of annotation via lights[BDP+94, NA94b, NA94a, PN93, PN94, PN95].

Niyogi and Adelson emphasize segmentation over long sequence of frames [NA94b, NA94a]. Their technique relies on recovering the boundaries of moving figures in the XT[NA94a]and recently [NA94b] XYT spatiotemporal solids, followed by fitting deformable splines to the contours. These splines are the elements of the articulated nonrigid model whose features aid recognition.

Polana and Nelson[PN93] characterize the temporal texture of a moving figure by "summing the energy of the highest amplitude frequency and its multiples". They use Fourier analysis. The results are normalized with respect to total energy so that the measure is 1 for periodic events and 0 for a flat spectrum. Their input is a sequence of 128 frames, each 128x128 pixels. Their analysis consists of determining the normal flow, thresholding the magnitude of the flow, determining the centroid of all "moving" points, and computing the mean velocity of the centroid. The motion in XYT of the centroid determines a linear trajectory. They use as motion signals reference curves that are "lines in the temporal solid parallel to the linear trajectory".

Polana and Nelson's more recent work[PN94, PN95] emphasizes the spatial distribution of energies around the moving figure. They compute spatial spatistics in a coarse mesh and derive a vector describing the relative magnitudes and periodicity of activity in the regions, over time. Their experiments demonstrate that the values so derived can be used to discriminate among differing activities.

Our work, in the spirit of Polana and Nelson, does not fit a model for the

moving figure but instead describes the variation over time of a set of features of a dense flow distribution.

2 Method

Image sequences are gathered while the subject is walking laterally before a static camera and processed offline. Motion stabilization could be accomplished a tracking system such as [LK93] that pursues a moving object. Only one subject moves in the field of view. Figure 1 shows the viewing conditions in our lab. Figure 2 shows eight images from a sequence of 18 (from 08 to 25) for this subject. Only one field in a frame is used, because of the substantial difference between the two fields caused by motion. This reduces the vertical resolution, and spatial averaging reduces the horizontal resolution, resulting in 256x256 images. The system processes subimages that are manually selected from the original image. Since the camera does not move, selecting the subimage only speeds up motion processing and does not affect the results.

2.1 Describing Flow

Unlike other methods, we use dense optical flow fields, generated by minimizing the sum of absolute differences between image patches [BLP89]. The algorithm is sensitive to brightness change caused by fluorescent lights, reflections, or shadows, so we further process the images to assess the total temporal variation over a region. Each pair of images is filtered with a Gaussian of $\sigma = 3$ to determine a local average brightness. We ignore motion where the average local brightness change is below a threshold. The algorithm determines displacements between pairs of images, both forward and backward in time, which then must pass a symmetric validation test. The validated flow values are then thresholded at 1.0 pixels per frame, yielding a *blob*, a set of moving points, together with their flow values. Figure 3 shows the blobs or moving points for eight images from sequence 06.

The motion of the object is a path in 3 space; we view its projection. We assume orthographic projection for our analysis since the effects of projection are negligible. The system determines the centroid of all moving points, where each point has unit mass; these centroids arise from the mass distribution in Fig. 3. A second weighted centroid is computed where each point is "weighted" by the

Figure 1: One image from the sequence

magnitude of the motion at the point, as shown in Fig. 4. The major component in the motion is the x-component, shown in Fig. 5.

The second moments of the two distributions provide the best-fitting ellipses for the the moving points, unweighted and weighted by the magnitude of the motion at the point. The ratio of the lengths of the major and minor axes of these ellipses is a scale-invariant measure of the distribution of motion, reflecting both the position and velocity of the moving points.

Another measure that includes not only the motion but the location of the motion is the "torque" around the centroid: at each point we compute the vector connecting the point to the centroid and then the dot product of that vector and the flow at the point. To make this measure scale-invariant we scale by the square of the length of the vertical axis of the ellipse.

The descriptions include:

- the y-coordinate of the centroid

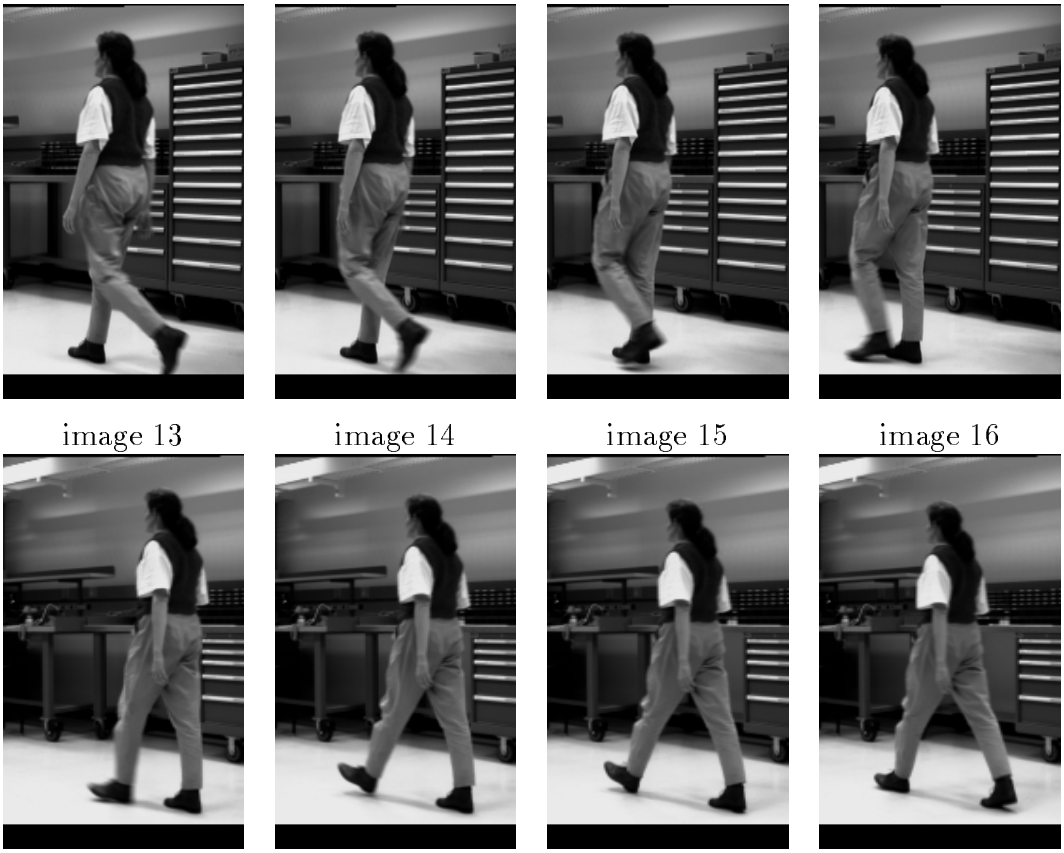

image 13 image 14 image 15 image 16

image 17 image 18 image 19 image 20

Figure 2: Image sequence 06

- the y-coordinate of the weighted centroid
- the difference of the y-coordinates of the two centroids
- the relative shape (or elongation)—the ratio of the length of the major axis to the minor axis of an ellipse
- the average velocity
- the distribution of velocities, weighted by magnitude and unweighted.
- the angle of the major axis of the ellipse with vertical

We generate several signals from the optical flow field. The centroid of the moving regions is an important reference point. The centroid of the moving points, weighted by the magnitude of the flow, is another. Their relative positions varies systematically over the sequence. Figure 6 displays the centroid as a box and the ellipse that fits the moving points in solid lines. The centroid of the weighted points is shown as a cross and its ellipse is shown in dashed lines.

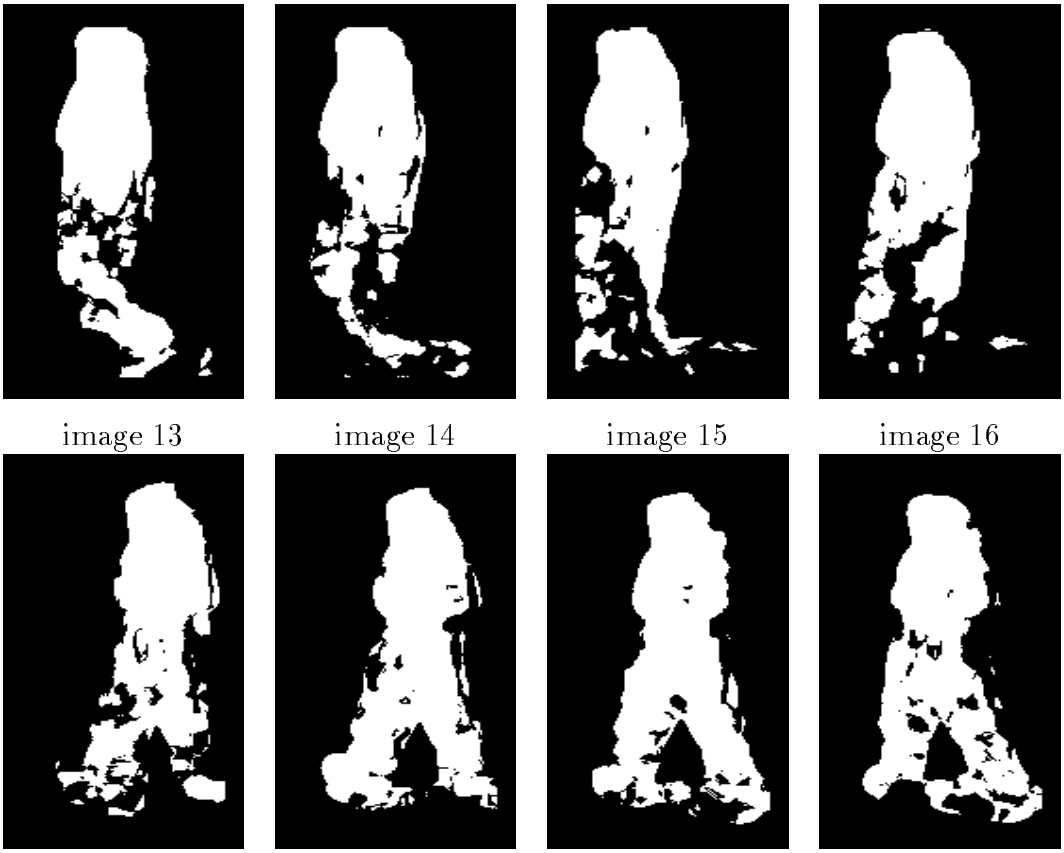

Figure 3: The moving points for sequence 06

2.2 Frequency and Phase Estimation

To estimate the frequency of the short time series, we use Barrodale and Erickson's least-squares linear prediction (LSLP) method and maximum entropy estimation [BE80a, BE80b]. LSLP[2] computes a set of coefficients that predicts future values of a time series, x_t, i.e. it computes m coefficents, $a_j, j = 1, .., m$, such that:

$$y_t = \sum_{j=1}^{m} a_j x_{t-j}, t = m+1, m+2, ..., n, \quad (1)$$

is an estimate of x_t. Equation (1) may be expressed as the system of linear equations:

$$B\mathbf{a} = \mathbf{y}, \quad (2)$$

[2]The following discussion is taken from [BE80a].

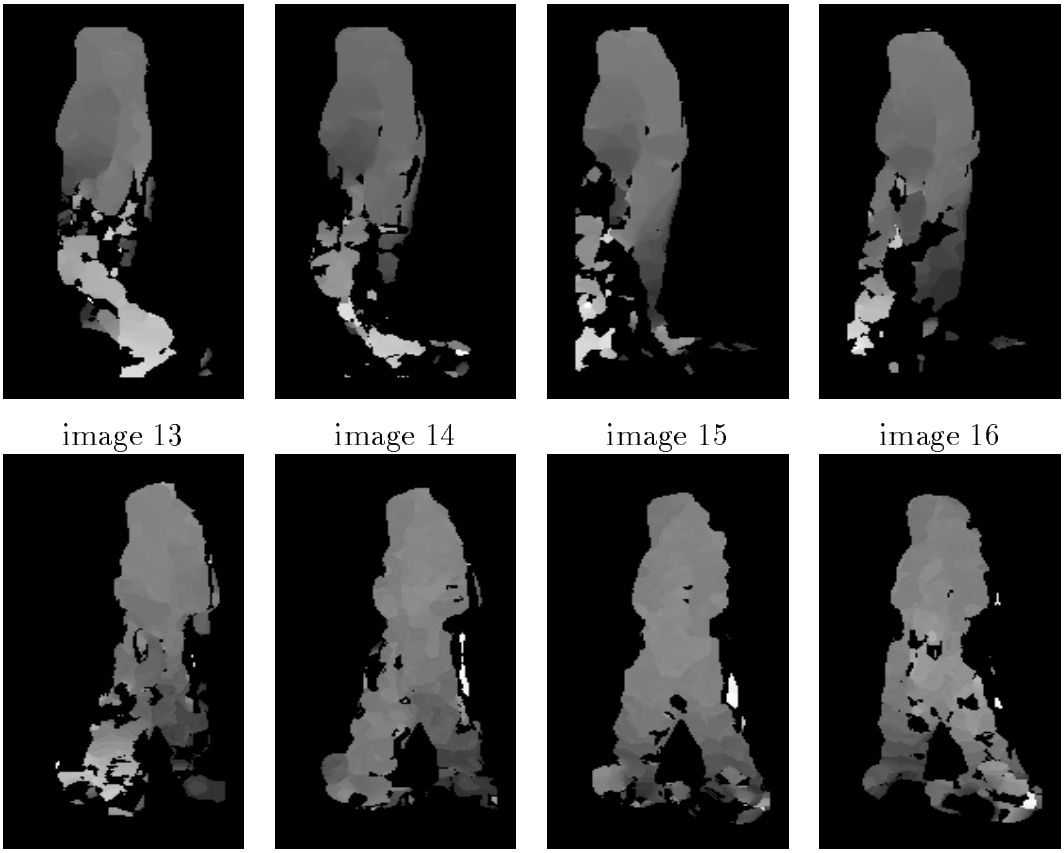

Figure 4: The magnitudes of velocities for sequence 06

where

$$B = \begin{pmatrix} x_m & x_{m-1} & \dots & x_1 \\ x_{m+1} & x_m & \dots & x_2 \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ x_{n-1} & x_{n-2} & \dots & x_{n-m} \end{pmatrix},$$

and

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_m \end{pmatrix}, \text{and } \mathbf{y} = \begin{pmatrix} x_{m+1} \\ x_{m+2} \\ . \\ . \\ . \\ x_n \end{pmatrix}.$$

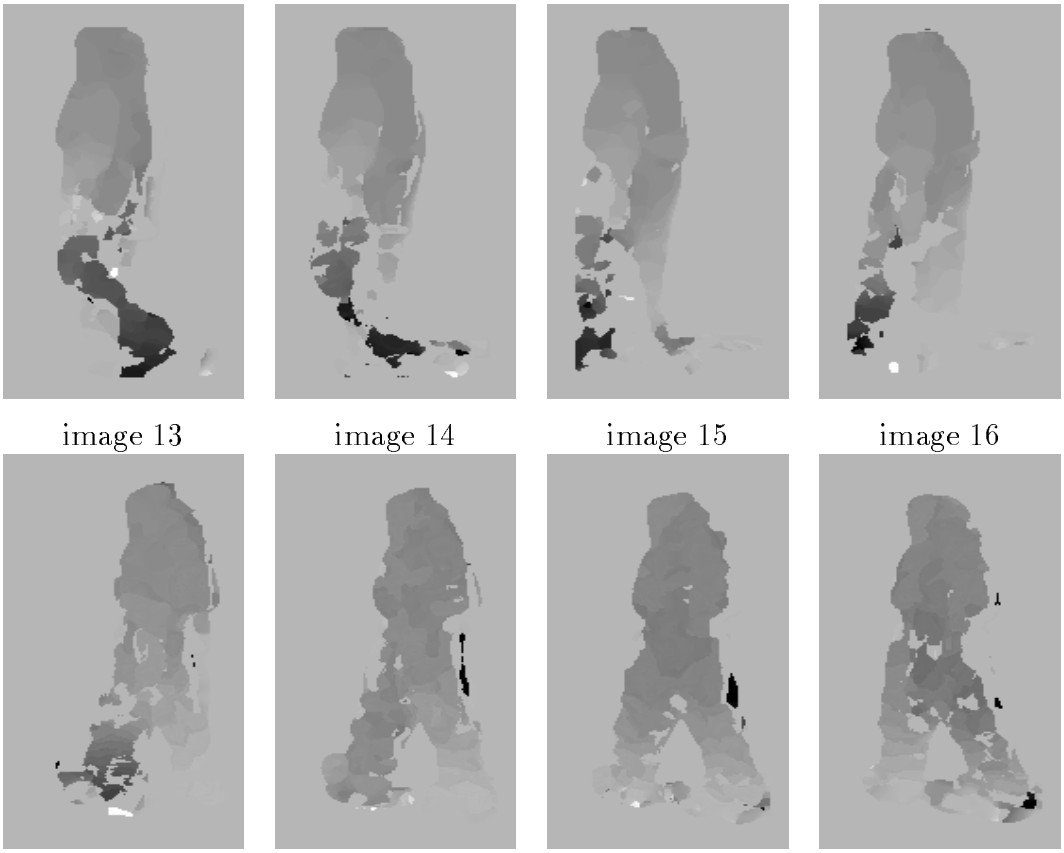

image 13 image 14 image 15 image 16

image 17 image 18 image 19 image 20

Figure 5: The x component of velocities for sequence 06

Solving for $\mathbf{a}$ from the symmetric positive definite system of equations:

$$B^T B\mathbf{a} = B^T\mathbf{y}, \tag{3}$$

gives a set of coefficents that minimizes the L_2 norm of the prediction error, i.e., the least-squares solution. Barrodale and Erickson also show variations of the method that yield coefficients optimized for reverse prediction and forward and backward prediction combined. We solve the forward and backward prediction variation using LU decomposition [PTVF92].

The coefficients provide the autoregressive parameters required for the maximum entropy spectrum of the time series using:

$$P(\omega) = \frac{P_m \Delta t}{|1 - \sum_{j=1}^{m} a_j e^{i\omega j \Delta t}|^2} \tag{4}$$

where Δt is the sampling period, $P_m = S_m/2(n-m)$, and S_m is the L_2 norm of the residuals. Five coefficients are sufficient to estimate the spectrum for the time

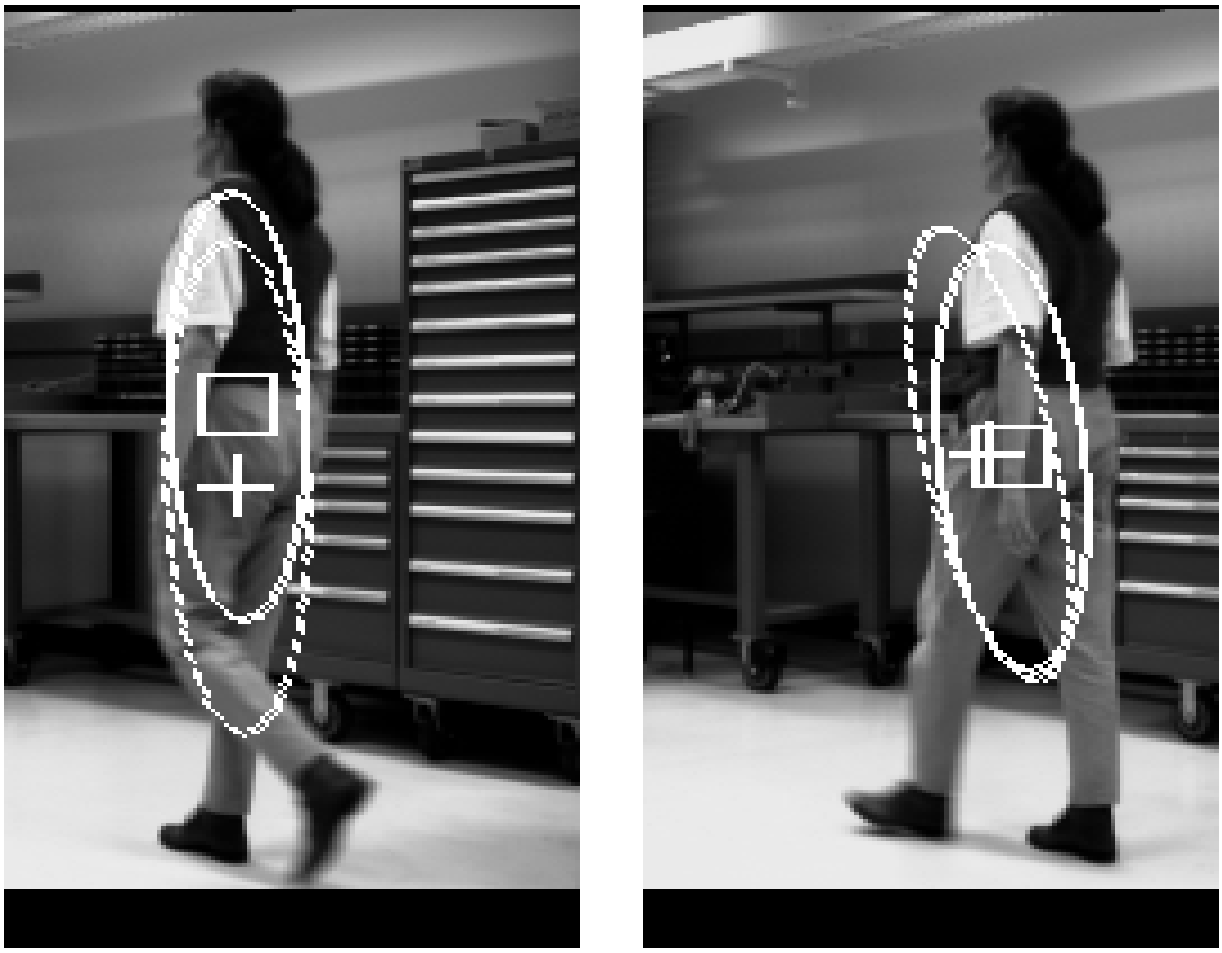

image 14 image 17

Figure 6: The centroids and fitting ellipses for images 14 and 17 in sequence 06

series considered here. We compute the spectrum from the coefficients for a set of frequency values using Equation (4), and find the frequency at which the spectrum is maximum. For example, Figure 7 shows a short time series that is the y position of the centroid of motion in an image sequence. Figure 8 is the LSLP maximum entropy spectrum based on five forward and backward prediction coefficients. The figure shows the frequency ω expressed as a ratio of f_s, the sampling frequency, where f_s is $1/\Delta t$. The spectrum shows a definite maximum that identifies the frequency of the motion. This method avoids problems associated with using discrete Fourier transforms to estimate spectra of a short time series.

Given the fundamental frequency of the time series, it is a simple matter to compute the phase of the signal. The coefficients of the Fourier representation of

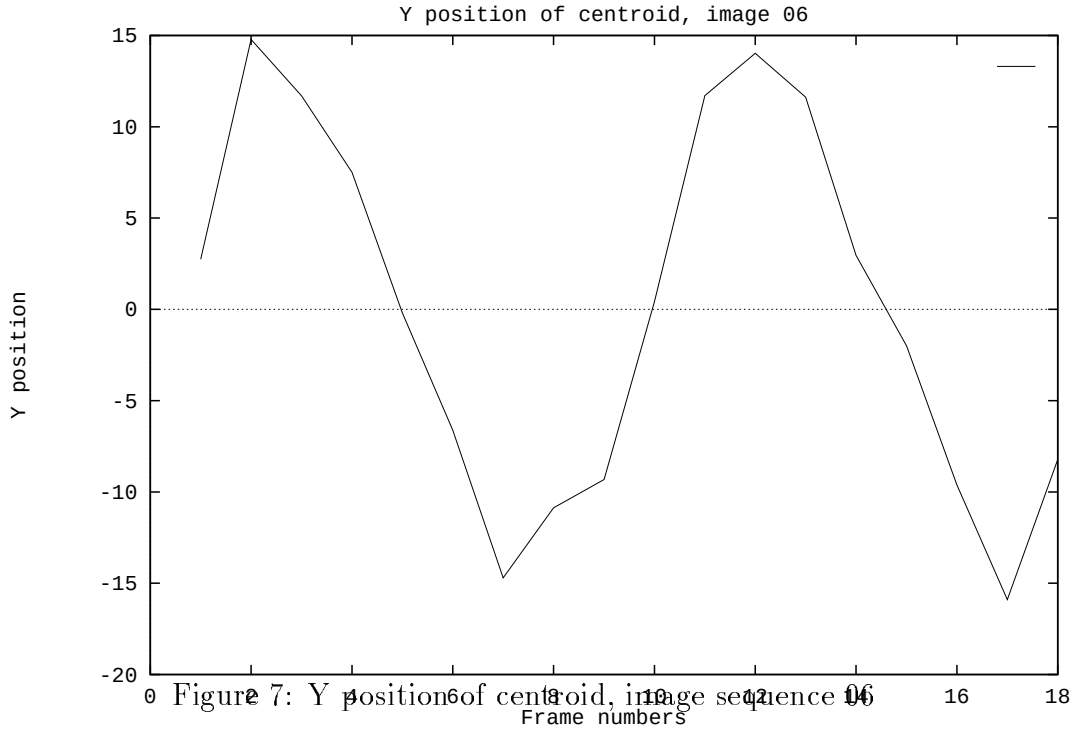

Figure 7: Y position of centroid, image sequence 06

an infinite time series are given by [OS89]:

$$X(e^{i\omega}) = \sum_{t=-\infty}^{\infty} x_t e^{-i\omega t} \tag{5}$$

Since we know the frequency of the maximum in the spectrum, w_{max}, we compute the Fourier coefficent for that frequency from the finite time series using:

$$X(e^{i\omega_{max}}) = \sum_{t=1}^{n} x_t e^{-i\omega_{max} t}. \tag{6}$$

The phase angle of the complex coefficent $X(e^{i\omega_{max}})$ gives the phase of a signal with respect to a fixed reference and allows comparison of the phase relationship between various signals from the same sequence.

3 Experimental results

We study the phase difference between the signals we identify: the y-coordinate of the centroid, the y-coordinate of the weighted centroid, and their difference. We use the first as a reference and take the difference of the other two with the first.

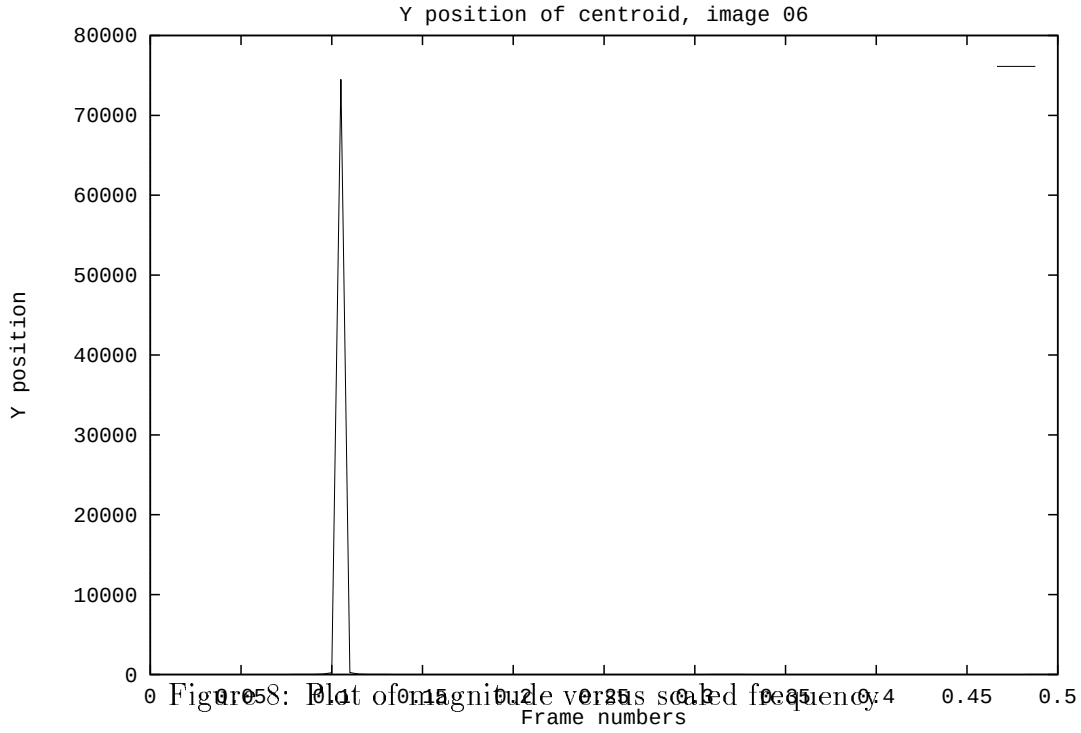

Figure 8: Plot of magnitude versus scaled frequency.

Figure 9 shows the paths of the centroids of the moving points over the entire image sequence, images 08 through image 25. Lines connect corresponding centroids (from the same image). The centroids lie on a solid curve and the centroids of the weighted points lie on a dashed curve. Unlike this figure, Fig. 7 shows the y-coordinate as a function of time.

The long-term trend may not be constant in y, but the result is such a low-frequency term that it will not appear in the frequency analysis.

The following table shows the frequency f_c of the centroid y-value, its phase ϕ_c, and the phase of the weighted points ϕ_{wc}, and their difference ϕ_{diff}. The frequency we find is the frequency of the arm-leg movement, and is twice the frequency of the full gait cycle.

Figure 9: The sequence of centroids from sequence 06, images 8 through 25, shown as paths in image space. Solid lines connect the centroids of blobs, and dashed lines connect the centroids of weighted points. Solid lines connect the two centroids from each image.

Seq.	f_c	ϕ_c	ϕ_{wc}	ϕ_{diff}
01	0.13	0.430	0.230	-.010
02	0.14	-.125	0.43	0.268
03	0.15	-.277	-.439	0.300
04	0.105	0.434	0.280	-.466
05	0.100	-.215	-.464	0.432
06	0.105	-.177	0.388	0.416
07	0.100	-.300	-.368	0.375
08	0.105	0.379	0.015	-.056

The frequency of the components of a particular gait is consistent across the various measures, as would be expected because of the periodic nature of gait. However, because the frequency is a measure of the speed of a gait, it cannot itself be used to identify a person. The first three image sequences are of a seven-year old boy, and their frequencies are higher, as expected.

There are several measures that can be extracted from this information. We use the phase change between the components of the flow. To determine the phase for the signals, we use the fundamental frequency computed from the sequence of the centroids' y-values. Its phase is the reference phase value. We then compute the difference between the phase values for each signal. The differences between the phase of centroid's y-value ϕ_c and the phase of the weighted points ϕ_{wc} is feature 1, $F1$, and the difference between the reference and ϕ_{diff} is feature 2, $F2$.

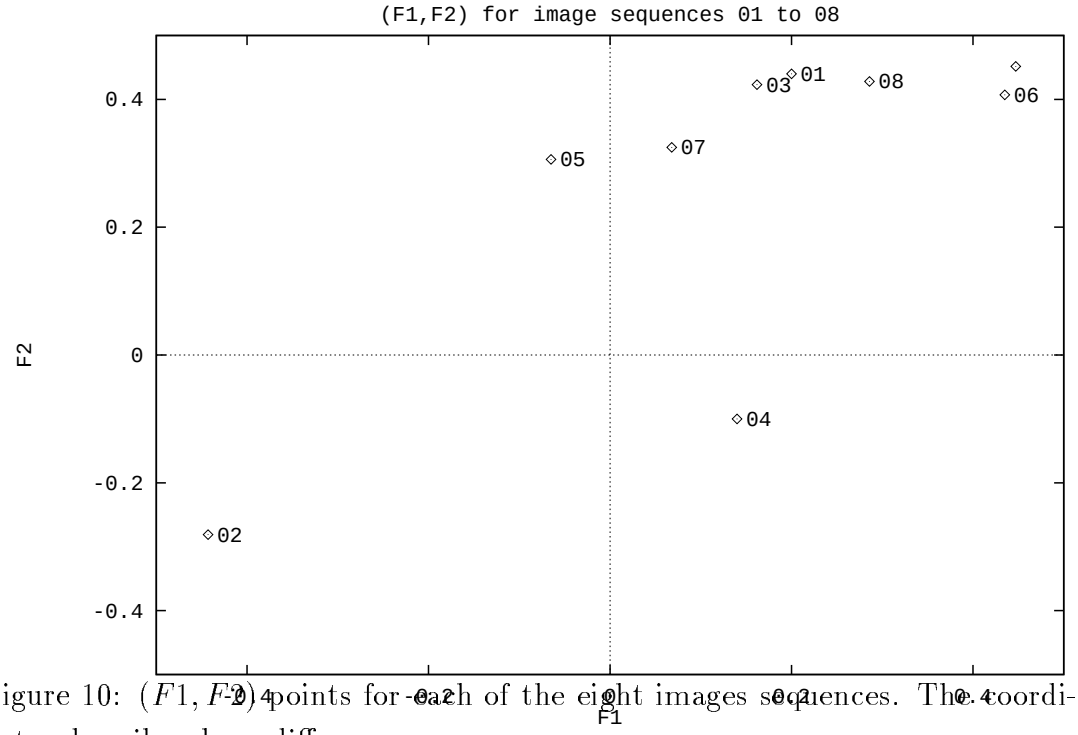

Figure 10: $(F1, F2)$ points for each of the eight images sequences. The coordinates describe phase differences.

Phase is represented as fractions of 2π between -0.5 and 0.5. The following table shows phase differences for the weighted centroid and the centroid difference.

	F1	F2
Seq.	$\phi_c - \phi_{wc}$	$\phi_c - \phi_{diff}$
01	0.200	0.440
02	-.443	-.281
03	0.162	0.423
04	0.140	0.900
05	0.249	0.353
06	0.435	0.407
07	0.068	0.325
08	0.364	0.435

Figure 10 shows these points $(F1, F2)$ in a 2-D space that is toroidal: both the $F1$ and $F2$ coordinates wrap around since they refer to phase. Figure 10 shows these points $(F1, F2)$ in a 2-D space that is toroidal: both the $F1$ and $F2$ coordinates wrap around since they refer to phase.

The following table lists the phases of the signals arising from the axes: ϕ_a; the axes of the weighted point distribution: ϕ_{wa}; and the torque: ϕ_t.

Seq.	ϕ_a	ϕ_{wa}	ϕ_t
01	-.026	-.048	-.465
02	0.157	0.169	0.300
03	0.320	0.254	-.152
04	-.223	0.060	-.095
05	0.405	0.312	?.???
06	0.478	0.413	-.068
07	0.180	0.272	-.009
08	-.022	-.065	0.457

The differences between the phase of centroid's y-value ϕ_c and the phase of ratio of the axes of the blob ϕ_a is feature 3, $F3$; the phase difference of ratio of the axes of the weighted points ϕ_{wa} is feature 4, $F4$; and the difference between the reference and the phase of the torque ϕ_t is feature 5, $F5$. The following table describes $F3, F4, F5$. Figure 11 shows these points $(F3, F4)$.

	F3	F4	F5
Seq.	$\phi_c - \phi_a$	$\phi_c - \phi_{wa}$	$\phi_c - \phi_t$
01	0.456	0.478	-.105
02	-.170	-.182	-.425
03	0.403	0.469	-.429
04	-.343	0.494	-.471
05	0.380	0.473	?.???
06	0.345	0.410	-.109
07	-.480	0.428	-.291
08	0.401	0.444	-.078

Other features we have computed include the average velocity of the figure, by averaging the flow vectors (not by differencing the position of the centroids). This feature does not separate the gaits, since there is little variation, ranging only from -.344 to -.124.

Sequences 01, 02, and 03 are of the same person (child) and 05 and 08 are of an adult. 04, 06, and 07 are three other persons. 05 and 08 do not show strong similarity, but 01 and 03 are quite similar in $(F1, F2)$ as well as in $(F3, F4)$, and $F5$. Sequence 02 is an awkward motion sequence; often being under observation

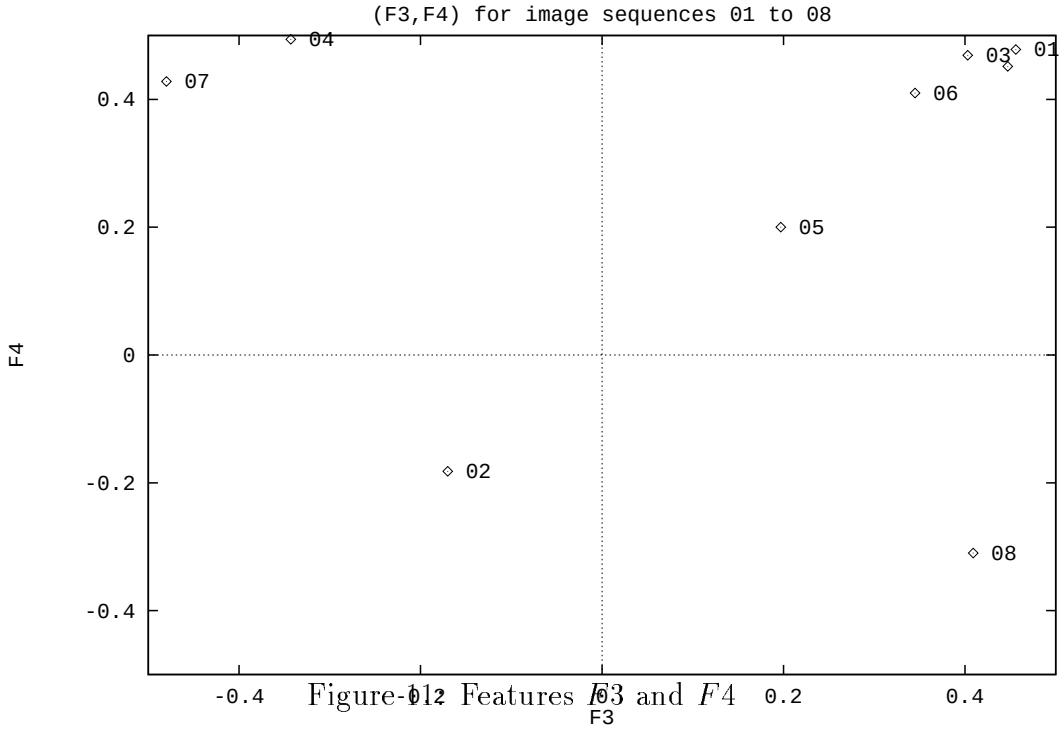

Figure-12 Features $F3$ and $F4$

causes unusual gaits[PB81]: "When one thinks about his or her walking patterns, tension is developed and rhythm and coordination are upset."

We use the vector of features $(F1, F2, F3, F4)$ to discriminate among the gait sequences. The following table shows the squared differences between the feature vectors.

JL FIX TABLE

Seq.	01	02	03	04	05	06	07	08
01	0.00	1.76	0.00	0.93	0.23	0.07	0.91	0.63
02	1.76	0.00	1.61	0.86	0.77	1.86	1.10	1.39
03	0.00	1.61	0.00	0.83	0.18	0.08	0.80	0.62
04	0.93	0.86	0.83	0.00	0.58	0.82	0.21	1.51
05	0.23	0.77	0.18	0.58	0.00	0.33	0.53	0.44
06	0.07	1.86	0.08	0.82	0.33	0.00	0.82	0.55
07	0.91	1.10	0.80	0.21	0.53	0.82	0.00	1.39
08	0.63	1.39	0.62	1.51	0.44	0.55	1.39	0.00

Sequence	Best match	Distance
01	03	0.00
02	04	0.86
03	01	0.00
04	07	0.21
05	01,08	0.02
06	08	0.02
07	04	0.21
08	05	0.02

Sequences 01 and 03 are each other's best match, and 08 selects 05 as its best match, but 05 does not. Several other sequences match 05 better than 08.

4 Discussion

The analysis of moving figures begins with a distribution of moving points in two dimensions. We analyze the dense 2-D optical flow of the person and derive a series of measures of the position of the person and the distribution of the flow. We then determine the frequency and phase of these periodic signals. The relative phase of various measures become features of the motions. We use these features to characterize the gait of several persons.

The regularity of the resulting measures shows that a model-free approach based on the motion distribution is feasible. Further extensions might relax the constraint of absence of models to allow the two main moving substructures, the legs and arms, to be modeled independently. We expect that other features we can extract from these signals may allow discrimination between persons.

References

[BDP+94] A. G. Bharatkumar, K. E. Daigle, M. G. Pandy, Qin Cai, and J. K. Aggarwal. Lower limb kinematics of human walking with the medial axis transformation. In *IEEE Workshop on Non-Rigid Motion*, pages 70–76, 1994.

[BE80a] I. Barrodale and R. E. Erickson. Algorithms for least-squares linear prediction and maximum entropy spectral analysis—part I: Theory. *Geophysics*, 45(3):420–432, March 1980.

[BE80b] I. Barrodale and R. E. Erickson. Algorithms for least-squares linear prediction and maximum entropy spectral analysis—part II: FORTRAN program. *Geophysics*, 45(3):433–446, March 1980.

[BLP89] H. Bulthoff, J. J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Nature*, 337:549–553, February 1989.

[CS94] Claudette Cedras and Mubarak Shah. A survey of motion analysis from moving light displays. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994*, pages 214–221, 1994.

[LK93] James J. Little and Johnny Kam. A smart buffer for tracking using motion data. In *Proc. Workshop on Computer Architectures for Machine Perception*, pages 257–266, December 1993.

[LW82] Kathryn Luttgens and Katharine F. Wells. *Kinesiology: Scientific Basis of Human Motion.* Saunders College Publishing, Philadelphia, 1982.

[NA94a] Sourabh A. Niyougi and Edward H. Adelson. Analyzing and recognizing walking figures in XYT. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994*, pages 469–474, 1994.

[NA94b] Sourabh A. Niyougi and Edward H. Adelson. Analyzing gait with spatiotemporal surfaces. In *IEEE Workshop on Non-Rigid Motion*, pages 64–69, 1994.

[OS89] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1989.

[PB81] John Piscopo and James A. Baley. *Kinesiology, the science of movement*, volume New York. John Wiley and Sons, 1981.

[PN93] Ramprasad Polana and Randal Nelson. Detecting activities. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1993*, pages 2–7, 1993.

[PN94] Ramprasad Polana and Randal Nelson. Recognition of nonrigid motion. In *Proc. 1994 DARPA Image Understanding Workshop*, pages 1219–1224, 1994.

[PN95] Ramprasad Polana and Randal Nelson. Nonparametric recognition of nonrigid motion. Technical report, University of Rochester, 1995.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C (2nd Edition)*. Cambridge University Press, 1992.

[SD94] Steven M. Seitz and Charles R. Dyer. Affine invariant detection of periodic motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994*, pages 970–975, 1994.