VIDERE

Videre: Journal of Computer Vision Research

Quarterly Journal
Winter 1998, Volume 1, Number 2
The MIT Press

Article 1

Recognizing People by Their Gait: The Shape of Motion

James J. Little
Jeffrey E. Boyd

Videre: Journal of Computer Vision Research (ISSN 1089-2788) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $30.00, Institutions $125.00. Canadians add additional 7% GST. Prices subject to change without notice.

Recognizing People by Their Gait: The Shape of Motion

James J. Little,[1] Jeffrey E. Boyd[2]

The image flow of a moving figure varies both spatially and temporally. We develop a model-free description of instantaneous motion, the *shape of motion*, that varies with the type of moving figure and the type of motion. We use that description to recognize individuals by their gait, discriminating them by periodic variation in the shape of their motion. For each image in a sequence, we derive dense optical flow, $(u(x, y), v(x, y))$. Scale-independent scalar features of each flow, based on moments of the moving point weighted by $|u|$, $|v|$, or $|(u, v)|$, characterize the spatial distribution of the flow.

We then analyze the periodic structure of these sequences of scalars. The scalar sequences for an image sequence have the same fundamental period but differ in phase, which is a phase feature for each signal. Some phase features are consistent for one person and show significant statistical variation among persons. We use the phase feature vectors to recognize individuals by the shape of their motion. As few as three features out of the full set of twelve lead to excellent discrimination.

Keywords: action recognition, gait recognition, motion features, optic flow, motion energy, spatial frequency, analysis

1 Introduction

Our goal is to develop a model-free description of image motion, and then to demonstrate its usefulness by describing the motion of a walking human figure and recognizing individuals by variation in the characteristics of the motion description. Such a description is useful in video surveillance where it contributes to the recognition of individuals and can indicate aspects of an individual's behavior. Model-free descriptions of motion could also prove useful in vision-based user interfaces by helping to recognize individuals, what they are doing, and nuances of their behavior.

The pattern of motion in the human gait has been studied in kinesiology using data acquired from moving light displays. Using such data, kinesiologists describe the forward propulsion of the torso by the legs, the ballistic motion of swinging arms and legs, and the relationships among these motions [23, 30]. Similarly, in computer vision, model-based approaches to gait analysis recover the three-dimensional structure of a person in a model and then interpret the model. The literature on moving light displays provides an introduction to modeling and moving figures [11]. Unuma, Anjyo, and Takeuchi [42] show the value of a structural model in describing variations in gaits. They use Fourier analysis of joint angles in a model to synthesize images of different types of gaits, e.g., a happy walk versus a tired walk.

Alternatives to the model-based approach emphasize determining features of the motion fields, acquired from a sequence of images, without structural reconstruction. Recent theoretical work demonstrates the recoverability of affine motion characteristics from image sequences [38]. It is therefore reasonable to suggest that variations in gaits are recoverable from variations in images sequences and that a model-free approach to gait analysis is viable. Moreover, during periodic motion the varying spatial distribution of motion is apparent. Capturing this variation and analyzing its temporal variation should lead to a useful characterization of periodic motion.

Hogg [16] was among the first to study the motion of a walking figure using an articulated model. There have recently been several attempts to recover characteristics of gait from image sequences, without the aid of annotation via lights [35, 5, 27, 28, 31, 32, 3, 4]. Niyogi and Adelson [27, 28] emphasize segmentation over a long sequence of frames. Their technique relies on recovering the boundaries of moving figures in the xt domain [27] and recently [28] 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 [31, 32] 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

[1] Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4. little@cs.ubc.ca

[2] Department of Electrical and Computer Engineering, University of California, La Jolla, CA 92093-0407. jeffboyd@ece.ucsd.edu

mitpress.mit.edu/videre.html

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 128 × 128 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 [32, 33] emphasizes the spatial distribution of energies around the moving figure. They compute spatial statistics 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.

Shavit and Jepson [39, 40] use the centroid and moments of a binarized motion figure to represent the distribution of its motion. The movement of the centroid characterizes the external forces on an object, while the deformation of the object is computed from the dispersion (the eigenvalues of the covariance matrix) or ratio of lengths of the moment ellipse.

Bobick and Davis [6] introduced the Motion Energy Image (MEI), a smoothed description of the cumulative spatial distribution of motion energy in a motion sequence. They match this description of motion against stored models of known actions. Bobick and Davis [7] enhanced the MEI to form a motion-history image (MHI), where pixel intensity is a function, over time, of the energy in the current motion energy (binarized) and recent activity, which they extend in later work [14]. We will discuss these two representations further in Section 2.2.

Baumberg and Hogg [3] present a method of representing the shape of a moving body at an instance in time. Their method produces a description composed of a set of principal spline components and a direction of motion. In later work, Baumberg and Hogg [4] add temporal variation by modeling the changing shape as a vibrating plate. They create a vibration model for a "generic" pedestrian and then are able to measure the quality of fit of the generic data to another pedestrian.

Liu and Picard [22] detect and segment areas of periodic motion in images by detecting spectral harmonic peaks. The method is not model based and identifies regions in the images that exhibit periodic motion.

Recently more elaborate models, often including kinematics and dynamics of the human figure, are used to track humans in sequences [36, 9, 19, 18, 43].

Our work, in the spirit of Polana and Nelson, and Baumberg and Hogg, is a model-free approach making no attempt to recover a structural model of a human subject. Instead we describe the shape of the motion with a set of features derived from moments of a dense flow distribution [20]. Our goal is not to fingerprint people, but to determine what content of motion aids recognition. We wish to recognize gaits, both types of gaits as well as individual gaits. The features are invariant to scale and do not require synchronization of the gait or identification of reference points on the moving figure.

The following sections describe the creation of motion features and an experiment that determines the variation of the features over a set of walking subjects. Results of the experiment show that features acquired by our process exhibit significant variation due to different subjects and are suitable for recognition of people by subtle differences in their gaits,

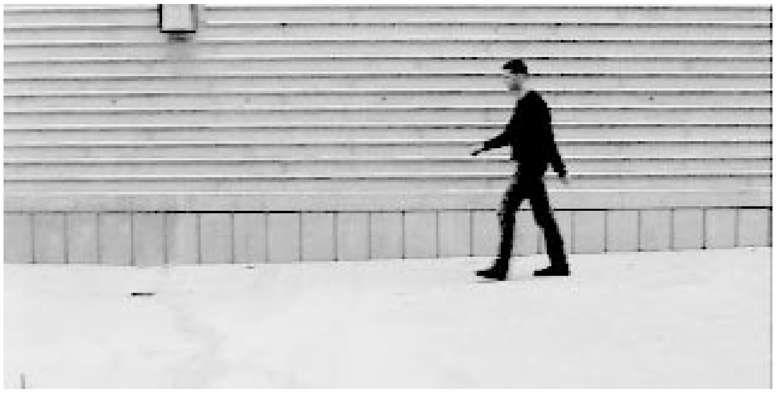


Figure 1. Sample image from experimental data described in Section 3 (number 23 of 84 images, sequence 3 of subject 5).

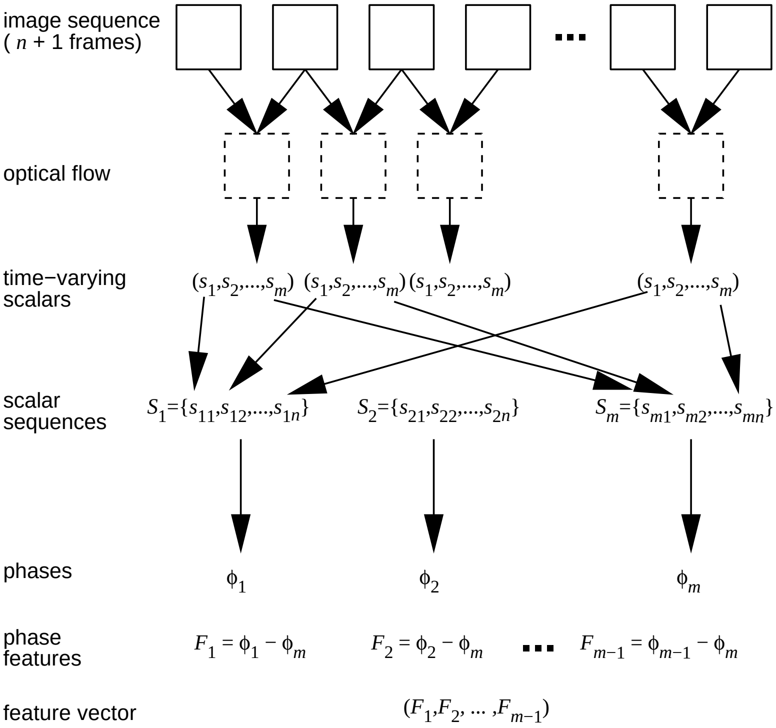


Figure 2. The structure of the image analysis. Each image sequence produces a vector of $m - 1$ phase values.

as identified by phase analysis of periodic variations in the shape of motion.

2 Motion Feature Creation

Image sequences are gathered while the subject is walking laterally before a static camera and processed offline. Motion stabilization could be accomplished by a tracking system that pursues a moving object, e.g., Little and Kam [21]. However, our focus is on the motion, so we restrict the experimental situation to a single subject moving in the field of view before a static camera. Figure 1 shows an example of the images used, image number 23 of 84 in a sequence taken from the experimental data described in Section 3.

Figure 2 illustrates the data flow through the system that creates our motion features. We begin with an image sequence of $n + 1$ images and then derive n dense optical flow images. For each of these optical flow images we compute m characteristics that describe the shape of the motion (i.e., the spatial distribution of the flow), for example, the centroid of the moving points, and various moments of the flow distribution. Some of these are locations in the image, but we treat all as time-varying scalar values. We arrange the values to form a time series for each scalar. A walking person undergoes periodic motion, returning to a standard position after a certain time period that depends on the frequency of the gait. Thus we analyze the periodic structure of these time series, and determine the fundamental frequency of the variation of each scalar. The set of time series for an image sequence share the same frequency, or simple multiples of the fundamental, but their phases vary. To make the data from different sequences comparable, we subtract a reference phase, ϕ_m, derived from one of the scalars. We characterize each image sequence by a vector, $F = (F_1, \ldots, F_{m-1})$, of $m - 1$ relative phase features. The phase feature vectors are then used to recognize individuals.

2.1 Tracking and Optical Flow

The motion of the object is a path in three dimensions; we view its projection. Instead of determining motion of three-dimensional elements of a figure, we look for characteristics of the periodic variation of the two-dimensional optical flow.

The raw optical flow identifies temporal changes in brightness; however, illumination changes such as reflections, shadows, moving clouds, and inter-reflections between the moving figure and the background, as well as reflections of the moving figure in specular surfaces in the background, pollute the motion signal. To isolate the moving figure, we manually compute the average displacement of the person through the image sequence and then use only the flow within a moving window traveling with the average motion. Within the window there remain many islands of small local variation, so we compute the connected components of each F_j and eliminate all points not in sufficiently large connected regions. The remaining large components form a mask within which we can analyze the flow. This reduces the sensitivity of the moment computations to outlying points.

Figure 3 shows six subimages from the sequence corresponding to Figure 1. We will refer to the subimages as images from here on and will display our results in subimages, for compactness. All processing is carried out in the coordinates of the original frames.

Unlike other methods, we use dense optical flow fields, generated by minimizing the sum of absolute differences between image patches [10]. The algorithm is sensitive to brightness change caused by reflections, shadows, and changes of illumination, so we first process the images by computing the logarithm of brightness, transforming the multiplicative effect of illumination change into an additive one. Filtering by a Laplacian of Gaussian (effectively a bandpass filter) removes the additive effects.

The optical flow algorithm, for each pixel, searches among a limited set of discrete displacements for the displacement $(u(x, y), v(x, y))$ that minimizes the sum of absolute differences between a patch in one image and the corresponding displaced patch in the other image. The algorithm finds a best-matching patch in the second image for each patch in the first. The algorithm is run a second time, switching the roles of the two images. For a correct match, the results will likely agree. In order to remove invalid matches, we compare the results at each point in the

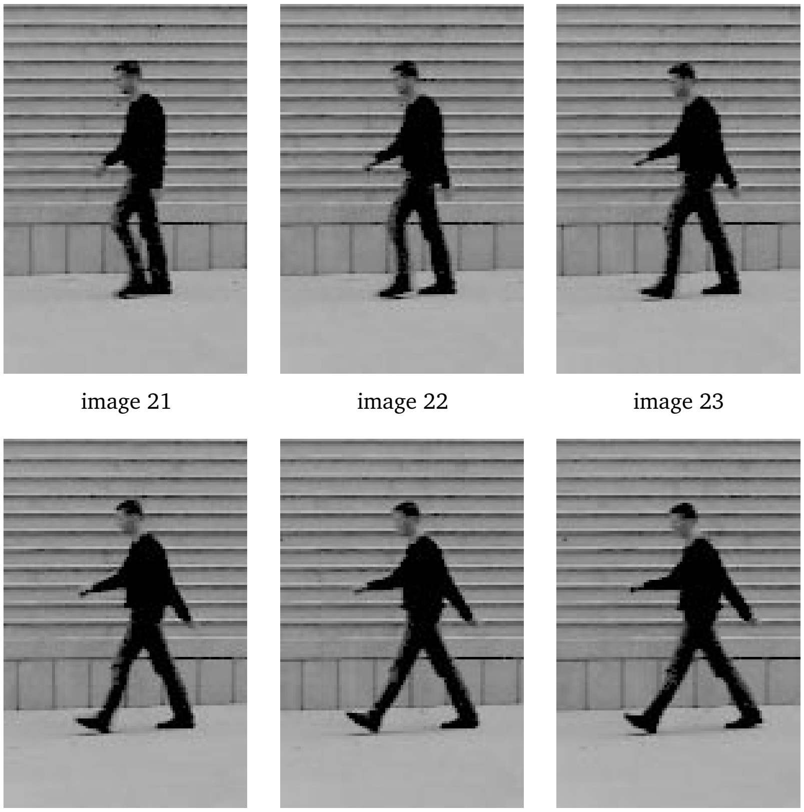


image 21

image 22

image 23

image 24

image 25

image 26

Figure 3. Subimages 21 through 26 for the sequence corresponding to Figure 1.

first image with the result at the corresponding point in the second. The second point should match to the first: the sum of displacement vectors should be approximately zero [25, 15]. Only those matches that pass this validation test are retained.

The results could be interpolated to provide subpixel displacements, but we use only the integral values. In effect, the minimum displacement is 1.0 pixels per frame; points that are assigned non-zero displacements form a set of *moving points*. Let $T(u, v)$ be defined as

$$T(u, v) = \begin{cases} 1, & \text{if } |(u, v)| \geq 1.0 \\ 0, & \text{otherwise} \end{cases}$$

$T(u, v)$ segments moving pixels from non-moving pixels. Figure 4 shows the moving points for the images in Figure 3.

2.2 Spatial Distribution of Flow

The flow component of the system provides dense measurements of optical flow for a set of points in the image. Instead of finding the boundary of this set [3, 4], we use all the points and analyze their spatial distribution. We use the points with unit values, as signified by T (as shown in Figure 4), as well as weighted by the magnitude of the motion, $|(u, v)|$, at every point, as shown in Figure 5, and weighted by $|u|$ and $|v|$, as shown in Figures 6 and 7.

To describe the spatial distribution, we compute the centroid of all moving points. The x and y coordinates of a centroid are two scalar measures of the distribution of motion. We also compute the second moments of each spatial distribution. The moment of inertia about an axis [41] describes the distribution by the average of the product of the

Figure 4. The moving points for images in Figure 3. (White is moving and black is stationary.)

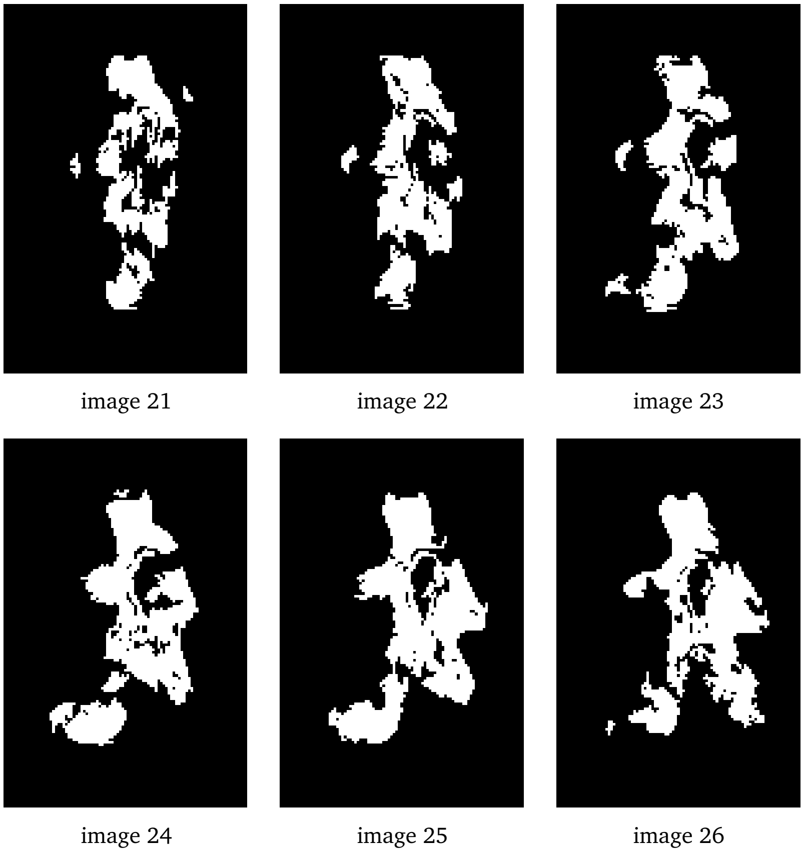


image 21

image 22

image 23

image 24

image 25

image 26

Figure 5. The magnitudes of the flow, $|(u, v)|$, for images in Figure 3.

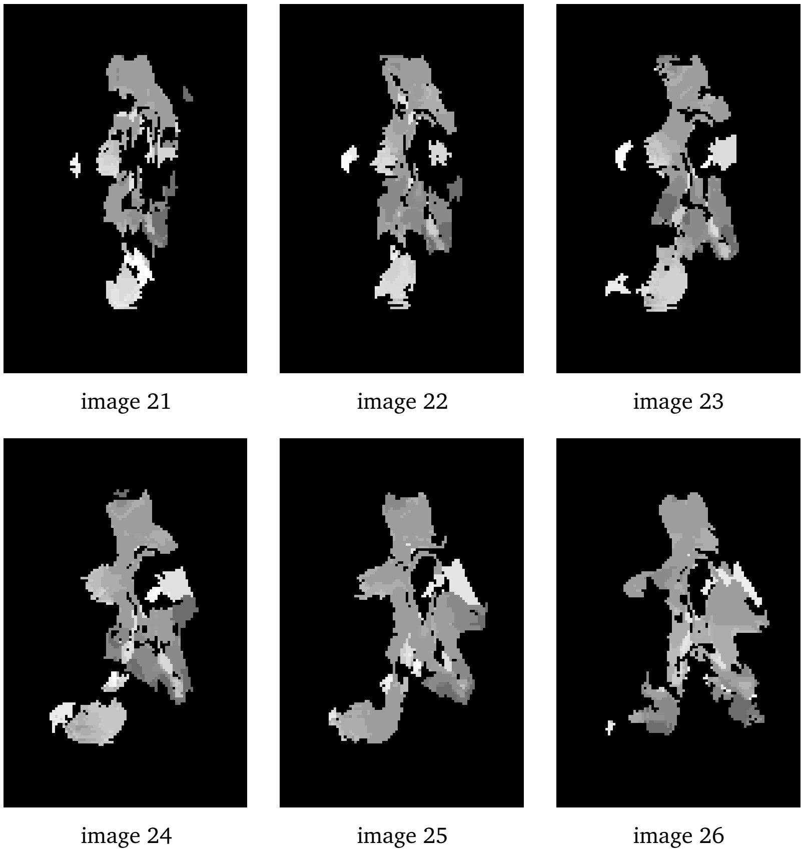


image 21

image 22

image 23

image 24

image 25

image 26

Figure 6. The x component of flow, u, for images in Figure 3.

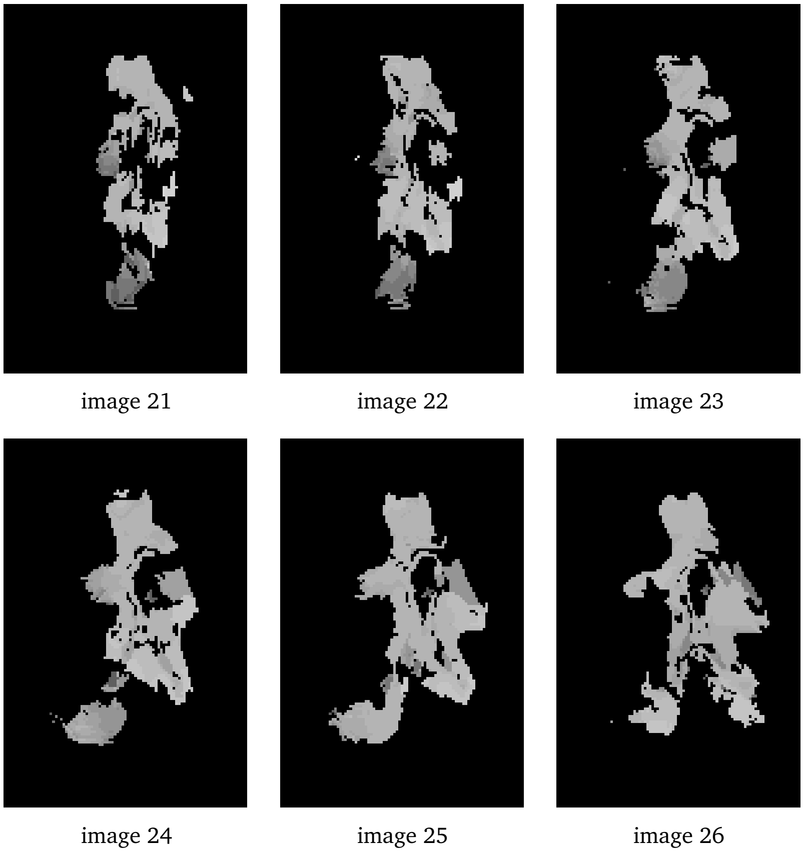


image 21

image 22

image 23

image 24

image 25

image 26

Figure 7. The y component of flow, v, for images in Figure 3.

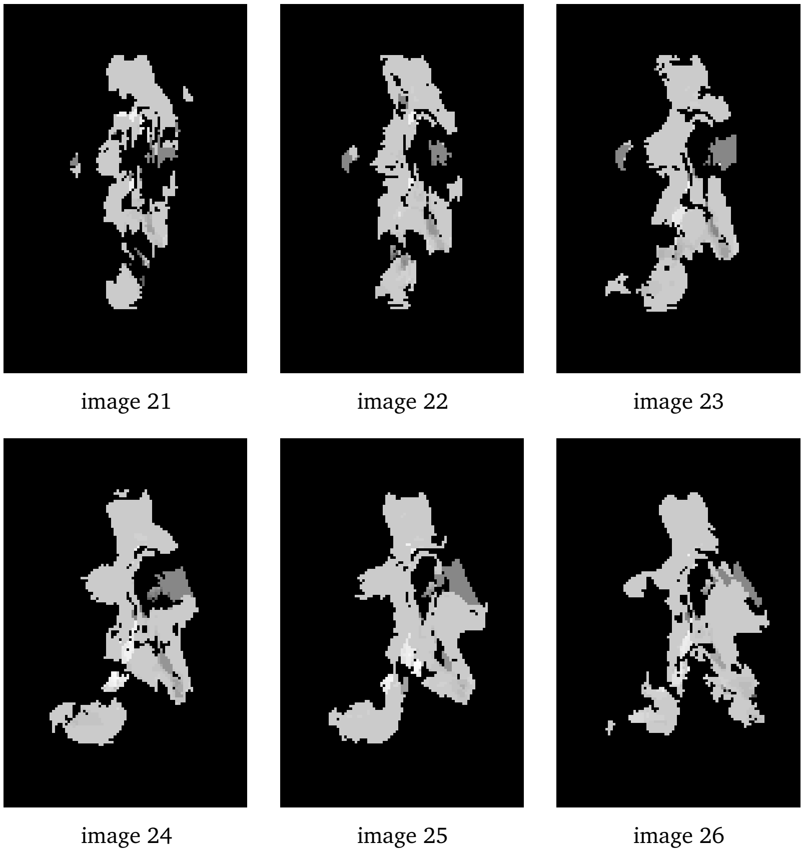


image 21

image 22

image 23

image 24

image 25

image 26

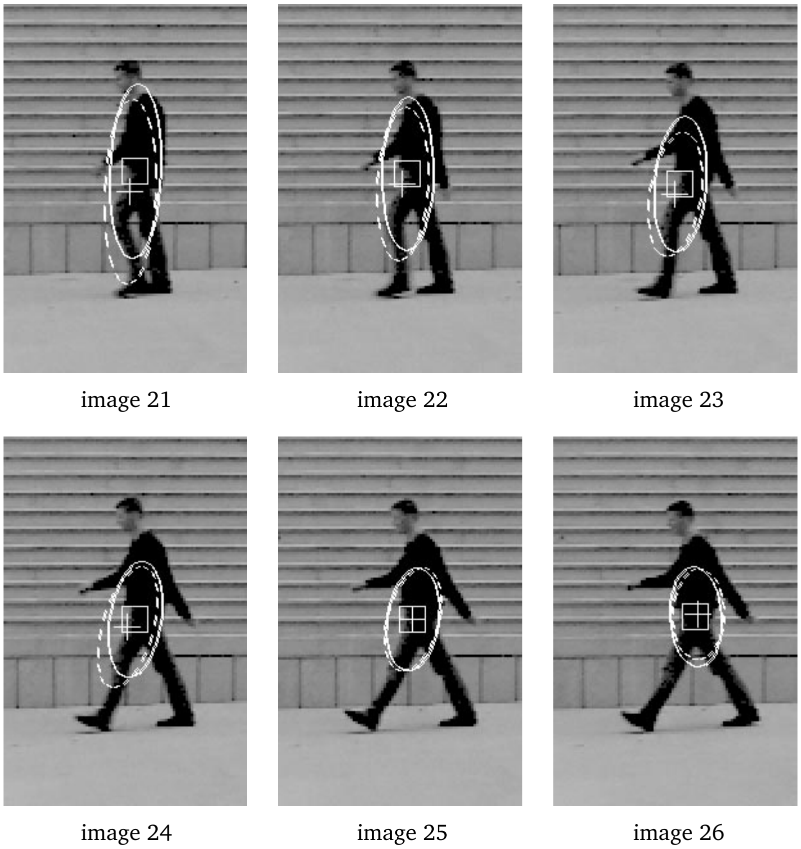


image 21

image 22

image 23

image 24

image 25

image 26

Figure 8. The centroid and moment ellipse for T (box and solid lines) and for $T|(u, v)|$ (cross and dashed lines) for images in Figure 3.

mass at each point by its distance from the axis. The moment varies with the direction of the axis; it can be succinctly described by an ellipse whose major and minor axes show the minimum and maximum values of the moment.

The *shape of motion* is the distribution of flow, characterized by several sets of measures of the flow: the moving points (T) and the points in T weighted by $|(u, v)|$, $|u|$, and $|v|$. The features of the flow include the centroids and second moments of these distributions. It characterizes the flow at a particular instant in time.

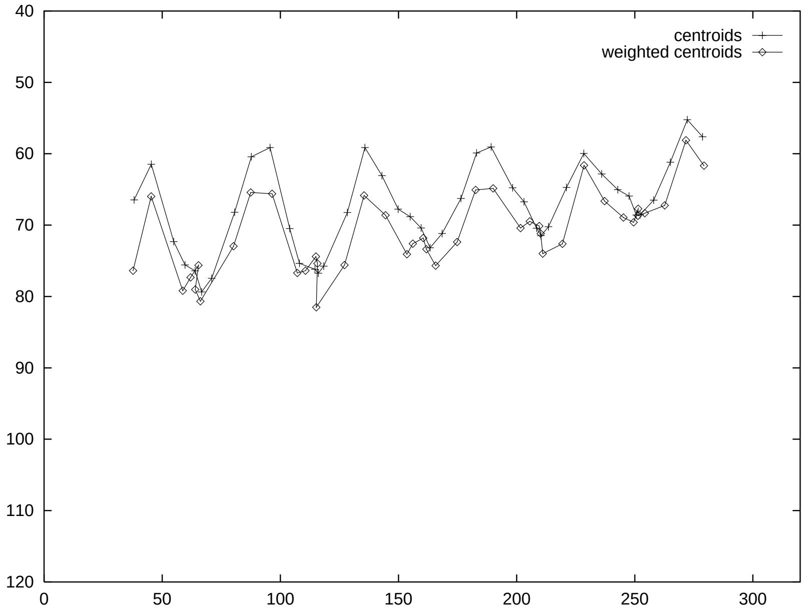
The shape of motion varies systematically during a motion sequence. The relative positions of the centroids of T and $T|(u, v)|$ vary systematically over the sequence. Figure 8 displays the centroid of T as a box and the moment ellipse for T in solid lines superimposed on the image sequence. The centroid of $T|(u, v)|$ is shown as a cross, and its ellipse is shown in dashed lines. 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.[1] Figure 9 shows the centroids of these two distributions plotted in the coordinates of the original image sequence.

The scalar measures we extract are shown in Table 1 (summations are over the image).

Each image I_j in an image sequence generates $m = 13$ scalar values, s_{ij}, where i varies from 1 to m, and j from 1 to n. We assemble scalar

[1] Hogg in 1983 noted the periodic variation of the size of the bounding box of a moving figure.

Figure 9. Plot of (x_c, y_c) and (x_{wc}, y_{wc}) for the full example image sequence, as cross and diamond, respectively.



sequences for each time-varying scalar, S_i. The next section describes how we compute the frequency and phase for each of these signals.

The Motion Energy Image (MEI) [6] is arrived at by: binary threshold of motion displacements computed by thresholding the pixelwise summed squared difference between each image and the first, over an entire sequence. The features characterizing the MEI are a set of the seven Hu moments, which are translation, scale, and rotation invariant [17], plus terms sensitive to orientation and correlation of x and y, and a measure of compactness. Likewise the MEI is one image, now grey-valued to represent the recent history of motion at a location, that represents an entire sequence. The critical difference in our approach is that the shape of motion is a description of the instantaneous distribution of motion at one point in a sequence, rather than an integration of the motion of an entire sequence into a pair of images. We then observe the variation of the description and compute features derived over time. Shavit and Jepson [40] likewise observe the variation in shape of the moment ellipse but use it to derive conclusions about forces affecting the object.

2.3 Frequency and Phase Estimation

A human gait has a single, fundamental driving frequency, as a consequence of the fact that the parts of the body must move in a cooperative manner. For example, for every step forward taken with the left foot, the right arm swings forward exactly once. Since all components of the gait, such as movements of individual body parts, have the same fundamental frequency, all signals derived by summing the movements of these parts must also have that frequency. It is possible for higher frequency harmonics to be introduced, but only in integer multiples of the fundamental. Although the frequency of the scalar sequences derived from a gait must be the same, the phases of the signals vary. We find the phase of each signal after first finding the fundamental frequency.

The time series we generate contain relatively few cycles and are very noisy. Both these factors confound Fourier transform techniques because

Table 1. Summary of scalar measures describing the shape of the optical flow.

Description	Label	Formula
x coordinate of centroid	x_c	$\sum xT / \sum T$
x coordinate, centroid of $\lvert(u, v)\rvert$ distribution	x_{wc}	$\sum x\lvert(u, v)\rvert T / \sum \lvert(u, v)\rvert T$
y coordinate, centroid of $\lvert(u, v)\rvert$ distribution	y_{wc}	$\sum y\lvert(u, v)\rvert T / \sum \lvert(u, v)\rvert T$
x coordinate of difference of centroids	x_d	$x_{wc} - x_c$
y coordinate of difference of centroids	y_d	$y_{wc} - y_c$
aspect ratio (or elongation), ratio of length of major axis to minor axis of an ellipse	a_c	$\lambda_{max}/\lambda_{min}$, where λs are eigenvalues of second moment matrix for distribution
elongation of weighted ellipse	a_{wc}	as in a_c, but for weighted distribution
difference of elongations	a_d	$a_c - a_{wc}$
x coordinate, centroid of $\lvert u\rvert$ distribution	x_{uwc}	$\sum x\lvert u\rvert T / \sum \lvert u\rvert T$
y coordinate, centroid of $\lvert u\rvert$ distribution	y_{uwc}	$\sum y\lvert u\rvert T / \sum \lvert u\rvert T$
x coordinate, centroid of $\lvert v\rvert$ distribution	x_{vwc}	$\sum x\lvert v\rvert T / \sum \lvert v\rvert T$
y coordinate, centroid of $\lvert v\rvert$ distribution	y_{vwc}	$\sum y\lvert v\rvert T / \sum \lvert v\rvert T$
y coordinate of centroid	y_c	$\sum yT / \sum T$

of the inevitable presence of sidelobes in the spectrum. To avoid this problem, we turn to maximum entropy spectrum estimation [37]. This technique can be summarized as follows.

1. Find a set of coefficients that predicts values of a time series from a set of previous values.
2. Build a linear shift-invariant filter that gives the difference between the predicted values and the actual signal.
3. If the coefficients predict the underlying sinusoidal signal correctly, then the output of the filter must be noise.
4. The spectrum of the noise and the z-transform of the filter are known, so the spectrum of the underlying sinusoids can be derived.

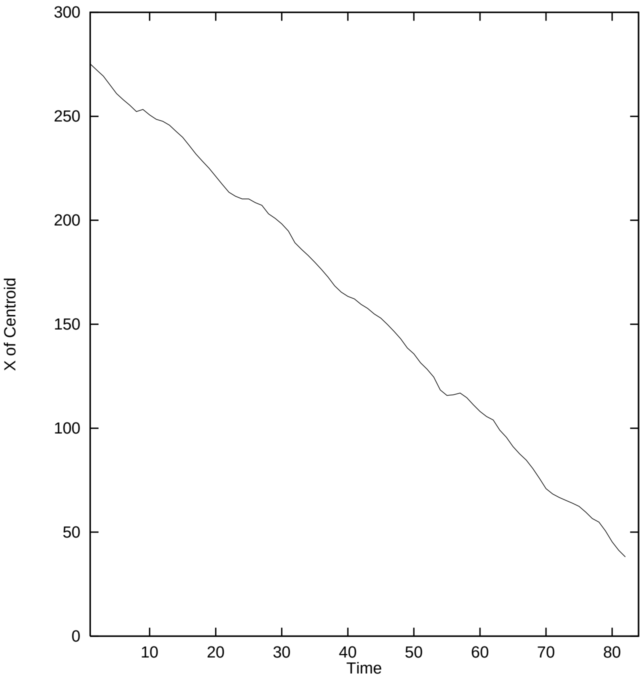
We use the least-squares linear prediction (LSLP) method of Barrodale and Erickson [1, 2] to find the prediction coefficients. From there it is a simple matter to compute the spectrum.

Let $\{x_t\}$ be a time series of length n. Then $\{\hat{x}_t\}$ are the values of the sequence predicted by the m linear coefficients a_j, $j = 1, \ldots, m$, i.e.,

$$\hat{x}_t = \sum_{j=1}^{m} a_j x_{t-j}, \qquad t = m+1, m+2, \ldots, n \tag{1}$$

predicts x_t. To find the coefficients that give the best prediction, Barrodale and Erickson find the least-squares solution of

Figure 10. Plot of x_c, without linear component subtracted, for image sequence shown in Figure 3.



$$C\mathbf{a} = \mathbf{y}, \tag{2}$$

where

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

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

Solving for **a** from the symmetric positive definite system of equations

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

gives a set of coefficients that minimizes the L_2 norm of the prediction error, which is the residual of Equation (2). 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 (to exploit the data in the short series as much as possible [37] using LU decomposition [34]. Before applying LSLP, we subtract the linear background, as estimated by linear regression, from the signal. We can do this because the subject walks with a constant velocity. (See Figures 10 and 11.)

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

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

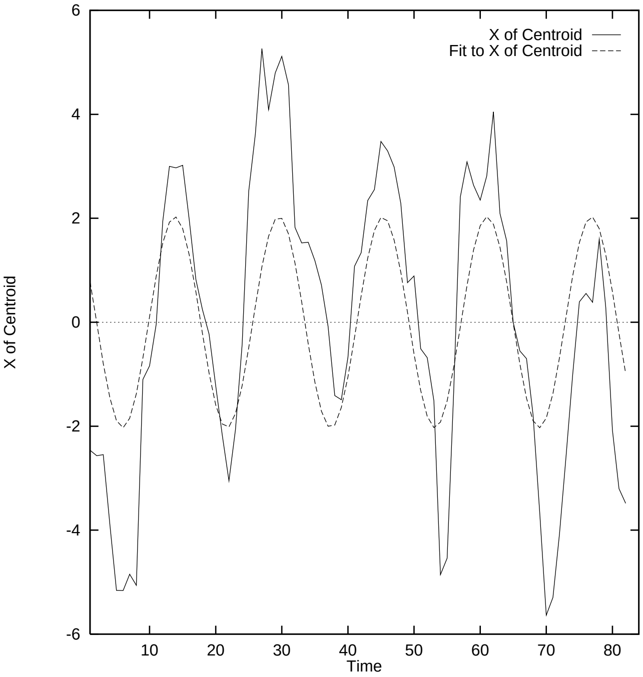


Figure 11. Plot of x_c, with the linear component subtracted, and the sinusoid representing the fundamental frequency from the LSLP fit for the sequence in Figure 3.

where Δt is the sampling period, $P_m = S_m/2(n-m)$, and S_m is the L_2 norm of the prediction error of Equation (2). Note that $\{\hat{x}_t\}$ is never computed. Within the constant scaling factor of $P_m \Delta t$, only the coefficients $\{a_j\}$ are needed to compute the spectrum. The number of coefficients required for an accurate spectrum depends on the content of the signal and the amount of noise. A pure sinusoid with no noise requires only two coefficients. As the noise increases, the required number of coefficients also increases. However, if too many coefficients are used then the spectrum begins to model the noise and shows erroneous peaks. Twenty coefficients are used to estimate the spectrum for the time series considered here. This proved to be a reliable number and only in the noisiest sequences was it necessary to use a different number to avoid modeling the noise. To get the fundamental frequency of the gait, 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, ω_{max}. This is the fundamental frequency.

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 an infinite time series are given by Oppenheim and Schafer [29]:

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

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

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

Figure 12. LSLP spectrum of y_c, for image sequence shown in Figure 3; the fundamental frequency is at 0.063.

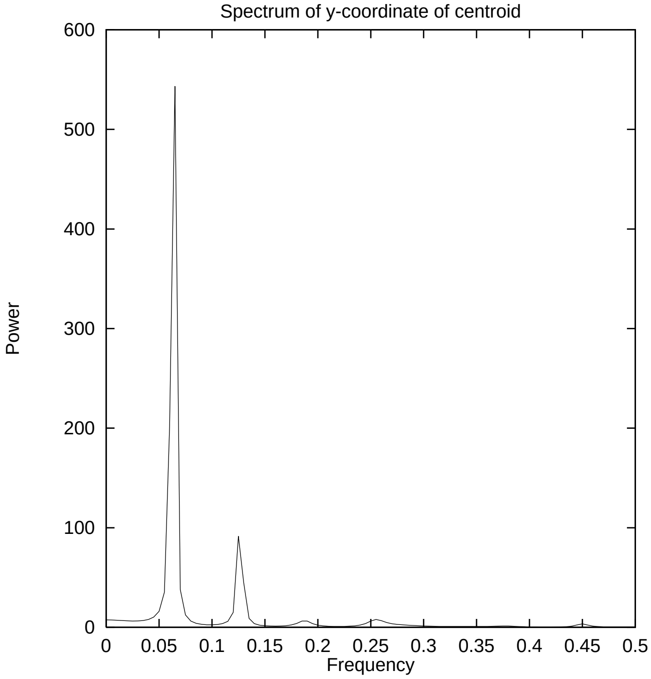
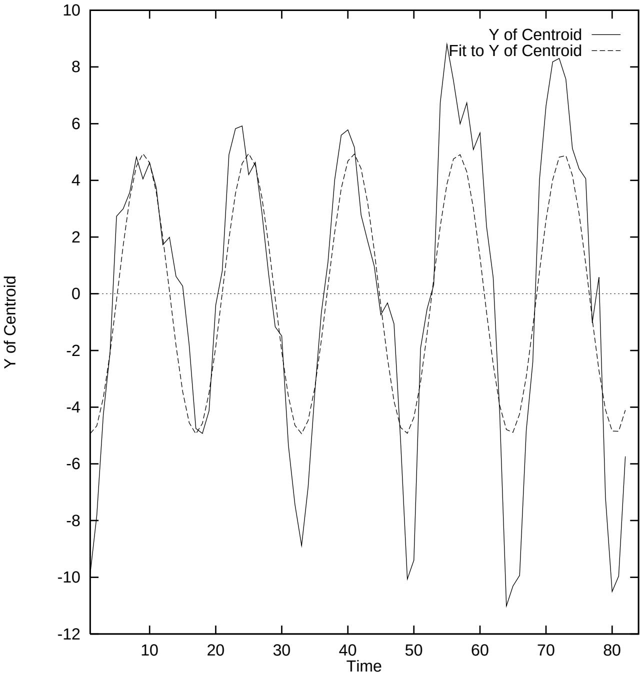


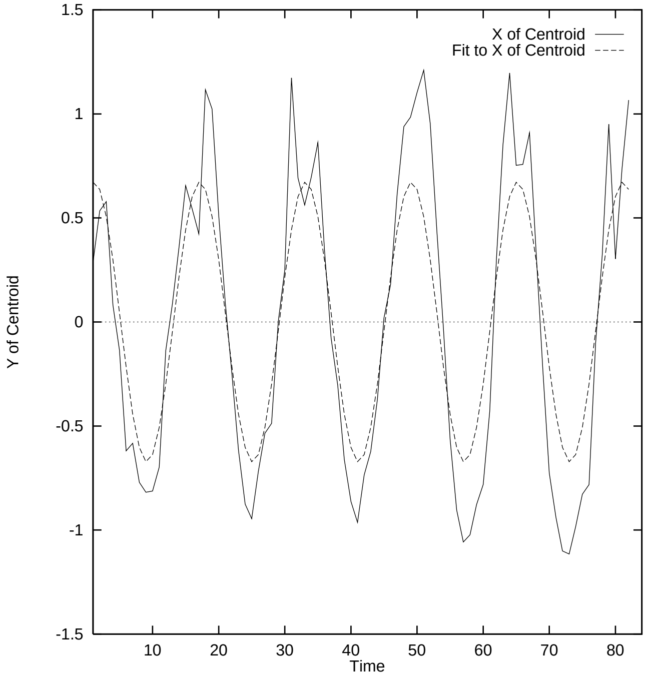
Figure 13. Plot of y_c, with linear component subtracted, and LSLP fit for image sequence shown in Figure 3.



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

For example, Figure 10 shows the x_c signal for the sequence shown in Figure 3. Figure 11 shows x_c with the linear background removed to reveal the sinusoidal signal. Superimposed on this signal is a pure sinusoid with frequency ω_{max} at the phase given by Equation (6). Figure 12 is the LSLP maximum entropy spectrum of the y_c signal based on twenty forward and backward prediction coefficients. The frequency ω

Figure 14. Plot of a_c, with linear component subtracted, and LSLP fit for image sequence shown in Figure 3.



is expressed as a ratio of f_s, the sampling frequency, where f_s is $1/\Delta t$. The spectrum, Figure 12, shows a definite maximum that identifies the frequency of the gait. Figures 13 and 14 show two additional examples of signals, y_c and a_c, and their LSLP fits.

2.4 Phase Features

The phase computed by Equation (6) depends arbitrarily on where the sequence begins in the subject's gait. In order to get a feature that has no arbitrary dependencies, it is necessary to select one of the signals as a phase reference for all the others. The measurement of y_c was empirically determined to be the most reliable over all the image sequences that we sampled in experimentation. The frequency computed for it is one step, for example, from the left footfall to the right footfall, which takes approximately 16 frames for most image sequences. We choose the fundamental frequency, ω_{max}, from the spectrum of y_c and compute the other phase measurements, fixing the frequency.

The frequency computed independently for the other scalar measures either differs slightly or is a multiple of the frequency of y_c, as shown in Table 2.

Figure 15 shows the y_{wc} signal and its LSLP fits, both at the fundamental, 0.063, and its own best LSLP fit, at frequency 0.125. The spectrum of that signal, y_{wc} (which appears in Figure 16), shows a strong peak at 0.125 (which is approximately a harmonic of 0.063), and also shows a peak comparable to the peak in the original spectrum (Figure 12). The strong harmonic appears only in the signals of the weighted distribution, and only in the y elements. We believe this is due to occlusions as the rapidly moving limbs disappear behind the body. Nevertheless, the fit at the fundamental is also quite good.

We use the phase of y_c as the reference and subtract it from the other phase measurements for an image sequence to create the phase feature vector. We can thus avoid having to identify a reference position (say, for example, the left footfall) in each image sequence. Variation in the

Table 2. ω_{max} computed for scalar sequences corresponding to image sequence in Figure 3.

x_c	0.0635
y_c	0.0635
x_{wc}	0.063
y_{wc}	0.125
x_d	0.063
y_d	0.063
a_c	0.0625
a_{wc}	0.0625
a_d	0.064
x_{uwc}	0.063
y_{uwc}	0.1255
x_{vwc}	0.02
y_{vwc}	0.0625

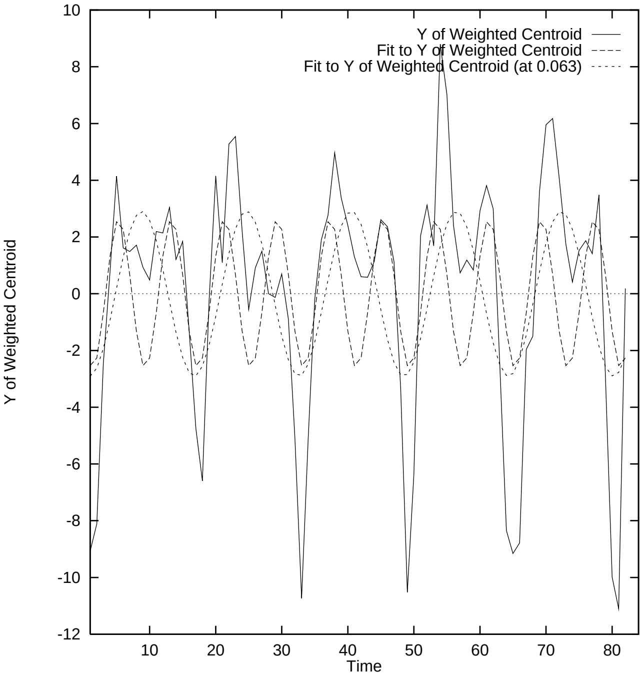


Figure 15. Plot of y_{wc}, with linear component subtracted, and LSLP fits for image sequence shown in Figure 3, both at 0.063 and 0.125.

measurement of the reference propagates to each of the other phase features, so a reliable reference is essential to minimize variation in the feature vector. Our thirteen scalar features ($m = 13$) thus provide twelve measurements per image sequence. Figure 17 shows a plot of the fits to all of x_c, y_c, a_c, and y_{vwc}, and illustrates the phase relationships among them.[2]

At this point we have completed the analysis of the image sequences as shown in Figure 2. We have computed the shape of motion for each pair of images, giving a set of thirteen measurements for each optical flow image. Each scalar measurement, collected for the sequence, yields a time series that is then analyzed by LSLP to produce a phase value, at the fundamental frequency determined from y_c. These thirteen values

[2] The relative magnitudes of the signals may also be useful information, but we have not investigated their use.

Figure 16. LSLP spectrum of y_{wc}, for image sequence shown in Figure 3; the fundamental frequency is at 0.063, but the dominant frequency here is 0.125.

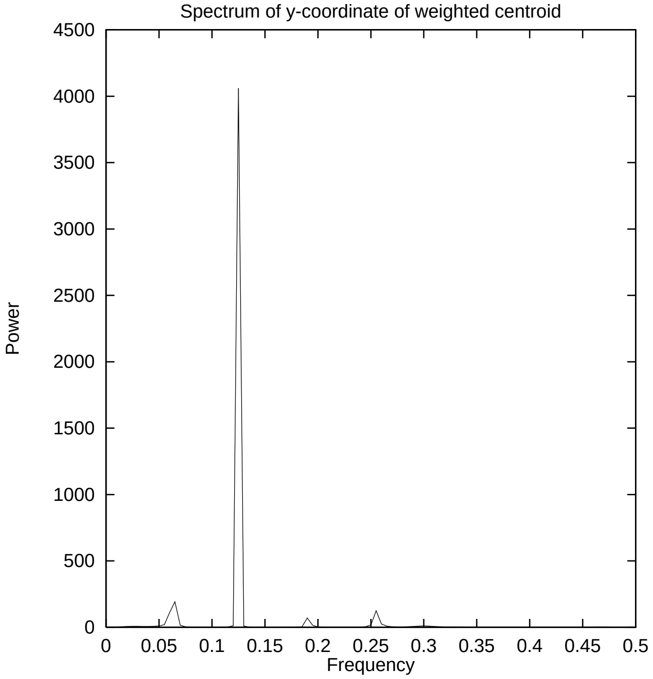
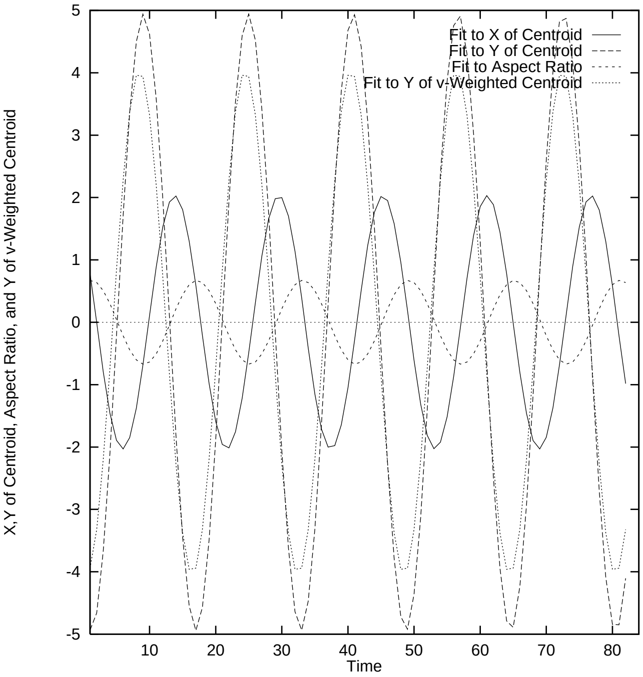


Figure 17. Plots of LSLP fits to x_c, y_c, a_c, and y_{vwc}, image sequence 3 of subject 5. x_c and a_c are almost in phase, while y_c is almost 0.5 cycles out of phase, and y_{vwc} is somewhat out of phase.

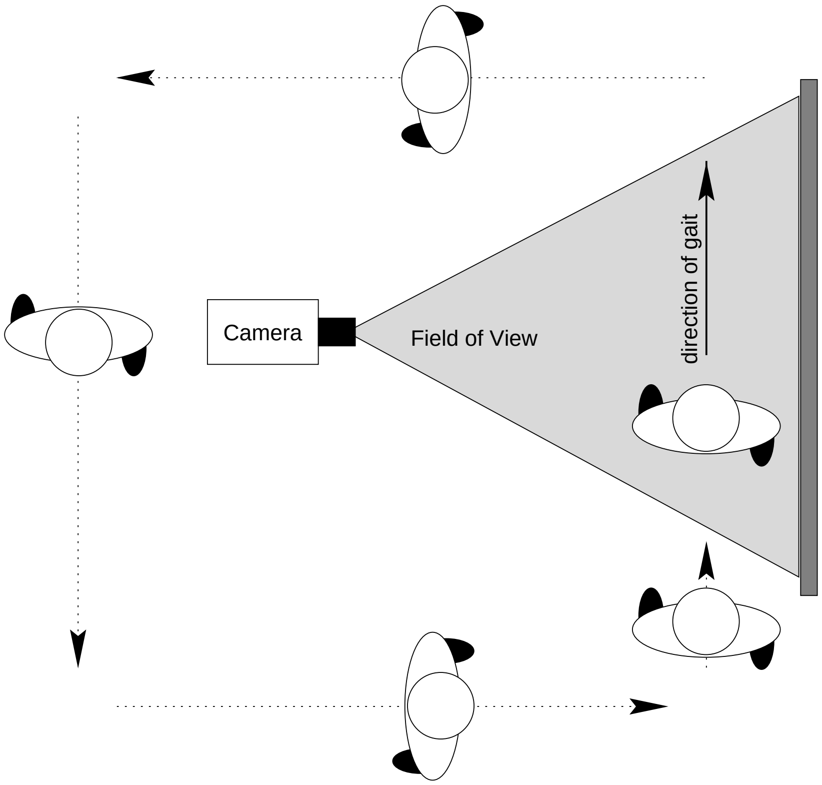


are reduced to twelve, by subtracting the reference phase for y_c, yielding a phase vector for an image sequence. The phase vector is independent of image scale or temporal sampling period, as well as the relative speed of the gait, as long as it remains a normal walking gait.

3 Experiment

To verify the usefulness of the proposed description, we analyzed a set of image sequences of a set of six walking subjects viewed from a

Figure 18. Schematic of experimental apparatus, plan view.



static camera before a fixed background, computing the shape of motion and deriving phase features for each sequence. We then used the phase features in a simple recognition test: we computed the mean phase vector for each subject and then classified each sequence by minimum Euclidean distance from the mean vector. As part of this test, we determined the subset of the twelve features that were most effective for recognition.

In our experiment we sampled the gaits of six people using the apparatus depicted in Figure 18. A camera fixed on a tripod points towards a fixed non-reflecting static background. We achieved diffuse lighting by acquiring the data outdoors and in the shade. The subjects walked in a circular path such that on one side of the path they passed through the field of view of the camera and passed behind the camera on the other side. The subjects walk this path for approximately fifteen minutes while the images are recorded on video tape.

After discarding at least the first two passes recorded to avoid anomalies caused by awareness of the camera, seven image sequences of each of the selected subjects were digitized from the tape, i.e., there are six people and seven sequences per person for a total of 42 sequences. The length of each sequence varies with the amount of time each person takes to traverse the field of view, but the average length is about 80 frames at 30 frames per second.

Images digitized from the tape have a resolution of 480 × 640 pixels in 24-bit full color. Before computing the optical flow, we convert the picture to a gray scale (we use the Y component of a YUV transformation of the RGB values) and subsample it, by averaging, to half the original resolution. The resulting images have a resolution of 240 × 320. Figure 1 shows an example frame of the lower-resolution resampled black and white images. We could have used only one field of the image, but the gain of vertical resolution using a full frame offsets the temporal smoothing produced by using both fields. We use only the lower 160 pixels, cutting off the top (which contains only sky). If the camera were closer we would have had a better view of the moving figures, at the cost of shorter sequences. We opted to get long sequences to improve frequency and phase estimation. Since the step is approximately 10-20 frames, we need at least that, preferably at lease two steps to include the full cycle. We have used at least 60 frames in all our experiments but have not experimented with shorter sequences.

4 Results

We analyze the phase features in two ways. First, analysis of variance allows us to determine whether or not there is any significant variation in the phase features among the subjects. Second, we test the matches between each sequence and the remaining sequences to show successful recognition. The following two sections describe the results based on these analyses.

4.1 Analysis of Variance

Analysis of variance (ANOVA) is a statistical method used to indicate the presence of a significant variation in a variable related to some factor. Here the variables are the phase features and the factor in which we are interested is the person walking in front of the camera. Our analysis uses a single-factor ANOVA as described by Winer [45]. The method analyzes only a single variable and must be repeated for each phase feature. We used the commercial statistical software package StatView 4.5 to perform the analysis.

Care is necessary because most statistical software expects a continuous random variable to exist on a line, but phase features exist on a circle, i.e., the phase wraps around causing problems for conventional computations. For example, suppose we have two phase measurements of $+175°$ and $-175°$. The mean of these numbers computed in the conventional way is $0°$, but because the phase wraps around, the correct mean is $180°$ (or $-180°$). The incorrect mean of $0°$ leads to erroneous variances and confounds ANOVA. We were able to use the commercial software by transforming any feature with a distribution crossing $180°$ by rotating the reference for that feature by $180°$, performing the analysis, and then rotating the result back. To perform the rotation we used the following formula:

$$\theta_{new} = \begin{cases} \theta - 180° & \text{if } \theta \geq 0 \\ \theta + 180° & \text{if } \theta < 0 \end{cases}.$$

This approach proved to be simple and effective while allowing us to capitalize on the reliability of commercial software.

We now show the analysis for the single phase feature, x_c, summarized in Table 3. The numbers given indicate the phase as a fraction of a circle's circumference. Therefore, all phases are in the range $[-0.5, 0.5]$. ANOVA computes the F statistic (the ratio of the between-person variance to the within-person variance) for each feature, and an associated probability. The probability indicates the likelihood that the null hypothesis is true. In this case, the null hypothesis is that the mean feature values for all the people are the same, namely

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6,$$

Table 3. Summary of experimental measurements of x_c phase feature. Each column is a different person.

Person	1	2	3	4	5	6
Seq. 1	−0.266	−0.270	−0.220	−0.264	−0.284	−0.345
Seq. 2	−0.271	−0.274	−0.275	−0.230	−0.319	−0.317
Seq. 3	−0.283	−0.271	−0.238	−0.255	−0.282	−0.303
Seq. 4	−0.264	−0.273	−0.244	−0.237	−0.288	−0.327
Seq. 5	−0.289	−0.258	−0.258	−0.255	−0.257	−0.309
Seq. 6	−0.313	−0.263	−0.270	−0.260	−0.208	−0.304
Seq. 7	−0.278	−0.264	−0.246	−0.265	−0.233	−0.318

Table 4. Results of Scheffe's post-hoc test for x_c. Each entry is the probability that the null hypothesis is true for the persons indicated on the row and column. The star (*) indicates a probability of the null hypothesis that is less than the arbitrary significance level of 5%. The null hypothesis is no significant difference between persons.

	2	3	4	5	6
1	0.93	0.20	0.26	0.91	0.06
2		0.75	0.84	>0.999	<0.01*
3			>0.99	0.78	<0.01*
4				0.86	<0.01*
5					<0.01*

Table 5. F-values and $P(H_0)$ for all phase features: low probabilities indicate that there is significant variation.

Variable	F-value	$P(H_0)$
x_c	10.38	<0.0001
y_c		Reference
x_{wc}	6.54	0.0002
y_{wc}	21.86	<0.0001
x_d	4.77	0.0019
x_d	4.77	<0.0001
y_d	31.16	<0.0001
a_c	17.28	<0.0001
a_{wc}	15.90	<0.0001
a_d	43.86	<0.0001
x_{uwc}	6.92	0.0001
y_{uwc}	4.85	0.0017
x_{vwc}	16.45	<0.0001
y_{vwc}	18.21	<0.0001

where μ_i is the mean for person i. For the data shown in Table 3, the F-value is 10.376, with degrees of freedom 5 and 36, yielding $P(H_0) < 0.0001$. It is therefore reasonable to conclude that at least one of the means is significantly different from the others.

While the F-value gives a reliable test of the null hypothesis, it cannot indicate which of the means is responsible for a significantly low probability. For this we use Scheffe's post-hoc test [12]. This test looks at all possible pairs of means and produces a probability that the null hypothesis, $H_0 : \mu_i = \mu_j$, for each pair is true. Table 4 summarizes Scheffe's post-hoc test for x_c. Asterisks in the table indicate values that are significant for an arbitrary significance level of 5%. Scheffe's test is fairly conservative, compensating for spurious significant results that occur with multiple comparisons [12]. Results in Table 4 indicate that significant variation in x_c occurs because person 6 differs from the others.

Table 5 summarizes the F-values for all of the phase features. Two features, y_d and a_d, exhibit a large number of significant probabilities.

Table 6. Results of Scheffe's post-hoc test for y_d. Each entry is the probability that the null hypothesis is true for the persons indicated on the row and column. The star (*) indicates a probability less than the arbitrary significance level of 5%.

	2	**3**	**4**	**5**	**6**
1	<0.0001*	0.0113*	<0.0001*	0.0025*	<0.0001*
2		<0.0001*	0.5588	<0.0001*	0.0081
3			0.0015*	0.9957	0.2444
4				0.0072*	0.3899
5					0.5328

Table 7. Results of Scheffe's post-hoc test for a_d. Each entry is the probability that the null hypothesis is true for the persons indicated on the row and column. The star (*) indicates a probability less than the arbitrary significance level of 5%.

	2	**3**	**4**	**5**	**6**
1	0.9756	0.0017*	<0.0001*	0.9591	0.0390*
2		0.0162*	<0.0001*	0.6108	0.0047*
3			0.0048*	0.0001*	<0.0001*
4				<0.0001*	<0.0001*
5					0.2543

The post-hoc tests for these features are shown in Tables 6 and 7. All features showed some significant variations with y_{wc}, y_d, a_d, and y_{vwc} showing the greatest variation.

4.2 Recognition

To test the usefulness of these features, we use the phase vectors in a simple recognition test. Each vector of twelve relative phase measurements is treated as a point in a twelve-dimensional space. We have seven image sequences for each of six subjects. The statistical analysis shows that there is significant variation among the subjects, in the phase features. Figure 19 plots a scattergram of two features, y_d and a_d; each symbol denotes the value of the phase features for a particular image sequence

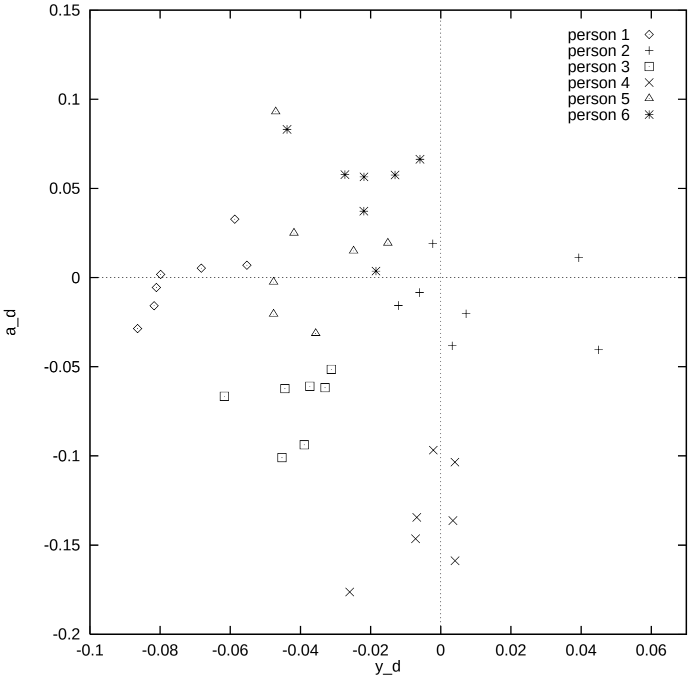
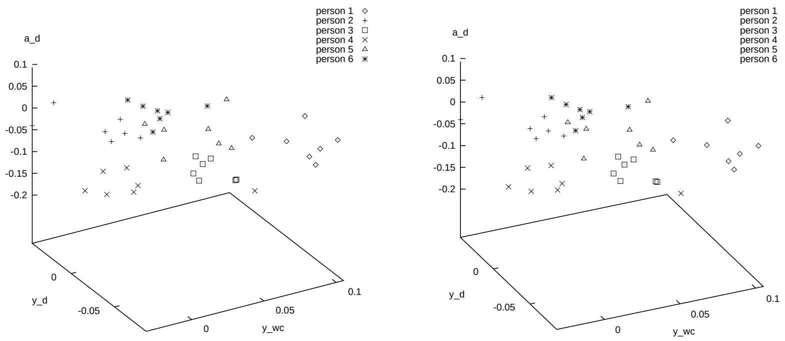


Figure 19. Plot of scatter of y_d versus a_d.



left right

Figure 20. Stereo plot of features a_d, y_d, and y_{wc}.

Table 8.

Description	**Label**	**Number(s)**
x coordinate of centroid	(x_c,y_c)	1
x and y coordinates of centroid of $\|(u, v)\|$ distribution	(x_{wc},y_{wc})	(2,3)
x and y coordinates of difference of centroids	(x_d,y_d)	(4,5)
aspect ratio (or elongation)	a_c	6
elongation of weighted ellipse	a_{wc}	7
difference of elongations	a_d	8
x and y coordinates of centroid of $\|u\|$ distribution	(x_{uwc},y_{uwc})	(9,10)
x and y coordinates of centroid of $\|v\|$ distribution	(x_{vwc},y_{vwc})	(11,12)

of a particular subject. It is clear from the scattergram that these features should be useful for discriminating among the subjects; it would be easy to introduce linear discrimination boundaries separating the feature points for the image sequences of one subject from its neighbors. Figure 20 shows a stereo plot of the three most statistically significant features, y_{wc}, y_d, and a_d.

Another way to visualize the data is to plot the phase vectors as twelve points in the phase range from -0.5 to 0.5. The twelve phase features are listed in Table 8.

Figure 21 shows several phase vectors for one subject. Most of the phase values vary little between image sequences. Figure 22 collects three phase vectors from image sequences of three different subjects. In comparison with Figure 21, there is substantial variation among the phase features across subjects.

In our recognition tests, we use the class mean of the seven feature vectors as an *exemplar* for the class. Figure 23 superimposes the phase vectors from two image sequences of one subject with its class exemplar. Finally, Figure 24 shows the class exemplars for all six subjects. The phase values show repeatability for a single subject and variation between subjects.

Figure 21. Phase vectors for two image sequences of subject 1.

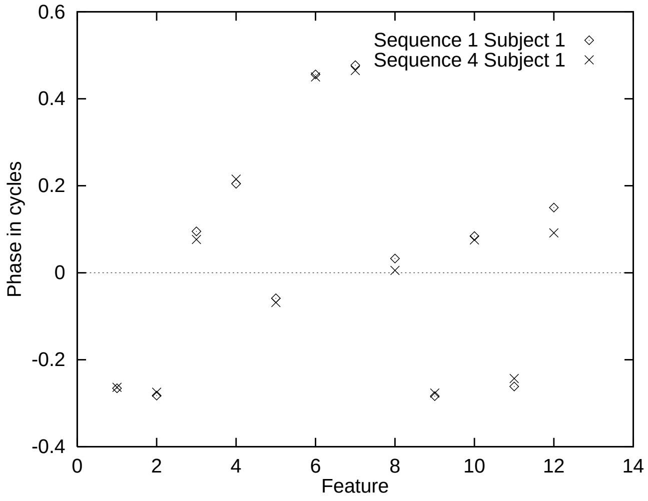
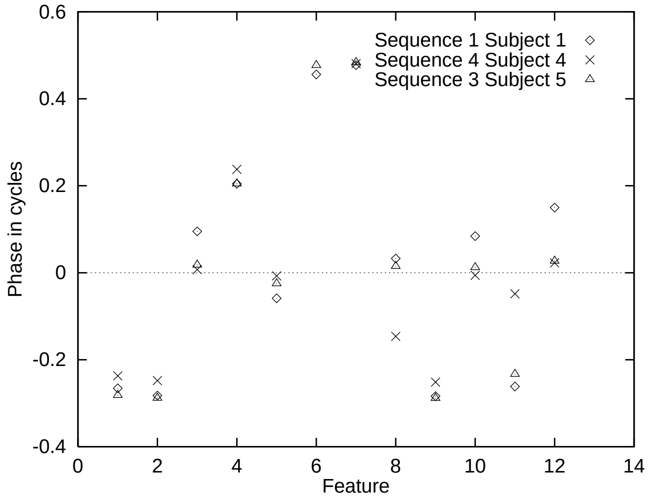


Figure 22. Phase vectors for image sequences of subjects 1, 4, and 5.



We tried three different simple classification methods: nearest-neighbor, k-nearest-neighbors, and nearest-neighbor with class exemplars. To compare phase vectors, we sum the squared elementwise differences of phase angles, adjusted for the phase wraparound at $0.5 = -0.5$. In the nearest-neighbor test (NN), each phase vector is classified as belonging to the class of its nearest neighbor feature vector. In the k-nearest-neighbor test, $k = 3$ (3NN), we find the three nearest neighboring vectors, and choose the class of the majority, or, if no majority, simply the nearest neighbor. The exemplar method (exNN) classifies a vector as the class of its nearest-neighbor exemplar or class mean. In all our tests the exemplar method behaves best so we will report only its results.

We have six people, each with seven image sequences. For a small number of examples, such as our 42, the literature suggests computing an unbiased estimate of the true recognition rate using a *leaving-one-out* crossvalidation method [44]. We leave one example out, train on the rest (compute exemplars), and then classify the omitted element using these exemplars. We perform this experiment for each of the 42 examples, and report the number of correct classifications.

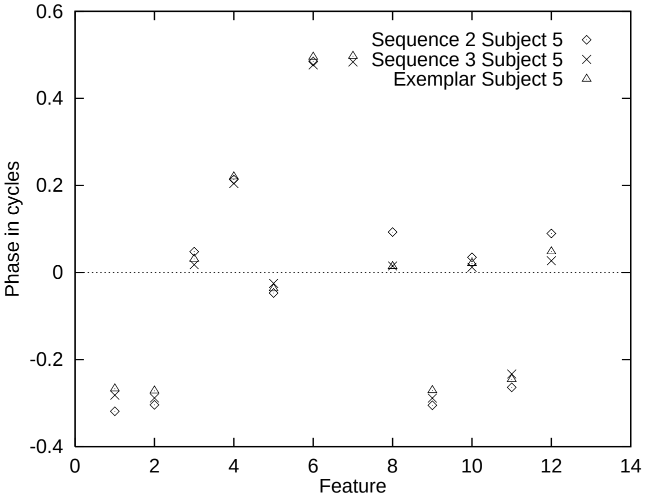


Figure 23. Phase vectors for two image sequences of subject 5, plus the exemplar for the subject.

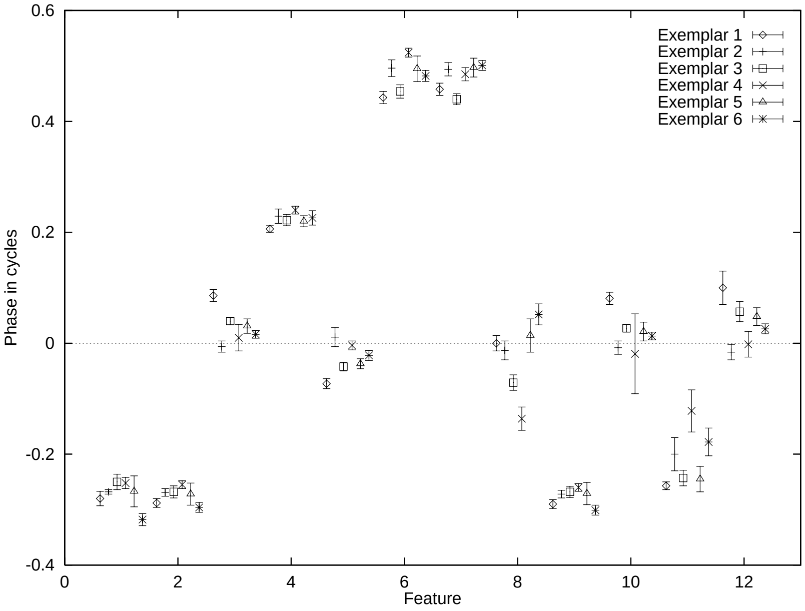


Figure 24. Phase vectors for the exemplars of all six subjects. Error bars show the standard deviation for each subject.

Using the full feature vector we achieved a recognition rate of 90.5%, but we can do slightly better using a subset of the features. We performed a full combinatorial experiment in which all subsets of features were tested for recognition rate using various crossvalidation methods. Table 9 shows the results obtained by leaving one vector out of the training and testing it against the exemplars, for the full possible range of features.

To make the results even stronger, we performed crossvalidation using fewer of the data for training and testing more examples. We used the simple procedure of treating the seven examples for each person as an index set: (1234567), and selecting one, two, or three sequences (by index) for each person, training on the remaining six, five, or four, and testing. We use all possible choices of size one, two, or three from seven, for each of six people. The number of tests is, for $s = 7$ sequences, $p = 6$ people, and k test vectors, $\binom{s}{k} * p * k$: for $k = 1$, $42 = \binom{7}{1} * 6 * 1 = 7 * 6 * 1$; for $k = 2$, $252 = \binom{7}{2} * 6 * 2 = 21 * 6 * 2$; and for $k = 3$, $630 =$

Table 9. Results of recognition, using leave-one-out crossvalidation, with the number of features used and percentage correct.

Number	Features	Percent correct
1	a_d	66.7
2	a_d, y_{vwc}	85.7
3	y_{wc}, a_d, y_{vwc}	90.5
4	y_d, a_c, a_d, y_{vwc}	92.9
5	x_d, y_d, a_{wc}, a_d, y_{vwc}	95.2
6	x_d, y_d, a_{wc}, a_d, x_{uwc}, y_{vwc}	95.2

Table 10. Subsets of size $k = 1$: results of recognition, with the number of features used and percentage correct.

Number	Features	Percent correct
1	a_d	66.7
2	y_{wc}, a_d	85.7
3	y_{wc}, a_d, y_{vwc}	90.5
4	y_d, a_{wc}, a_d, y_{vwc}	95.2
5	y_d, a_{wc}, a_d, x_{uwc}, y_{vwc}	95.2

Table 11. Subsets of size $k = 2$: results of recognition, with the number of features used and percentage correct.

Number	Features	Percent correct
1	a_d	61.9
2	a_d, y_{vwc}	85.3
3	y_{wc}, a_d, y_{vwc}	88.5
4	y_d, a_{wc}, a_d, y_{vwc}	91.7
5	x_c, y_d, a_c, a_d, y_{vwc}	93.7

$\binom{7}{3} * 6 * 3 = 35 * 6 * 3$. Again we perform full combinatorial experiments and find the best features. Table 10 shows the results for $k = 1$, Table 11 for $k = 2$, and Table 12 for $k = 3$.

Even when almost half of the data are omitted, the recognition rates remain over 90% when five features are used out of twelve. Only three features are needed for excellent success rates.

The analysis of variance predicts that the features will have the following approximate significance:

$$a_d > y_d > y_{wc} > y_{vwc} > a_c > x_{vwc} >$$
$$a_{wc} > x_c > x_{uwc} > x_{wc} > y_{uwc} > x_d.$$

The scattergram of y_d and a_d indicates why these features are actually useful in recognition (Figure 19); the separation of subjects using these features is quite good. When we consider triples of features, using exNN, the subsets of features that showed best recognition rates were (y_{wc}, a_d, y_{vwc}), (y_d, a_d, y_{vwc}), and (y_{wc}, y_d, a_d). These correspond well with the

Table 12. Subsets of size $k = 3$: results of recognition, with the number of features used and percentage correct.

Number	Features	Percent correct
1	a_d	60.5
2	a_d, y_{vwc}	84.6
3	y_{wc}, a_d, y_{vwc}	87.6
4	y_d, a_{wc}, a_d, y_{vwc}	89.8
5	x_c, y_d, a_c, a_d, y_{vwc}	92.2

features that statistical analysis predicts will be useful. The F-value computed by ANOVA indicates whether or not a single feature is useful for classification. It may be possible to transform multiple features to obtain better classification features [13], but ANOVA indicates the minimum that we should expect.

We have tested using the direct Euclidean distance and also the Mahalanobis distance (which scales the difference in each coordinate by the inverse of the variance in that dimension) [24], but the Euclidean works better. With this little data the variance estimates are unreliable.

5 Discussion

5.1 Comparison with Other Methods

The other techniques for representing human motion have been applied to recognizing activities [26], but only Niyogi and Adelson [27] have specifically tried to recognize individuals by their motion. Our imaging situation is exactly the same they used to achieve a recognition rate of up to 83%. There is no reason to expect that their method would not work well with this data.

Niyogi and Adelson acquire contours by examining spatiotemporal solids of adjoined images. At a particular height y, the bounding contours found in xyt form a trace over a sequence in t, yielding a vector $x(y,t)$. They then adjust for overall x translation, find the period by looking at the maximum x extent, and linearly interpolate the $x(y,t)$ signals at a fixed relative height.

Finally, they use a simple Euclidean distance or a more robust weighted distance between $x(y,t)$ vectors for recognition.

Our system achieves higher recognition rates, but it should be possible to equal our results with their system. Our system is much more general, however, in that it is not model based, and could apply equally to the motion of animals. Our results show that it is unnecessary to assume a model-based interpretation of the moving figure. Instead, the shape of motion retains much of the relevant information about the motion of the figure and can be analyzed to recover phase features that allow discrimination among individuals. We do expect that discrimination would suffer when the database of individuals became large; simple motion cues could identify types of gaits, but would no longer uniquely identify an individual.

5.2 Effects of Flow Computation

The results presented in this paper reflect flow computed using constant spatial support for flow correspondence. However, in the course of our investigations we tried computing the flow with varying amounts of spatial support. Limited spatial support gives a high spatial resolution in the flow but is more susceptible to noise. In trials with limited spatial support we had to segment the foreground from the background to mask background noise. Broader support obviates the requirement for foreground separation but reduces the spatial resolution of the flow. Recognition was equally effective, provided that flow parameters were kept constant, but there were variations in the results that were worth noting.

The coarse, low-noise flow (broad support) yielded phase features that had lower between-subject variations. In other words, the subjects looked more alike. At the same time, however, variations for individual subjects dropped, and recognition still worked. Changing the resolution of the flow changed the set of features showing the most-significant variation. Features that were useful for fine flow were not necessarily viable for coarse flow, and vice versa. The key factor is that whatever combination of flow algorithm and parameters is used, that combination must be held constant for recognition to work.

The figures appearing in this paper are for flow with relatively high spatial resolution. Elements of a figure such as moving arms or legs have high velocities, and change their spatial orientation during movement, unlike the trunk. Thus, flow values of the center of the body are less subject to noise. With better spatiotemporal resolution, the limbs are better detected in the flow. This suggests that the features of flow that depend on the spatial distribution of velocities would be more important. This is exactly what we found. In these tests, the phase features that were most effective in recognition included a_d, y_d, y_{wc}, and y_{vwc}.

5.3 Intuitive Understanding of Phase Features

After statistical analysis and testing of classification methods, we must consider why these phase features allow recognition. At present we have no provable explanation, but we offer the following speculation based on our experimental results and intuition.

The phase features we compute are scale independent. They should not vary with the distance between the camera and the subject. No scaling of the data is required. The phase of the scalar signals remains unchanged if the subject is small or large in the camera field of view.

We expected that the difference signals, x_d, y_d, and a_d, would work very well for recognition. These signals can convey information about the relative magnitudes of the two signals they are derived from, and perhaps in doing so may convey information about the build of the subject. The phase of a sum or difference of two sinusoids gives an indication of their relative magnitude. The phasor diagram in Figure 25 illustrates this point.[3] The figure shows the phasor subtraction of two signals, A and B, to yield the difference C. Although the phase relationship between A and B is the same in both (a) and (b) of Figure 25, the difference in the relative magnitude causes the phase of the difference, C, to be different. Therefore, phase features are not only scale independent, but also phases of signal differences are sensitive to the

[3] A phasor, or phase vector, is a rotating vector that represents a quantity that varies sinusoidally. They are useful in the analysis of systems in which the signals share a common frequency but vary in magnitude and phase, e.g., alternating current electrical systems.

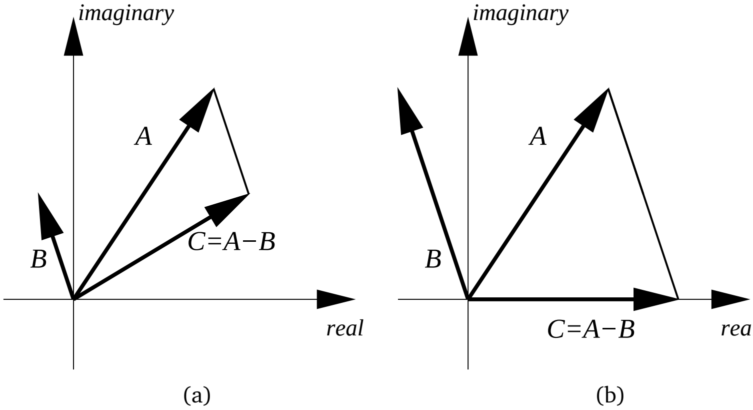


Figure 25. Phasor diagram illustrating the dependence of phasor difference on relative magnitude: (a) example of phasor subtraction $C = A - B$, and (b) altering only the relative magnitude of A and B gives a different phase for the difference.

relative magnitudes of the scalar signals. This means that people of differing builds but with similar rhythms might still be distinguished by our analysis.

Our expectations were largely realized in our results: a_d and y_d were excellent for recognition where high spatiotemporal resolution was used, but x_d was poor. Our results suggest that, while our notions about the usefulness of the difference signals may be valid, we still need to understand better the effects of flow calculations on the features.

5.4 Limitations of Results

Although the analysis of our results is valid, we are limited in our ability to extrapolate them. Our sample size was small (only six subjects) so we cannot conclude much about gaits in the entire human population. In addition to the small sample size, no steps were taken to ensure a random sample. However, the excellent recognition results and strong statistical evidence of variation between subjects attest to the value of shape-of-motion features.

As mentioned in Sections 5.1 and 5.2, the parameters used in the computation of optical flow can affect the values of the phase features. Although we believe that varying parameters and methods for computing optical flow will produce features useful for recognition, we do not yet understand enough to predict which features will be the best.

It is entirely possible, even likely, that there are correlations among the various features. We have not analyzed the data extensively to identify these correlations.

In an effort to control the conditions of the experiment, we considered only pedestrians walking across the camera field of view. There is no reason to expect that shape-of-motion features are invariant to viewing angle, so we expect recognition—using the features we have acquired—to fail if the view is not approximately perpendicular to the direction of the gait. Some of our preliminary results suggest that recognition is possible for pedestrians walking towards and away from the camera. Our experiments included only three sequences each for two subjects; we were able to classify five out of six for a sequence in which the subject walks across the field of view, but with some motion towards the camera. For sequences walking towards and away from the camera, we correctly classified all six sequences.

However, the exemplars are not similar in the two cases. When motion is perpendicular to the viewing direction, self-occlusion is maximal, but the figure's relative motion is largest. Motion along the viewing direction eliminates self-occlusion, but the relative motion is less. In surveillance applications where multiple cameras are employed, this is not likely to be a problem. One simply selects the images from the camera with the best viewing angle for recognition. Moreover, it is possible to identify which direction an individual is walking and use that information to select the appropriate set of exemplars tuned to the direction of movement.

The controlled experimental situation eliminates other considerations, such as moving backgrounds, tracking cameras, and occlusions.

5.5 Future Work

The experiment described here attempts to eliminate confounding factors from the experiment by acquiring all image sequences under identical circumstances. There remains the task of determining what effect other factors may have, such as viewing angle, clothing of subjects, and time.

For practical application of shape-of-motion features, we need to know the useful range of camera views over which the recognition will work for a single set of exemplars. A useful experiment would be to determine the sensitivity of the features to viewing angle. The results would enable a multicamera surveillance system to select an optimal view for recognition.

People have ways of subjectively describing the way a person walks. A person may have a bouncy walk or perhaps a shuffle. One may infer that a person is excited or lethargic based on such observations. A future experiment may determine if the phase features of walking can be correlated to subjective descriptions of the way a person walks. The result would be a feature space that is segmented based on subjective descriptions. Some human-figure animation work in computer graphics by Unuma et al suggest that this may be possible [42]. They model a human gait using a set of periodic joint-angle functions for several subjectively different gaits. They then interpolate Fourier decompositions of these signals to generate a variety of gait types. Based on this idea, our proposed model-free recognition method may be able not only to recognize, but describe too.

Model-free recognition may be applied to the domain of image databases. The method could search a large database of image sequences for gaits of various types. For example, one could search a database for sequences that contain periodic motion. Then, if the moving region has all the correct phase relationships, one may conclude that the sequence contains a person walking. Further analysis may allow one to search for people who walk like a given person, or exhibit certain characteristics in their gait.

Moving light displays (MLDs) have seen extensive use in gait analysis, and experiments with them have shown that recognition by humans is possible using the MLD images alone. The focus of computer analysis of MLDs has been on the construction of a kinematic model of the subject. Our results suggest that this may not be necessary for recognition. Optical flow computed from MLD images may be viewed as a point sampling of the full flow field. If the sampling is sufficient to estimate the scalars used in our recognition system, then model-free recognition is possible. In related work [8], we have examined the relationships between the MLD flow and full flow images and found that the phase features have values, when derived from the MLD images, that are similar to the values using full gray-value images. This suggests that there is no need to determine an articulated model to interpret MLD images.

6 Summary

The spatial distribution of optical flow, the shape of motion, yields model-free features whose variation over time is periodic. The essential difference among these features is the relative phase between the periodic signals. Our experiments and analysis demonstrate that these phase measurements are repeatable for particular subjects and vary significantly between subjects. The variation makes the features effective for recognizing individuals by characteristics of their gait, and the recognition is relatively insensitive to the means for computing optical flow. The phase analysis applies directly to periodic motion events, but the flow description, the shape of motion, applies to other motions.

7 Acknowledgments

We wish to thank the graduate students at the UCSD Visual Computing Laboratory for their participation as subjects in our experiments. Also, Dr. J. Edwin Boyd provided valuable advice on the use of ANOVA. Thanks also to Don Murray for valuable suggestions about the manuscript.

References

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

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

[3] Baumberg, A. M. and Hogg, D. C. Learning flexible models from image sequences. Technical Report 93.36, University of Leeds School of Computer Studies, 1993.

[4] Baumberg, A. M. and Hogg, D. C. Learning spatiotemporal models from training examples. Technical Report 95.9, University of Leeds School of Computer Studies, 1995.

[5] Bharatkumar, A. G., Daigle, K. E., Pandy, M. G., Cai, Q., and Aggarwal, J. K. Lower limb kinematics of human walking with the medial axis transformation. In *IEEE Workshop on Nonrigid Motion*, pages 70–76, 1994.

[6] Bobick, A. F. and Davis, J. W. An appearance-based representation of action. In *Proc. 13th International Conference on Pattern Recognition*, 1996.

[7] Bobick, A. F. and Davis, J. W. Real-time recognition of activity using temporal templates. In *Workshop on Applications of Computer Vision, 1996*.

[8] Boyd, J. E. and Little, J. Global vs. segmented interpretation of motion: Multiple light displays. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 18–25, 1997.

[9] Bregler, C. Learning and recognizing human dynamics in video sequences. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1997*, pages 568–574, 1997.

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

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

[12] Cohen, P. R. *Empirical Methods for Artificial Intelligence*. The MIT Press, Cambridge, MA, 1995.

[13] Cui, Y., Swets, D., and Weng, J. Learning-based hand sign recognition using shoslif-m. In *Proc. 5th International Conference on Computer Vision*, pages 631–636, 1995.

[14] Davis, J. W. and Bobick, A. F. The representation and recognition of human movement using temporal templates. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1997*, pages 928–934, 1997.

[15] Fua, P. A parallel stereo algorithm that produces dense depth maps and preserves image features. Technical Report 1369, INRIA, 1991.

[16] Hogg, D. C. A program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

[17] Hu, M. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, IT-8(2), 1962.

[18] Hunter, E. A., Kelly, P. H., and Jain, R. C. Estimation of articulated motion using kinematics of constrained mixture densities. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 10–17, 1997.

[19] Ju, S. X., Black, M. J., and Yacoob, Y. Cardboard people: A parameterized model of articulated motion. In *2nd Int. Conf. on Automatic Face- and Gesture-Recognition*, pages 38–44, Killington, Vermont, 1996.

[20] Little, J. J. and Boyd, J. E. Describing motion for recognition. In *IEEE Symposium on Computer Vision*, pages 235–240, 1995.

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

[22] Liu, F. and Picard, R. W. Detecting and segmenting periodic motion. Technical Report 400, MIT Media Lab Perceptual Computing Section, 1996.

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

[24] Mahalanobis, P. On the generalized distance in statistics. *Proc. Natl. Inst. Science*, 12:49–55, 1936.

[25] Marr, D. and Poggio, T. Cooperative computation of stereo disparity. *Science*, 194(4262):283–287, 1976.

[26] Nelson, R. C. and Polana, R. Qualitative recognition of motion from temporal texture. *Journal of Visual Communication and Image Representation*, 5:172–180, 1994.

[27] Niyogi, S. A. and Adelson, E. H. Analyzing and recognizing walking figures in XYT. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994*, pages 469–474, 1994.

[28] Niyogi, S. A. and Adelson, E. H. Analyzing gait with spatiotemporal surfaces. In *IEEE Workshop on Nonrigid Motion*, pages 64–69, 1994.

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

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

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

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

[33] Polana, R. and Nelson, R. Nonparametric recognition of nonrigid motion. Technical Report TR575, University of Rochester, 1995.

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

[35] Rohr, K. Towards model-based recognition of human movements in image sequences. *Computer Vision, Graphics, and Image Processing*, 59(1):94–115, 1994.

[36] Rowley, H. A. and Rehg, J. M. Analyzing articulated motion using expectation maximization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1997*, pages 935–941, 1997.

[37] S. Lawrence Marple, J. *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1987.

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

[39] Shavit, E. and Jepson, A. Motion understanding using phase portraits. In *IJCAI Workshop: Looking at People*, 1993.

[40] Shavit, E. and Jepson, A. Qualitative motion from visual dynamics. In *IEEE Workshop on Qualitative Vision*, pages 82–88, 1993.

[41] Teh, C. H. and Chin, R. T. On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:496–513, 1988.

[42] Unuma, M., Anjyo, K., and Takeuchi, R. Fourier principles for emotion-based human figure animation. In *Proceedings of SIGRAPH 95*, pages 91–96, 1995.

[43] Wachter, S. and Nagel, H.-H. Tracking of persons in monocular image sequences. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 1–9, 1997.

[44] Weiss, S. M. and Kulikowski, C. A. *Computer Systems That Learn*. Morgan Kaufmann, 1991.

[45] Winer, B. J. *Statistical Principles in Experimental Design*. McGraw-Hill, New York, 1971.

Editors in Chief

Editorial Board

Action Editors