Discrimination of Motion Based on Traces in the Space of Probability Functions over Feature Relations *

Sudeep Sarkar and Isidro Robledo Vega†
Computer Science and Engineering Department
University of South Florida
Tampa, FL 33620
{sarkar, irobledo}@csee.usf.edu

Abstract

In this paper we demonstrate that it is possible to discriminate between high level motion types such as walking, jogging, or running based on just the change in the relational statistics among the detected image features, without the need for object models, perfect segmentation, or tracking. Instead of the statistics of the feature attributes themselves, we consider the distribution of the statistics of the relations among the features. We represent the observed distribution of feature relations in an image as a point in a space where the Euclidean distance is related to the Bhattacharya distance between probability functions. Different motion types sweep out different traces in this Space of Probability Functions (SoPF). We demonstrate the effectiveness of this representation on image sequences of human in motion, gathered using a digital video camera. We show that it is not only possible to distinguish between motion types but also to discriminate between persons based on the SoPF traces.

1. Introduction

High level complex motion analysis need not be contingent on object recognition [5, 16]. Rather, motion analysis can aid recognition of objects, thus giving rise to an approach called *motion based recognition*. One of the problems in motion analysis is *motion recognition*, which can be used for motion based recognition. Some of the tasks in motion recognition involve distinguishing rigid motion from non-rigid motion [15], recognizing human motion types such as walking, running, jogging, etc. [12], or computing the periodicity of motion [11, 14, 6]. Some motion recognition strategies require feature correspondences in terms of optic flow fields [11, 3, 8] or object parts [7]. These methods typically suffer from non-robustness with respect to noise, image resolution, and large motion. To tackle this, some approaches rely on more area based measures, such as image or object self-similarity [6, 14].

In another area of computer vision, namely object recognition, there has been recent push for the use of statistical methods for recognition instead of the traditional model based alignment and matching methods. For instance, there have been in-depth studies into statistical features that are appropriate for modeling natural images [10]. Multidimensional histograms of feature distributions have been used for recognition and shape modeling [1, 13]. Strong models in terms of Markov Random Fields have been proposed for statistical modeling of images [17]. In this paper, we propose a strategy based on statistical modeling of images to bear on the problem of motion recognition. To be precise, we present the formalism of Space of Probability Functions (SoPF) over *relational* parameters, in which traces of the observed relational statistics are used to recognize high level motion.

With motion, the statistics of the relationships among the 2D image features change. This change or non-stationarity in relational statistics is not random, but follow the motion pattern. The shape of the probability function governing the distribution of the inter-feature relations, which can be estimated by the normalized histogram of observed values, change as parts of the object move. We suggest the use of a space over these probability functions, which we refer to as the SoPF (Space of Probability Functions), to study the trend of change in their shapes. Distances in this space are related to the Bhattacharya distance between probability mass functions. Each motion type creates a trace in this space. The attractive aspects of this approach are that (i) it does not require perfect segmentation of the object from the background, (ii) it does not require feature tracking, (iii) it is amenable to learning, and (iv) there is no assumption about single pixel movement between frames.

*This work was supported by the National Science Foundation grants No. IIS-9907141 and EIA-9729904.
†His PhD studies are being supported by CONACYT-SEP-Mexico.

0-7695-1272-0/01 $10.00 © 2001 IEEE

In the next section, we outline the concept of the relational distribution of an image, followed by a section describing the Space of Probability Functions (SoPF). We present results on motion sequences of three persons, each executing three different kinds of movements (walking, jogging, and running).

2. Relational Distributions

We view an image as an assemblage of low level features, such as edge pixels, corners, straight lines, or region patches. The structure perceived in an image is determined more by the relationship among features than by the individual feature attributes themselves. Our goal is to device a mechanism to capture this structure so that we can use its variation with time to recognize high level motion. Graphs (or hypergraphs) have been the most commonly used mechanism of capturing relationships among features. However, the study of variation of a graph over time requires solving the correspondence problem between features, which is a computationally difficult problem. We avoid the need for correspondence by using just the statistical distribution of the relational attributes observed in the image.

Let $\mathcal{F} = \{f_1, \cdots, f_N\}$ represent the set of features in an image. Let $\mathbf{F_k}$ represent a random $k-$tuple of features. The relationship among these k-tuple features is denoted by R_k. Thus, the 2-ary relationship between features, which is the most commonly used form, will be represented by R_2. Notice that, low-order correlation is captured by smaller values of k and higher-orders of correlation are captured by larger values of k. In a set of N primitive features there are of course C_k^N possible $k-$tuples. Let these relationships R_k be characterized by a set of attributes $\mathbf{A_k} = \{A_k(1), \cdots, A_k(M)\}$. Then the shape of the object is represented by joint probability functions: $P_k(\mathbf{A_k} = \mathbf{a_k}) = P_k(a_k(1), \cdots, a_k(M)) = P_k(\mathbf{a_k})$, where $a_k(i)$ is the (discretized, in practice) value taken by the relational attribute $A_k(i)$. A feature in an image can be looked upon as a random outcome governed by these probability functions. The attribute of each feature is a random outcome of $P_1(\mathbf{a_1})$. Each pair of image features is a random outcome drawn from $P_2(\mathbf{a_2})$. And, each image feature triple is a random outcome of $P_3(\mathbf{a_3})$. We term the set of probability functions $\{P_k(\mathbf{a_k}) | k = 1, \cdots, L\}$ as the *Relational Distributions*. We estimate these distributions, based on a training set of images, by the normalized histogram of the attributes of every k-tuple ($k \geq 1$) of observed features. Other sophisticated strategies based on parameterized functional forms and kernel densities can also be adopted. However, the simpler histogram based strategy is sufficient for now.

2.1. Edge Based Features

We illustrate our ideas in this paper by considering edge pixels detected in an image as the features. Other features

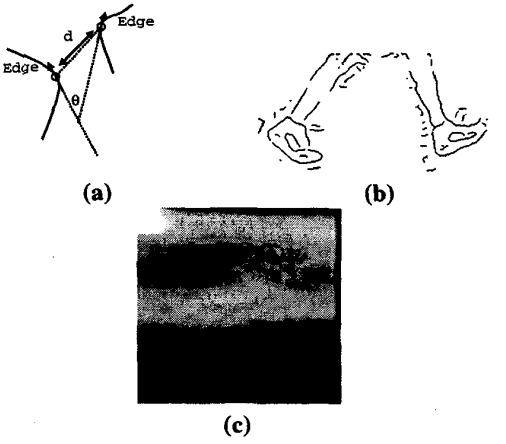

Figure 1. Edge pixel based 2-ary relational distribution. (a) The two attributes characterizing relationship between two edge pixels. (b) Edge pixels in an image. (c) The distribution $P_2(d/D, \theta)$, where D captures the image size. $P_2(0,0)$ is the top left corner of the image. Brighter values denote higher probabilities.

types such as the neurally inspired keys [9] or those based on the Gaussian derivatives [13] will be subjects of future studies. We detect edges in a image by the Canny edge detector followed by hysteresis thresholding. Each edge pixel, f_i, is associated with the gradient direction, θ_i, at that point. To capture the structure between edge pixels, we use the difference in edge orientations and the distance between the two edge pixels as the attributes, $\{A_2(1), A_2(2)\}$, of R_2. We normalize the distance between the pixels by the image size (D) to make it scale invariant. Note that our choice of the attributes is such that the probability representation is invariant with respect to scale, image plane rotation, and translation. Fig. 1(a) depicts the attributes that are computed between the two pixels. And, Fig. 1(c) shows the $P_2(a_1, a_2)$ for the edge image shown in Fig. 1(b), where high probabilities are shown as brighter values. Note the concentration of high values in certain regions of the probability event space.

To capture the relational distribution over triples of edge pixels, $P_3(a_1, a_2, a_3, a_4)$, we use four attributes, as illustrated in Fig. 2(a). Since all pairs of distances in the triplet are not independent on each other, attributes over all pairs would not constitute an independent set of attributes. Hence, we consider the pairs of pixels that are connected by the maximum spanning tree over them, which for Fig. 2(a) are d_{12} and d_{13}. Fig. 2(b) shows the four dimensional probability density $P_3(d_{12}/D, d_{13}/D, \theta_{12}, \theta_{13})$ shown as a 2D image for the edge image in Fig. 1(b). Note that most of the

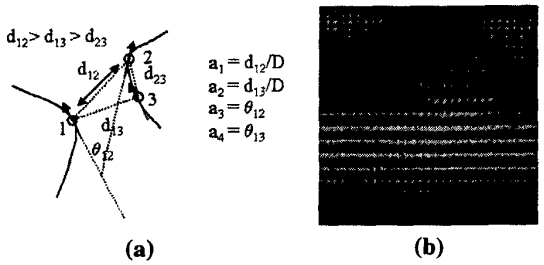

Figure 2. Edge pixel based 3-ary relational distribution. (a) The four attributes characterizing the relationship between three edge pixels. (b) The four dimensional relational distribution $P_3(d_{12}/D, d_{13}/D, \theta_{12}, \theta_{13})$ shown as a 2D image for the edge image in Fig. 1(b). The columns correspond to the row scanned version of the $(d_{12}/D, d_{13}/D)$ subspaces. The rows correspond to the row scanned version of the $(\theta_{12}, \theta_{13})$ subspaces. Only non-zero rows are shown. $P_3(0,0,0,0)$ is the top left corner of the image.

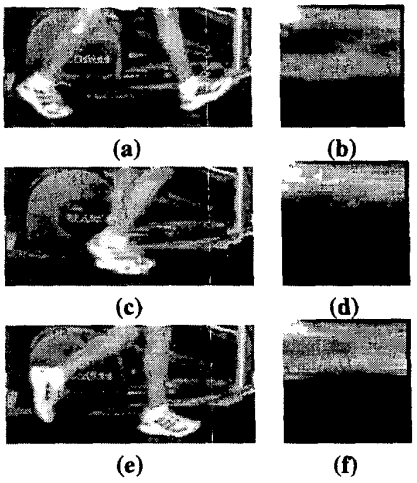

Figure 3. Some configurations of legs in motion with their corresponding 2-ary relational distributions.

probabilities are concentrated only in certain portions of the event space. As compared to the 2-ary distribution, the 3-ary distribution has a more concentrated shape. Entropy of the 2-ary distribution is 0.92 and that for the 3-ary is 0.83.

3. Space of Probability Functions (SoPF)

As the parts of an articulated object move, the relational distributions will change. Motion will introduce non-stationarity in the relational distributions over the image. Fig. 3 shows some example 2-ary relational distributions for some leg configurations. Notice how the modes of the probability functions, identified as the bright regions in the images, change with leg motion. By quantifying the evolution of the nature of these non-stationarities, we should be able to infer the nature of articulated motion. We should not only be able to distinguish periodic motion from non-periodic one, but we should also be able to make judgment about the nature of motion, e.g. walking versus jogging.

We model the relational distributions using a space of probability mass functions, referred to as the SoPF. Each relational distribution of an image is a point in this space. We construct this space such that the Euclidean distance between points representing two relational distributions, $P_{k,1}(\mathbf{a_k})$ and $P_{k,2}(\mathbf{a_k})$, is related to the Bhattacharya distance. The Bhattacharya distance between the two mass functions is given by $-\ln \sum \sqrt{P_{k,1}(\mathbf{a_k}) P_{k,2}(\mathbf{a_k})}$.

We arrive at the SoPF by principal component analysis. Let $P_{k,t_i}(\mathbf{a_k})$ represent the relational distribution at time t_i. We describe the *square root* of each relational distribution as a linear combination of orthogonal basis functions as follows. (We use the square root so that we arrive at a space where the distances are related to the Bhattacharya distance, which we will shortly prove.)

$$\sqrt{P_{k,t_i}(\mathbf{a_k})} = \sum_{i=1}^{N} c_i(t_i)\Phi_i(\mathbf{a_k}) + \mu(\mathbf{a_k}) + \eta(\mathbf{a_k}) \quad (1)$$

where $\Phi_i(\mathbf{a_k})$'s are orthonormal functions. The function $\mu(\mathbf{a_k})$ is a mean function defined over the attribute space. And $\eta(\mathbf{a_k})$ is a function capturing small random noise variations with zero mean and small variance. The orthonormal functions, Φ_i's, span a function space that captures the variations in the P_ks. We refer to this space as the SoPF.

Given a training set of mass functions, $\{P_{k,i}(\mathbf{a_k})|i = 1, \cdots, N\}$, this SoPF can be setup using the well-defined theory of principal component analysis (PCA), for whose details we refer the reader to standard texts. The dimensions of the SoPF are given by the eigenvectors of the covariance of the square root of the given relational distributions. The variance along each dimension is proportional to the eigenvalues associated with it. In practice, we can consider the subspace spanned by a few dominant vectors associated with the large eigenvalues. We have found that for human motion just 10 eigenvectors are sufficient.

The relational distribution at time t, is characterized by the coordinates $c_i(t)$'s in the SoPF. These coordinates can be obtained by projecting the given density onto each of the basis functions.

$$c_i(t) = \sum_{\mathbf{a_k}} \Phi_i(\mathbf{a_k}) \left(\sqrt{P_{k,t}(\mathbf{a_k})} - \mu(\mathbf{a_k}) \right) \quad (2)$$

Distance between two relational distributions, $P_{k,1}(\mathbf{a_k})$ and $P_{k,2}(\mathbf{a_k})$, in the SoPF will be given by the Euclidean distance between the coordinates representing the two, $\{c_{i1}|i = 1 \cdots N\}$ and $\{c_{i2}|i = 1 \cdots N\}$. Thus,

$$\text{Distance}(P_{k,1}(\mathbf{a_k}), P_{k,2}(\mathbf{a_k})) = \sum_{i=1}^{N}(c_{i1} - c_{i2})^2 \quad (3)$$

Using Eq. 1 and the orthogonal property of the spanning functions, $\Phi_i(\mathbf{a_k})$'s, we can show that the above distance is related to the Bhattacharya distance.

$$\begin{aligned} &\text{Distance}(P_{k,1}(\mathbf{a_k}), P_{k,2}(\mathbf{a_k})) = \textstyle\sum_{i=1}^{N}(c_{i1} - c_{i2})^2 \\ = \;& \textstyle\sum_{\mathbf{a_k}} \left(\sqrt{P_{k,1}(\mathbf{a_k})} - \sqrt{P_{k,2}(\mathbf{a_k})}\right)^2 \\ = \;& \textstyle\sum_{\mathbf{a_k}} \left(P_{k,1}(\mathbf{a_k}) + P_{k,2}(\mathbf{a_k}) - 2\sqrt{P_{k,1}(\mathbf{a_k})P_{k,2}(\mathbf{a_k})}\right) \\ = \;& 2 - 2\textstyle\sum_{\mathbf{a_k}} \sqrt{P_{k,1}(\mathbf{a_k})P_{k,2}(\mathbf{a_k})} \\ = \;& 2(1 - e^{-\text{Bhattacharya Distance}(P_{k,1}, P_{k,2})}) \end{aligned} \quad (4)$$

Note that Black and Jepson [2] also used PCA in the context of tracking and matching moving objects. However, our use of PCA is very different, we use PCA over the space of relational statistics whereas their use of PCA is over the image pixel space.

4. Motion Discrimination

Articulated motion sweeps a path or trace through the SoPF, denoted by the coordinates $\mathbf{c(t)} = \{c_1(t), \cdots, c_k(t)\}$. There are various sophisticated techniques such as those based on Hidden Markov Models, dynamic Bayesian networks, and State Space trajectories [4] that can be used to recognize particular motion types. One could use these formalisms and use the SoPF coordinates as inputs. In this paper, however, we adopt a simpler measure of warped distance between two traces to demonstrate the viability of using the traced paths to discriminate between motion types.

The warped distance is based on Dynamic Time Warping. It is computed over one motion cycle by warping one trace onto the other and then computing the difference. Let two traces over one motion cycle be denoted by $\{\mathbf{c_1}(i\delta t)|i = 1, \cdots, m\}$ and $\{\mathbf{c_2}(j\delta t)|j = 1, \cdots, n\}$, where δt is the time difference between image frames. Without loss of generality, let us assume that the first trace has lesser number of points, i.e. $m \leq n$. The distance between these traces is computed by first constructing a continuous curve, $\mathbf{C_1}(t)$ from the first trace by assuming linear interpolation between the coordinate points. Next we stretch this curve such that the first and the last coordinates match with the second curve, i.e. $\mathbf{C_1}(mt/n)$. Then we compute the distance between the second trace coordinate points and this stretched curve.

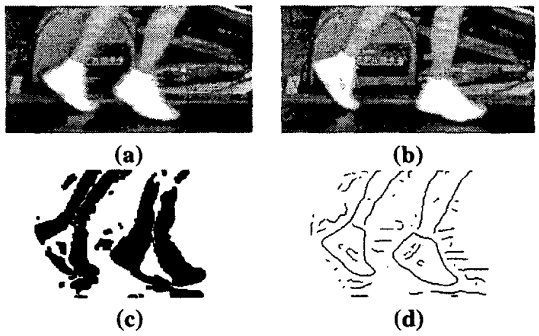

Figure 4. Two consecutive frames from a running sequence are shown in (a) and (b). The thresholded difference image is shown in (c). (d) The segmented motion edges.

$$\text{WD}(\mathbf{c_1}(i\delta t), \mathbf{c_2}(j\delta t) = \frac{1}{m}\sum_{j=1}^{n}\sum_{k=1}^{10}(\mathbf{C_2}^k(j\delta t) - \mathbf{C_1}^k(\frac{m}{n}j\delta t))^2 \quad (5)$$

The warped distance measure responds to changes in shapes of the traces over each motion cycle but does not change with the speed with which each cycle is executed. Thus, the distance between a fast walk and a slow walk would tend to be small as compared to the distance between a walk and a run cycle.

5. Human Motion Discrimination

We demonstrate the effectiveness of the presented ideas in the context of discriminating between walking, jogging, and running motion patterns. For our image sequence database, we imaged three persons performing the three types of motion on a treadmill using a Canon Optura digital video camera with progressive scan CCD. The frames were sampled at a rate of 30 frames per sec. Fig. 3 shows some typical frames. Due to temporal quantization noise, some of the image frames for running motion had motion blur. Also, the some frame to frame movement are over several pixels as is seen in the consecutive frames shown in Figs. 4 (a) and (b).

We compute the relational distribution for each image over only the edge pixels that are in motion, which are identified as follows. First, we apply the Canny edge detector over each image frame. Then, we compute the thresholded image difference between two consecutive frames, which identifies image parts that are possibly in motion. Fig. 4 (c) shows an example thresholded difference image. We consider only those Canny edge pixels that are in these motion regions (Fig. 4 (d)).

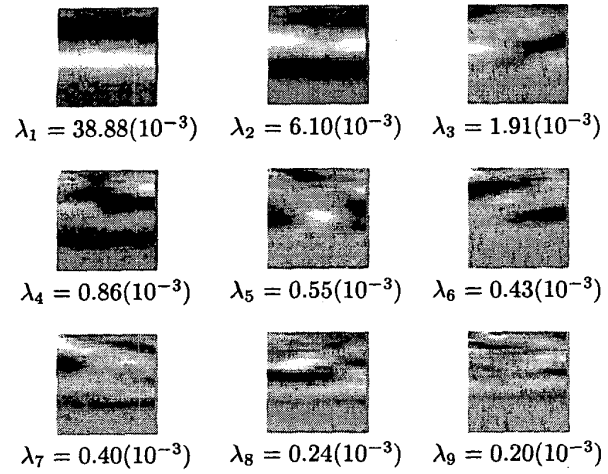

Figure 5. Ten most dominant dimensions of SoPF, with the corresponding eigenvalues quantifying the associated variation shown below each image. The vertical axis represents d/D and θ is along the horizontal. Brighter values indicate larger values.

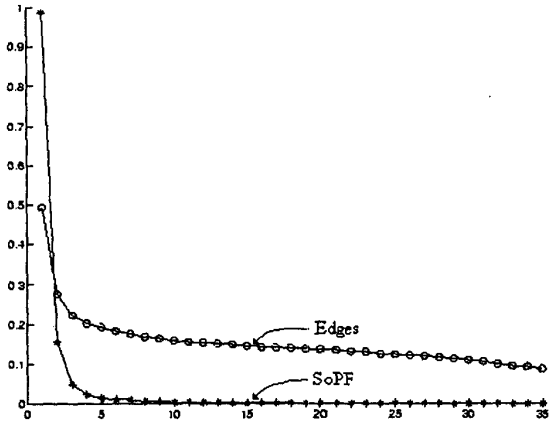

Figure 6. Comparison of the largest eigenvalues associated with the edge images of people in motion and those associated with the SoPF of 2-ary relational distributions of the same images. Values were energy normalized.

We used the 2-ary relational distributions, $P_2(d/D, \theta)$, to build the SoPF. One cycle of each motion type for each person is our training set. A total of 306 frames, was used to build the SoPF. The eigenvectors of the SoPF associated with the 10 largest eigenvalues are shown in Fig. 5 as gray level images. The vertical axes of the images plot the distance attribute, d/D, and the angle is along the horizontal. From the banded pattern in the two most dominant eigenvector we can see that they emphasize differences in the distance attribute between two features. Differences in orientation are emphasized by the other eigenvectors.

Fig. 6 shows the sorted eigenvalues for the 2-ary SoPFs as the '*'ed plot. Notice that most of the energy of the variation in the relational distribution is captured by the few large eigenvalues. For the results in this paper, we used the eigenvectors associated with the 10 largest eigenvalues, which as we see are sufficient.

Fig. 7 shows the variation of the coordinate, $c_0(t)$, associated with the largest eigenvalues of a training cycle for each of the three persons over each motion type. Each plot shows the variation over three motion cycles overlaid on each other. We can make the following observations from the figures. First, the differences in the nature of the variation for any person and a motion type over different cycles is small. Second, $c_0(t)$ captures mostly the periodic nature of the variation. Form of the variation between motion types and between persons is small. For the first and the third persons the second peak is smaller than the first one in the walking traces. Also, the amplitude of variation for the jogging motion of the third person is lower than the other two. Since the eigenvector associated with this coordinate emphasize the variation in distance between features, the amplitude of variation of $c_0(t)$ is related to the stride lengths. In other words, the third person's jogging stride is shorter than the other two. Another aspect worth pointing out is that the plots for running tend to be more "pointed" than for the other two motion types.

Fig. 8 plots the variation of the second coordinate $c_1(t)$, which has larger variation amongst different types of motions and persons. Differences in walking style of the second person from the first and the third show up in the nature of the variation. The running style of each person is different from the other two, as is also evident from the plots.

5.1. Distinguishing motion types

To test whether we can reliably distinguish between walking, jogging, and running across persons, we grouped the data into three classes, each representing each motion type and containing SoPF traces from all persons. We compute the warped distance (Eq. 5) between traces of each motion type (intra-class distance) and between traces of different motion types (inter-class distances). The first row of Table 1 list the means of the inter- and intra-class distances along with estimates of the variances of these estimates. The mean inter-class distance (30.52e-3) is almost double of the mean intra-class distance (15.82e-3).

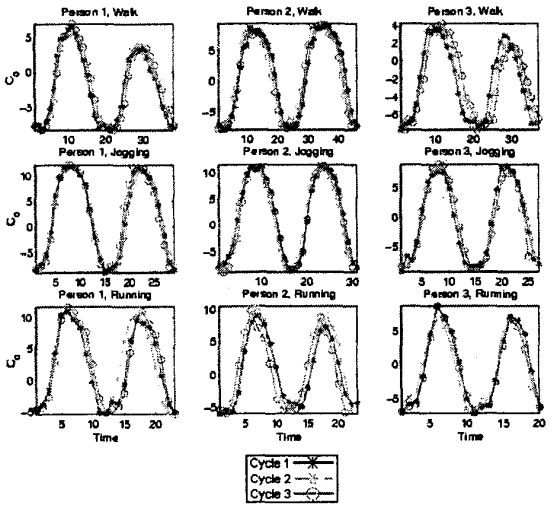

Figure 7. Variation of $c_0(t)$ within each motion cycle for each of the three persons and motion types.

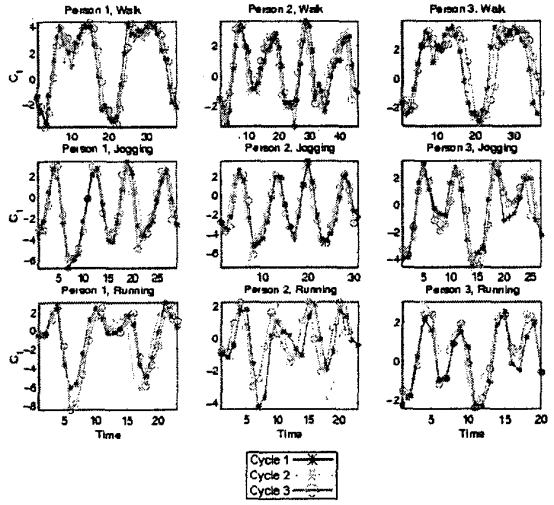

Figure 8. Variation of $c_1(t)$ within each motion cycle for each of the three persons and motion types.

Table 1. Distance between the traces through SoPF. The second and the fourth columns list the mean distances. The third and fifth columns list the variances in respective mean estimates.

Distinguishing	Trace Distance (10^{-3})			
	Intra-Class		Inter-Class	
	μ	σ_μ	μ	σ_μ
Motion Types	15.82	0.83	30.52	0.72
Motion Types				
of Person 1	2.42	0.39	41.38	2.75
of Person 2	3.76	0.69	19.80	0.75
of Person 3	4.78	1.46	15.45	0.70
Persons based				
on Walk	2.11	0.39	22.88	1.72
on Jogging	5.65	1.47	20.12	1.73
on Running	3.22	0.60	22.69	1.25

Table 2. Distance between the traces through SoPF for a sequence with moderate amount of segmentation noise.

Distinguishing	Trace Distance (10^{-3})			
	Intra-Class		Inter-Class	
	μ	σ_μ	μ	σ_μ
Motion Types	13.68	0.72	27.16	0.64
Motion Types				
of Person 1	2.12	0.36	37.32	2.46
of Person 2	3.34	0.62	18.05	0.75
of Person 3	4.23	1.31	14.30	0.58
Persons based				
on Walk	1.78	0.34	20.37	1.51
on Jogging	5.04	1.28	16.99	1.42
on Running	2.87	0.54	19.33	1.05

Table 3. Distance between the traces through SoPF for a sequence with large amount of segmentation noise.

Distinguishing	Trace Distance (10^{-3})			
	Intra-Class		Inter-Class	
	μ	σ_μ	μ	σ_μ
Motion Types	10.74	0.56	22.16	0.49
Motion Types				
of Person 1	1.79	0.30	30.33	1.97
of Person 2	2.91	0.53	15.17	0.64
of Person 3	3.81	1.06	12.68	0.44
Persons based				
on Walk	1.60	0.31	17.04	1.29
on Jogging	4.31	1.05	12.92	1.07
on Running	2.60	0.47	14.15	0.73

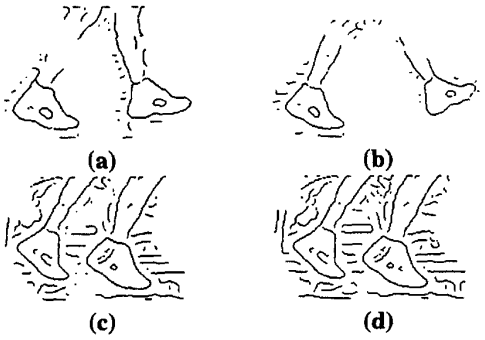

Figure 9. (a),(b): Some typical frames where the segmentation process misses significant portions of the legs. (c) An under segmented frame (corresponding to that in Fig. 4). (d) A more under segmented frame.

5.2. Motion discrimination on a per person basis

The SoPF traces can also be used to distinguish between motion types of a particular person. For each person we computed the distances between traces for the same motion type (intra-class distance) and that between traces for different motion types (inter-class distance). The second, third, and fourth rows of Table 1 list the mean distances for these two classes, along with estimates of their variances, for each of the three persons. We see that the mean inter-class distances are about 4 to 20 times larger than the mean intra-class distances.

5.3. Discriminating between persons

The next question we consider is the possibility of distinguishing persons based on SoPF traces of different gaits. To study this, for each motion type, we formed three classes of traces, one for each person. The inter- and intra-class mean distances between the traces over a cycle of motion is listed in the last three rows of Table 1. As we can see the inter-class mean distances are about 4 to 10 times larger than the intra-class mean distances. Thus, it is definitely possible to distinguish between persons based on SoPF traces.

5.4. Robustness with respect to segmentation errors

One of the attractive features of the presented approach is that it does not rely on perfect segmentation. Indeed, as outlined before, the segmentation process used is a rather crude one that identifies motion edges based on the image difference. The motion edges identified in such a manner contain a number of edge pixels from the background as was seen in Fig. 4(d). Sometimes, as shown in Fig. 9(a) and (b), even significant edges are missed if they are too close to the motion region boundary. Results presented so far were based on images that contained all these artifacts.

Table 4. Distance between the traces through SoPF of original size (O) and half scaled (HS) sequences.

	Trace Distance (10^{-3})		
Motion Cycles	O-Cycle 1	O-Cycle 2	O-Cycle 3
HS-Cycle 1	4.14	6.30	10.12
HS-Cycle 2	7.38	6.12	10.93
HS-Cycle 3	7.50	7.14	4.96

We also conducted a controlled study, where we relaxed our thresholds for identifying motion edges to include more edges. Fig. 9 (c) and (d) show motion edges identified for the frame shown in Fig. 4 for two different degrees of tolerances. More background edges are included in Fig. 9(d) than in Fig. 9(c), which is more that in Fig. 4(d). The various inter- and intra-class mean distances traces for the two noisy segmentations are listed in Tables 2 and 3. On comparing these distances with that listed in Table 1 we see that, although the gap between the inter- and intra-class means decrease with increasing segmentation noise, there is still enough discriminating power between the classes.

5.5. Robustness with respect to scale variations

To show that the traces in the SoPF are scale invariant we computed the distance between traces of three jogging cycles in their original sizes and their half scaled versions, all from the same person. The results are shown in Table 4, we can see the distances between traces are smaller that the mean intra-class distance shown in Table 1.

5.6. Higher order relational distributions

The results presented so far are using 2-ary relational distributions. One might ask if higher order relational distributions, such as 3-ary ones among three edge pixels, result in better discriminating power? Our studies, which due to limited space we cannot present here, indicate that although the difference between the inter- and intra-class mean distance reduces for 3-ary relations, the intra- and the inter-class distance variations are less than that for 2-ary relations. Thus, 3-ary relations should offer better discriminating power than 2-ary relations, but the probability space for 3-ary relations is much larger than for 2-ary and hence as a consequence is slower to compute. Complexity to compute 3-ary relation distributions is $\mathcal{O}(N^3)$, where N is the number of edge pixels, whereas the computational complexity for 2-ary distributions is $\mathcal{O}(N^2)$.

5.7. A few words about space and time

On first reading, it might appear that the space complexity of the relational distributions would be enormous and computing them would be time consuming. It is true

that the representational complexity increases with the N-aryness of the relations. However, we found that 2-ary and 3-ary relations suffice. To represent the 2-ary relations we used a 30x30 matrix which is much smaller than the images, which were 256x130 each. The 3-ary relations were represented by a 30x30x30x30 matrix. However, since the edges were structured, most of the 3-ary relation space was empty, hence, sparse matrix representations would be well suited, which we plan to explore. As far as time taken to compute the relational distributions are concerned, it took 6.9 seconds per image for the 2-ary relations and 38.24 min per image for the 3-ary relations by exhaustively enumerating all possible 3 combinations of edges on a Sun Ultra 30 Creator workstation running at 246 MHz with 500 Mb of RAM. This exhaustive enumeration, during run time, can be replaced by a stochastic sampling process, possibly by a Markov Chain Monte Carlo one, to estimate the coordinates in the SoPF space directly. This is presently under consideration.

6. Why not just PCA of the Edge images?

One might ask, why not just do a PCA of the edge images instead of the relational distribution of the edges? Our experience shows that the SoPF representation is more compact than the PCA space of the raw edges themselves. Besides, the computational complexity of the edge PCA is dependent of the image size used, hence the scale of the images, whereas the relational distributions have fixed sizes. In fact, since we have 265x130 images, we found it difficult to allocate enough memory to compute the eigenvalues and eigenvectors directly from these images using a Sun Ultra 30 Creator workstation running at 246 MHz with 256 Mb of RAM. So we then reduced the image size by half and it took almost 34 hrs to calculate the eigenspace. In contrast, the size of the relational distribution we used was 30x30, which is easier to compute. Fig. 6 shows the plot of the eigenvalues for the edge PCA as the 'o'ed plot. From this plot it is obvious that the Edge PCA space is a lot less compact than the SoPF space and that we need less number of dimensions.

7. Conclusions

In this paper we presented a statistical framework for motion analysis that tracks the variation of non-stationarity in the distributions of relations among image features in individual frames. We proposed the concept of a Space Of Probability Functions (SoPF) that allows us to capture the non-stationary variations. Among the attractive features of this approach are (i) no part level tracking or feature correspondence is necessary, (ii) motion segmentation of object need not be perfect, (iii) there is no need for object models, and (iv) single pixel movement between frames is not assumed. We presented results on human motion sequences of 3 persons with 3 kinds of motions, walking, jogging, and running. We realize that to make a definitive statement about performance we need to consider a larger image sequence database, but based on the testing that we have done we can claim that the method presented here is exciting.

References

[1] S. Belongie and J. Malik. Matching with shape contexts. In *Workshop on Content-Based Access of Image and Video Libraries*, pages 20–26, 2000.

[2] M. Black and A. Jepson. EigenTracking:robust matching and tracking of articulated objects using view-based representation. In *European Conference on Computer Vision*, pages 329–342, 1996.

[3] M. Black, Y. Yacoob, and S. Ju. Recognizing human motion using parameterized models of optical flow. In *Motion-Based Recognition*, page Ch. 11, 1997.

[4] A. Bobick and A. Wilson. A state based approach to the representation and recognition of gesture. *IEEE Trans. Pattern Anal. and Mach. Intel.*, 19(12):1325–1337, December 1997.

[5] R. Collins, A. Lipton, and T. Kanade. Introduction to the special section on video surveillance. *IEEE Trans. Pattern Anal. and Mach. Intel.*, 22(8):745–746, Aug. 2000.

[6] R. Cutler and L. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. Pattern Anal. and Mach. Intel.*, 22(8):781–796, Aug. 2000.

[7] N. Goddard. Human activity recognition. In *Motion-Based Recognition*, page Ch. 7, 1997.

[8] J. Little and J. Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2):xx–yy, 1998.

[9] D. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

[10] B. Olshausen and D. Field. Natural image statistics and efficient coding. *Network Computation in Neural Systems*, 7(2):333–339, 1996.

[11] R. Polana and R. Nelson. Detection and recognition of periodic, nonrigid motion. *International Journal of Computer Vision*, 23(3):261–282, 1997.

[12] R. Polana and R. Nelson. Temporal texture and activity recognition. In *Motion-Based Recognition*, page Ch. 5, 1997.

[13] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, Jan. 2000.

[14] S. Seitz and C. Dyer. Cyclic motion analysis using periodic trace. In *Motion-Based Recognition*, page Ch. 4, 1997.

[15] A. Selinger and L. Wixson. Classifying moving objects as rigid or non-rigid without correspondences. In *DARPA*, pages 341–347, 1998.

[16] M. Shah and R. Jain (Eds.). *Motion-Based Recognition*. Kluwer Publishers, 1997.

[17] S. Zhu. Embedding gestalt laws in markov random fields. *IEEE Trans. Pattern Anal. and Mach. Intel.*, 21(11):1170–1187, November 1999.