Matching Shape Sequences in Video with Applications in Human Movement Analysis.

Ashok Veeraraghavan, *Student Member, IEEE,* Amit Roy Chowdhury, *Member, IEEE,* and Rama Chellappa, *Fellow, IEEE*

Abstract

We present an approach for comparing two sequences of deforming shapes using both parametric models and non-parametric methods. In our approach, Kendall's definition of shape is used for feature extraction. Since the shape feature lives on a non-Euclidean manifold, we propose parametric models like the autoregressive model and autoregressive moving average model on the tangent space and demonstrate the ability of these models to capture the nature of shape deformations using experiments on gait-based human recognition. The non-parametric model is based on Dynamic Time-Warping. We suggest a modification of the Dynamic time-warping algorithm to include the nature of the non-Euclidean space in which the shape deformations take place. We also show the efficacy of this algorithm by its application to gait-based human recognition. We consider the shape deformations of a person's silhouette as a discriminating feature and provide recognition results using the non-parametric model. Our analysis leads to some interesting observations on the role of shape and kinematics in automated gait-based person authentication.

Index Terms

Shape, Shape Sequences, Shape Dynamics, Comparison of shape sequences, Gait recognition.

This work was supported by the NSF-ITR Grant 0325119.

Ashok Veeraraghavan and Rama Chellappa are with the University of Maryland, College Park. Amit Roy Chowdhury is with the University of California, Riverside

I. Introduction

Shape analysis plays a very important role in object recognition, matching and registration. There has been substantial work in shape representation and on defining a feature vector which captures the essential attributes of the shape. A description of shape must be invariant to location, scale and rotation. Several features describing a shape have been developed in the literature that provide for all or some of the above mentioned invariants and are very robust to errors in the silhouette extraction process. Most of these methods compare individual shapes on one or two frames. But, there has so far been very little work on attempting to capture the dynamics in this shape feature, as is available in a video and use this either directly for object recognition or for activity classification.

In typical video processing tasks the input is a video of an object or a set of objects that deform or change their relative poses. The essential information conveyed by the video can be usually captured by analyzing the boundary of each object as it changes with time. In this paper, we consider scenarios where the time variation of the shape of an object provides cues about the objects identity and/or the activity performed by the object and sometimes even about the nature of the interaction between different objects in the same scene. We describe both parametric and non-parametric methods to compute meaningful distance measures between two such sequences of deforming shapes. We illustrate our approach using gait analysis. We treat the silhouette of the individual during walking as a time sequence of deforming shapes. The methods provided are generic and can be used to characterize the time evolution of any set of landmark points, not necessarily on the silhouette of the object.

We begin by providing a brief literature review of the research in shape analysis. The interested reader may refer to comprehensive surveys of the field [1], [2]. Since the experimental results are for the problem of gait recognition we also provide a brief summary of prior work in gait-based person authentication. Special emphasis is given to understanding the role of shape and kinematics in gait recognition since our experiments lead to interesting observations on this issue.

A. Previous Work in Shape Analysis

Pavlidis [3] categorized shape descriptors into various taxonomies according to different criteria. Descriptors that use the points on the boundary of the shape are called external (or boundary)[4][5][6] while those that describe the interior of the object are called internal (or global)[7][8]. Descriptors that represent shape as a scalar or as a feature vector are called numeric while those like the medial axis transform that describes the shape as another image are called non-numeric descriptors. Descriptors are also classified as information preserving or not based on whether the descriptor allows accurate reconstruction of a shape.

1) Global Methods for shape matching: Global shape matching procedures treat the object as a whole and describe it using some features extracted from the object. The disadvantage of these methods is that it assumes that the image given must be segmented into various objects which by itself is not an easy problem. In general, these methods cannot handle occlusion and are not very robust to noise in the segmentation process. Popular moment based descriptors of the object such as [8],[10],[11] are global and numeric descriptors. Goshtasby [12] used the pixel values corresponding to polar coordinates centered around the center of mass of the shape, the shape matrix, as a description of the shape. Parui et. al. [13] used relative areas occupied by the object in concentric rings around the objects centroid as a description of the shape. Blum and Nagel [7] used the medial axis transform to represent the shape.

2) Boundary methods for shape matching: Shape matching methods based on the boundary of the object or on a set of pre-defined landmarks on the object have the advantage that they can be represented using a one dimensional function. In the early sixties, Freeman [14] used chain coding (a method for coding line drawings) for the description of shapes. Arkin et al. [15] used the turning function for comparing polygonal shapes. Persoon and Fu [5] described the boundary as a complex function of the arc length. The problem with a Fourier representation [5] is that the local information is lost in this method. Kashyap and Chellappa [4] used a circular autoregressive model of the distance from the centroid to the boundary to describe the shape. Srivastava et al. [16] propose differential geometric representations of continuous planar shapes.

Recently several authors have described shape as a set of finite ordered landmarks. Kendall [9] provided a mathematical theory for the description of landmark based shapes. Bookstein [17] and later Dryden and Mardia [18] have furthered the understanding of such landmark based shape descriptions. There has been a lot of work on planar shapes [19] and [20]. Prentice and Mardia [19] provided a statistical analysis of shapes formed by matched pairs of landmarks on the plane. They provided inference procedures on the complex plane and a measure of shape change in the plane. Berthilsson [21] and Dryden [22] describe a statistical theory for shape spaces. Projective shape and their respective invariants are discussed in [21] while shape models, metrics and their role in high level vision is discussed in [22]. The shape context [6] of a particular point in a point set captures the distribution of the other points with respect to it. [6] uses the shape context for the problem of object recognition. The softassign Procrustes matching algorithm [23] simultaneously establishes correspondences and determines the Procrustes fit.

3) Dynamics of shapes: The recent explosion in the area of shape discrimination and shape retrieval can be attributed to their enormous potential in the problems of object recognition and shape based image retrieval. Inspite of these recent developments there has been very few studies on the variation of object shape as a cue for object recognition and activity classification. Yezzi and Soatto [24] separate the overall motion from deformation in a sequence of shapes. They use the notion of shape average to differentiate global motion of a shape from the deformations of a shape. Vaswani et al. [25] used the dynamics of a configuration of interacting objects to perform activity classification. They apply the learned dynamics for the problem of detecting abnormal activities in a surveillance scenario. Recently, Liu and Ahuja [26] have proposed using autoregressive models on the Fourier descriptors for learning the dynamics of a 'dynamic shape'. They use this model for performing object recognition, synthesis and prediction. Mowbray and Nixon [27] use spatio-temporal Fourier descriptors to model the shape descriptions of temporally deforming objects and perform gait recognition experiments using their shape descriptor. In this paper, we provide a mathematical framework for comparing two sequences of shapes with applications in gait-based human identification and activity recognition.

B. Prior Work in Gait recognition

The study of human gait has recently been driven by its potential use as a biometric for person identification. We outline some of the methods in gait-based human identification.

1) Shape based methods: Niyogi and Adelson [28] obtained spatio-temporal solids by aligning consecutive images and use a weighted Euclidean distance to obtain recognition. Phillips et al. [29] provide a baseline algorithm for gait recognition using silhouette correlation. Han and Bhanu [30] use the gait energy image while Wang et al. use Procrustes shape analysis for recognition [31]. Foster et al. [32] use area based features. Bobick and Johnson [33] use activity specific static and stride parameters to perform recognition. Collins et al. build a silhouette based nearest neighbor classifier [34] to do recognition. Several researchers [35][36] have used Hidden Markov Models (HMM) for the task of gait based identification. Another shape based method for identifying individuals from noisy silhouettes is provided in [37].

2) Kinematics based methods: Apart from these image based approaches Cunado et al. [38] model the movement of thighs as articulated pendulums and extract a gait signature. But in such an approach robust estimation of thigh position from a video can be very difficult. In another kinematics based approach [39], trajectories of the various parameters of a kinematic model of the human body are used to learn a dynamical system. Tanawongsuwan and Bobick [48] have developed a normalization prcedure that maps gait features across different speeds in order to compensate for the inherent changes in gait features associated with the speed of walking. All the above methods have both static (shape) aspects and dynamic features used for gait recognition. Yet the relative importance of shape and dynamics in human motion has not been investigated. The experimental results of this work shed some light on this issue.

3) Prior work on role of shape and kinematics in human gait: Johansson [40] attached light displays to various body parts and showed that humans can identify motion with the pattern generated by a set of moving dots. Since Muybridge [41] captured photographic recordings of human and animal locomotion, considerable effort has been made in the Computer vision, Artificial Intelligence and Image Processing communities to the understanding of human activities from videos. A survey of work in human motion analysis can be found in [42].

Several studies have been done on the various cues that humans use for gait recognition. Hoenkamp[43] studied the various perceptual factors that contribute to the labeling of human gait. Medical studies[44] suggest that there are 24 different components to human gait. If all these different components are considered then it is claimed that the gait signature is unique. Since it is very difficult to extract these components reliably several other representations have been used. It has been shown [45] that humans can do gait recognition even in the absence of familiarity cues. Cutting and Kozlowski also suggest that dynamic cues like speed, bounciness and rhythm are more important for human recognition than static cues like height. Cutting and Proffitt [46] argue that motion is not the simple compilation of static forms and claim that it is a dynamic invariant that determines event perception. Moreover, they also found that dynamics was crucial to gender discrimination using gait. Therefore, it is intuitive to expect that dynamics also plays a role in person identification though shape information might also be equally important. Interestingly, Veres et al. [47] recently did a statistical analysis of the image information that is important in gait recognition and concluded that static information is more relevant than dynamical information. In the light of such developments, our experiments explore the importance of shape and dynamics in human movement analysis from the perspective of computer vision and analyze their role in existing gait recognition methodologies.

This paper is concerned with situations where the manner of shape change of an object provides clues about its identity and/or about the nature of the activity performed by the object. In such scenarios we need to be able to compute distances and compare two sequences of deforming shapes by considering the entire sequence as one entity instead of performing a frame-wise shape comparison. In this paper, we present methods for computing distances between such sequences of deforming shapes. The non-parametric method for comparing two shape sequences is an extension of the Dynamic time warping algorithm[49], initially used in the speech recognition literature. We propose a modification of the algorithm to account for the non-Euclidean nature of the shape-space. We also propose parametric models for learning the dynamics of the deformations of shape sequences. We can then compute distances between learned models in the appropriate parametric space in order to compute distances between shape sequences.

We suggest new gait recognition algorithms by computing the distances between two shape sequences. A sequence of a walking person is represented as a sequence of shapes and the distance between shape sequences is used to perform gait recognition. Experiments on gait recognition were performed 1) to show the efficacy of our shape sequence matching algorithms and 2) to learn the importance of the role of shape and kinematics in automatic gait recognition.

Section II provides a brief introduction to Kendall's landmark based shape descriptor used as a shape feature. In Section III and IV, we discuss our parametric and non-parametric methods for comparing shape sequences. In the experimental section V, we show the efficacy of our algorithms by providing recognition results using standard gait recognition databases. Finally section VI deals with conclusions and future work.

II. Kendall's Shape Theory - Preliminaries

A. *Definition of Shape*

"Shape is all the geometric information that remains when location, scale and rotational effects are filtered out from the object"[18]. We use Kendall's statistical shape as the shape feature in this paper. [18] provides a description of the various tools in statistical shape analysis. Kendall's representation of shape describes the shape configuration of k landmark points in an m-dimensional space as a $k \times m$ matrix containing the coordinates of the landmarks. In our analysis we have a 2 dimensional space and therefore it is convenient to describe the shape vector as a k dimensional complex vector.

The binarized silhouette denoting the extent of the object in an image is obtained. A shape feature is extracted from this binarized silhouette. This feature vector must be invariant to translation and scaling since the objects identity should not depend on the distance of the object from the camera. So any feature vector that we obtain must be invariant to translation and scale.

This yields the pre-shape of the object in each frame. Pre-shape is the geometric information that remains when location and scale effects are filtered out. Let the configuration of a set of k landmark points be given by a k-dimensional complex vector containing the position of the landmarks. Let us denote this configuration as X. Centered pre-shape is obtained by subtracting the mean from the configuration and then scaling to norm one. The centered pre-shape is given by

$$Z_c = \frac{CX}{\parallel CX \parallel}, \quad where \quad C = I_k - \frac{1}{k} 1_k 1_k^T, \tag{1}$$

where I_k is a k×k identity matrix and 1_k is a k dimensional vector of ones.

B. Distance between shapes

The pre-shape vector that is extracted by the method described above lies on a spherical manifold. Therefore a concept of distance between two shapes must include the non-Euclidean nature of the shape space. Several distance metrics have been defined in [18]. Consider two complex configurations X and Y with corresponding corresponding preshapes α and β. The full Procrustes distance between the configurations X and Y is defined as the Euclidean distance between the full Procrustes fit of α and β. Full Procrustes fit is chosen so as to minimize

$$d(Y, X) = \parallel \beta - \alpha s e^{j\theta} - (a + jb) 1_k \parallel, \tag{2}$$

where s is a scale, θ is the rotation and $(a + jb)$ is the translation. Full Procrustes distance is the minimum Full Procrustes fit i.e.,

$$d_F(Y, X) = \inf_{s,\theta,a,b} d(Y, X). \tag{3}$$

We note that the preshapes are actually obtained after filtering out effects of translation and scale. Hence, the translation value that minimizes the full Procrustes fit is given by $(a + jb) = 0$, while the scale $s = |\alpha^* \beta|$ is very close to unity. The rotation angle θ that minimizes the Full Procrustes fit is given by $\theta = arg(|\alpha^* \beta|)$.

The partial Procrustes distance between configurations X and Y is obtained by matching their respective preshapes α and β as closely as possible over rotations, but not scale. So,

$$d_P(X,Y) = \inf_{\Gamma \epsilon SO(m)} \parallel \beta - \alpha\Gamma \parallel . \tag{4}$$

It is interesting to note that the optimal rotation θ is the same whether we compute the full Procrustes distance or the partial Procrustes distance. The Procrustes distance $\rho(X,Y)$ is the closest great circle distance between α and β on the preshape sphere. The minimization is done over all rotations. Thus ρ is the smallest angle between complex vectors α and β over rotations of α and β. The three distance measures defined above are all trigonometrically related as

$$d_F(X,Y) = \sin\rho, \tag{5}$$

$$d_P(X,Y) = 2\sin(\frac{\rho}{2}). \tag{6}$$

When the shapes are very close to each other there is very little difference between the various shape distances. In our work we have used the various shape distances to compare the similarity of two shape sequences and obtain recognition results using these similarity scores. Our experiments show that the choice of shape-distance does not alter recognition performance significantly for the problem of gait recognition since the shapes of a single individual lie very close to each other. We show the results corresponding to the partial Procrustes distance in all our plots in this paper.

C. The tangent space

The shape tangent space is a linearization of the spherical shape space around a particular pole. Usually the Procrustes mean shape of a set of similar shapes(Y_i) is chosen as the pole for the tangent space coordinates. The Procrustes mean shape(μ) is obtained by minimizing the sum of squares of full Procrustes distances from each shape Y_i to the mean shape, i.e.,

$$\mu = \arg \inf_{\mu} \quad \Sigma d_F^2(Y_i,\mu). \tag{7}$$

The pre-shape formed by k points lie on a $k-1$ dimensional complex hypersphere of unit radius. If the various shapes in the data are close to each other then these points on the hypersphere will also lie close to each other. The Procrustes mean of this dataset will also lie close to these points. Therefore the tangent space constructed with the Procrustes mean shape as the pole is an approximate linear space for this data. The Euclidean distance in this tangent space is a good approximation to the various Procrustes distances d_F, d_P and ρ in shape space in the vicinity of the pole. The advantage of the tangent space is that this space is Euclidean.

The Procrustes tangent coordinates of a preshape α is given by

$$v(\alpha, \mu) = \alpha\alpha^*\mu - \mu|\alpha^*\mu|^2. \tag{8}$$

where μ is the Procrustes mean shape of the data.

III. Motivation for Shape Sequence Processing

There are several situations where we are interested in studying the way in which the shape of an object changes with time. The manner in which this shape change occurs provides clues about the nature of the object and sometimes even about the activity performed by the object. In [24] this shape change is considered to be a result of global motion and shape deformation. They separate the global motion by introducing a notion of temporal shape average and study the nature of both global motion of a shape and deformations. In [25] the manner of this shape change is captured parametrically using their tangent space projections. They also had an overview of how to model non-stationary shape sequences, but assumed stationarity in their examples. In this section we describe the motivation for our formulation and the scenarios that we are interested in tackling.

Consider the manner in which the shape of the lip changes when we speak. The manner in which the shape of the lip changes during speech provides significant information about the actual words that are being spoken. Consider the two words ‘arrange’ and ‘ranger’. If we take discrete snapshots of the shape of the lip during each of these words we see that the two sets of snapshots will be identical(or almost identical) though the ordering of the discrete snapshots

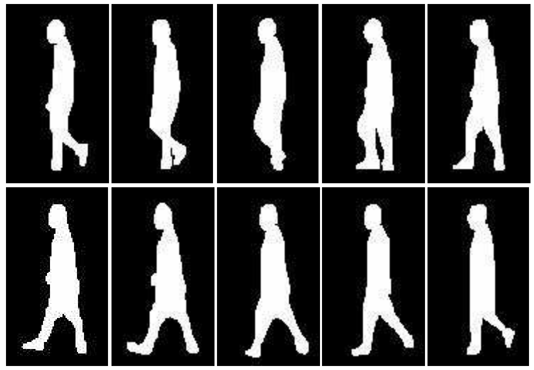


Fig. 1. Sequence of shapes as a person walks frontoparallely

will be very different for these two utterances. Therefore any method that inherently does not learn/use the dynamics information of this shape change will declare that these two utterances are very close to each other while in reality these are very different words. Therefore, in cases such as this, where shape change is critical to recognition, it is important to consider the entire shape sequence, i.e., the shape sequence is more important than the individual shapes at discrete time instants. There are many such cases where the nature of shape changes of silhouette of a human provides information about the activity performed by the human. Consider the images shown in Fig:1. It is not very difficult to perceive the fact that these represent the silhouette of a human while he is walking. These and many other examples can be thought of, where the shape change captured in the shape sequence provides information about the activity being performed.

Apart from providing information about the activity being performed, there are also several instances when the manner of shape changes provides valuable insights regarding the identity of the object. Even though the outline of the shape of both a lion and a cheetah are very similar (with four legs etc) especially in its profile view, the manner in which a lion and a cheetah move are so drastically different. The discrimination between two such classes is significantly improved if we take the manner of shape changes into account. Thus there are several situations where it is important to be able to learn the dynamics of shape changes or at the least to be able to compute meaningful distances between such shape sequences. Here, we present some parametric and non-parametric methods for tackling stationary shape sequences.

IV. Comparison of Shape Sequences

In this section we provide a method based on dynamic time warping to compute distances between shape sequences. We also provide methods based on autoregressive and autoregressive moving average models to learn the dynamics of these shape changes and use the distance measures between models as a measure of similarity between these shape sequences. The methods described here can be used generically for any landmark based description of shapes, not just to silhouettes.

A. Non-Parametric method for comparing shape sequences

Consider a situation where there are two shape sequences and we wish to compare how similar these two shape sequences are. We may not have any other specific information about these sequences and therefore any attempt at modeling these sequences is difficult. These shape sequences may be of differing length(number of frames) and therefore in order to compare these sequences we need to perform time normalization(scaling). A linear time scaling would be inappropriate because in most scenarios this time scaling would be inherently non-linear. Dynamic time warping(DTW) which has been successfully used by the speech recognition [49] community is an ideal candidate for performing this non-linear time normalization. However, certain modifications to the original DTW are also necessary in order to account for the non-Euclidean structure of the shape space.

1) Dynamic time warping: Dynamic time warping (DTW) is a method for computing a non-linear time normalization between a template vector sequence and a test vector sequence. These two sequences could be of differing lengths. The algorithm which is based on dynamic programming computes the best non-linear time normalization of the test sequence in order to match the template sequence, by performing a search over the space of all allowed time normalizations. The space of all time normalizations allowed is cleverly constructed using certain temporal consistency constraints. We list the temporal consistency constraints that we have used in our implementation of the DTW below.

- End point constraints: The beginning and the end of each sequence is rigidly fixed. For example if the template sequence is of length N and the test sequence is of length M then only time normalizations that map the first frame of the template to the first frame of the test sequence and also map the Nth frame of the template sequence to the Mth frame of the test sequence are allowed.
- The warping function(mapping function between the test sequence time to the template sequence time) should be monotonically increasing. In other words the sequence of 'events' in both the template and the test sequences should be the same.
- The warping function should be continuous.

Dynamic programming is used to efficiently compute the best warping function and the global warping error.

Pre-shape, as we have already discussed lies on a spherical manifold. The spherical nature of the shape-space must be taken into account in the implementation of the DTW algorithm. This implies that during the DTW computation the local distance measure used must take into account the non-Euclidean nature of the shape-space. Therefore, it is only meaningful to use the Procrustes shape distances described earlier. It is important to note that the Procrustes distance is not a distance metric since it is not commutative. Moreover, the nature of the definition of constraints make the DTW algorithm non-commutative even when we use a distance metric for the local feature error. If $A(t)$ and $B(t)$ are two shape sequences then, we define the distance between these two sequences $D(A(t), B(t))$ as

$$D(A(t), B(t)) = DTW(A(t), B(t)) + DTW(B(t), A(t)); \tag{9}$$

Such a distance between shape sequences is commutative. The isolation property i.e., $D(A(t), B(t)) = 0$ iff $A(t) = B(t)$, is enforced by penalizing all non-diagonal transitions in the local error metric.

B. Parametric models for shape sequences

In several situations, it is very useful to model the shape deformations over time. If such a model could be learned either from the data or from the physics of the actual scenario, then it would help significantly in problems such as identification and for synthesizing shape sequences. [26] learn the nature of shape changes of a fire sequence. They also synthesize new sequences of fire using the model that they learned. This section describes work with very similar objectives. We describe both autoregressive(AR) and autoregressive and moving average(ARMA) models on tangent space projections of the shape. We describe methods to learn these models from sequences and compute distances between models in this parametric setting. Our approach for parametric modeling differs from that of [26] in two important ways. The shape feature on which we build parametric models preserves locality while the Fourier descriptors that they use is a global shape feature. Therefore our method can in principle capture the dynamics of shape sequences locally and is better suited for applications where different local neighborhoods of the shape exhibit different dynamics. We use parametric modeling for modeling human gait, a very specific example where different local neighborhoods(different parts of the body) exhibit different dynamics. Moreover, we also extend the parametric modeling from AR to the ARMA model. The advantage of the ARMA model is that it can be used to characterize systems with both poles and zeros while the AR model can be used to characterize systems with zeros only.

1) AR Model on tangent space: The AR model is a simple time-series model that has been used very successfully for prediction and modeling especially in speech. The probabilistic interpretation of the AR model is valid only when the space is Euclidean. Therefore, we build an AR model on the tangent space projections of the shape sequence. Once the AR model is learned we can use this either for synthesis of a new shape sequence or for comparing shape sequences by computing distances between the model parameters.

The time series of the tangent space projections of the pre-shape vector of each shape is modeled as an AR process. Let, s_j, $j = 1, 2, ...M$ be the M such sequences of shapes. Let us denote the tangent space projection of the sequence of shape s_j (with mean of s_j as the pole) by α_j. Now, the AR model on the tangent space projections is given by,

$$\underline{\alpha_j}(t) = A_j \underline{\alpha_j}(t-1) + w(t) \tag{10}$$

where, w is a zero mean white Gaussian noise process and A_j is the transition matrix corresponding to the j^th sequence. For convenience and simplicity A_j is assumed to be a diagonal matrix.

For all the sequences in the gallery, the transition matrices are obtained and stored. Given a probe sequence, the transition matrix for the probe sequence is computed. The distances between the corresponding transition matrices are added to obtain a measure of the distance between the models. If A and B (for $j = 1, 2, ...N$) represent the transition matrices for the two sequences, then the distance between the models is defined as $D(A, B)$

$$D(A, B) = ||A_j - B_j||_F, \tag{11}$$

where $||.||_F$ denotes the Frobenius norm. The model in the gallery that is closest to the model of the given probe is chosen as the correct identity.

2) ARMA Model: We pose the problem of learning the nature of a shape sequence as one of learning a dynamical model from shape observations. We also regard the problem of shape sequence based recognition as one of computing the distances between the dynamical models thus learned. The dynamical model is a continuous state, discrete time model. Since the parameters of the models lie in a non-Euclidean space, the distance computations between the models is non-trivial. Let us assume that the time-series of tangent projections of shapes(about its mean as the pole) is given by $\alpha(t), t = 1, 2, , , , , \tau$. Then an ARMA model is defined as [50] [39]

$$\alpha(t) = Cx(t) + w(t); w(t) \sim N(0, R) \tag{12}$$

$$x(t+1) = Ax(t) + v(t); v(t) \sim N(0, Q). \tag{13}$$

Also, let the cross correlation between w and v be given by S. The parameters of the model are given by the transition matrix A and the state matrix C. We note that the choice of matrices A, C, R, Q, S is not unique. However, we can transform this model to the "innovation representation"[51] which is unique.

3) Learning the ARMA model: We use the tools from the system identification literature to estimate the model parameters.The estimation is closed form and therefore simple to implement. The algorithm is described in [51] and [52]. Given observations $\alpha(1), \alpha(2), ...\alpha(\tau)$, we have to learn the parameters of the innovation representation given by $\hat{A}, \hat{C}$ and $\hat{K}$($\hat{K}$:Kalman gain matrix of the innovation representation[51]). Note that in the innovation representation, the state covariance matrix $\lim_{t \to \infty} E[x(t)x^T(t)]$ is asymptotically diagonal. Let $[\alpha(1)\alpha(2)\alpha(3).....\alpha(\tau)]$ = $U \Sigma V^T$ be the singular value decomposition of the data. Then

$$\hat{C}(\tau) = U \tag{14}$$

$$\hat{A} = \Sigma V^T D_1 V (V^T D_2 V)^{-1} \Sigma^{-1} \tag{15}$$

where D_1 = [0 0;$I_{\tau-1}$ 0] and D_2 = [$I_{\tau-1}$ 0;0 0].

4) Distance between ARMA Models: Subspace angles [53] between two ARMA models are defined as the principal angles ($\theta_i, i = 1, 2, ...n$) between the column spaces generated by the observability spaces of the two models extended with the observability matrices of the inverse models [54]. The subspace angles between two ARMA models ($[A_1, C_1, K_1]$ and $[A_2, C_2, K_2]$ can be computed by the method described in [54]. Using these subspace angles $\theta_i, i = 1, 2, ...n$, three distances, Martin distance(d_M), gap distance(d_g) and Frobenius distance(d_F) between the ARMA models are defined as follows:

$$d_M^2 = \ln \prod_{i=1}^{n} \frac{1}{\cos^2(\theta_i)}, \tag{16}$$

$$d_g = \sin \theta_{max}, \tag{17}$$

$$d_F^2 = 2 \sum_{i=1}^{n} \sin^2 \theta_i. \tag{18}$$

The various distance measures do not alter the results significantly. We show the results using the Frobenius distance(d_F^2).

V. Experiments on Gait Recognition

We describe the various experiments we designed using the algorithms previously discussed in order to study gait-based human recognition. We also show an extension of the same analysis for the problem of activity recognition. The goals of the experiments were:

1) Show the efficacy of our algorithms in comparing shape sequences by applying it to the problem of automated gait recognition.
2) Study the role of shape and kinematics in automated gait recognition algorithms.
3) Make a similar study on the role of shape and kinematics for activity recognition.

Continuing our approach in [55] we use a purely shape based technique called the Stance Correlation to study the role of shape in automated gait recognition.

The algorithms for comparing shape sequences were applied on two standard databases. The USF database [29] consists of 71 people in the Gallery[1]. Various covariates like camera position, shoe type, surface and time were varied in a controlled manner to design a set of challenge experiments[2][29]. The results are evaluated using cumulative match scores[3](CMS) curves and the identification rate. The CMU database [34] consists of 25 people. Each of the 25 people perform four different activities(slow walk, fast walk, walking on an inclined surface and walking with a ball). For the CMU database we provide results for recognition both within an activity and across activities. We also provide some results on activity recognition on this dataset. Apart from these, we also provide activity recognition results on the MOCAP dataset(available from Credo Interactive Inc. and Carnegie Mellon University) which consists of different examples of various activities.

[1]A more expanded version is available on which we haven't yet experimented.However we do not expect our conclusions to alter significantly.

[2]Challenge Experiments:Probes A-G in increasing order of difficulty.

[3]Plot of percentage of recognition Vs rank.

A. *Feature Extraction*

Given a binary image consisting of the silhouette of a person, we need to extract the shape from this binary image. This can be done either by uniform sampling along each row or by uniform arc-length sampling. In uniform sampling, landmark points are obtained by identifying the edges of the silhouette in each row of the image. In uniform arc length sampling, the silhouette is initially interpolated using critical landmark points. Uniform sampling on this interpolated silhouette provides us with the uniform arc-length sampling landmarks. Once the landmarks are obtained, the shape is extracted using the procedure described in 2.1. The procedure for obtaining shapes from the video sequence is graphically illustrated in Figure 2. Note that each frame of the video sequence maps to a point on the spherical(hyper-spherical) shape manifold.

B. *Experiments on Gait Recognition*

1) Results on the USF Database: On the USF database we conducted experiments on recognition performance using these methods- Stance Correlation, DTW on shape space, Stance based AR(a slight modification of the AR model [55]) and ARMA model. Gait recognition experiments were designed for challenge experiments A-G. These experiments featured and tested the recognition performance against various covariates like camera angle, shoe type, surface change etc. Refer to [29] for a detailed description of the various experiments and the covariates in these experiments. Figure 3 shows the CMS curves for the challenge experiments A-G using the Dynamic time warping and the ARMA model. The recognition performance of the DTW based method is comparable to the state of art algorithms that have been tested on this data[35]. The performance of the ARMA model is lower since human gait is a very complex action and the ARMA model is unable to capture all these details.

In order to understand the significance of shape and kinematics in gait recognition, we ran the same experiments with other purely shape and purely dynamics based methods as described in [55]. Figure 4 shows the average CMS curves(average of the 7 Challenge experiments:Probes A-G) for the various shape and kinematics based methods.

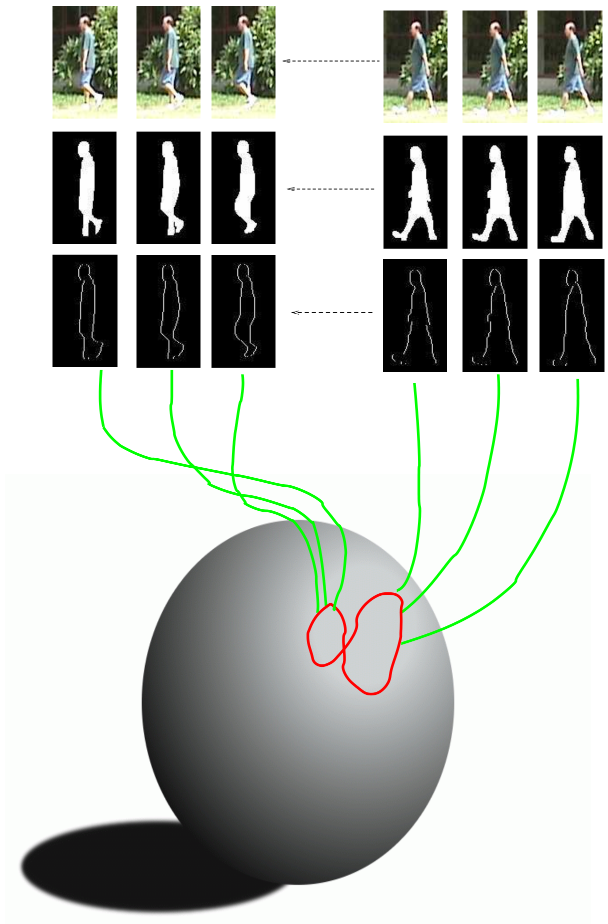
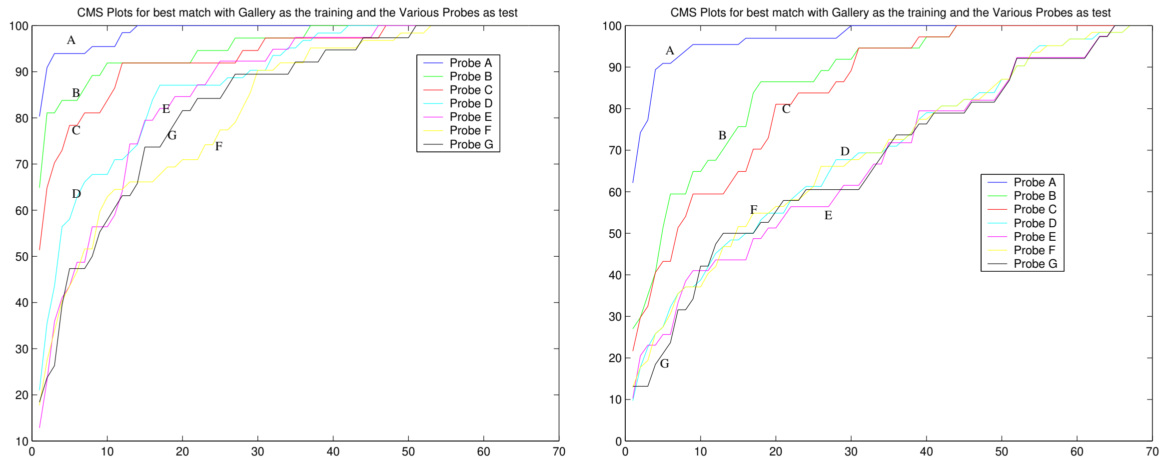


Fig. 2. Graphical illustration of the sequence of shapes obtained during gait

The following conclusions are drawn from Figure 4:

- The average CMS curve of the Stance Correlation method shows that shape without any kinematic cues provides recognition performance below baseline. The baseline algorithm is image correlation based and can be found in [29].
- The average CMS curve of the DTW method is better than that of Stance Correlation and close to baseline.
- The improvement in the average CMS curve in the DTW over that of the Stance Correlation method can be attributed to the presence of this implicit kinematics, because the algorithm tries to synchronize two warping paths.



(a) (b)

Fig. 3. Similarity matrix using(a)Dynamic Time Warping on shape space and (b)ARMA model on tangent space

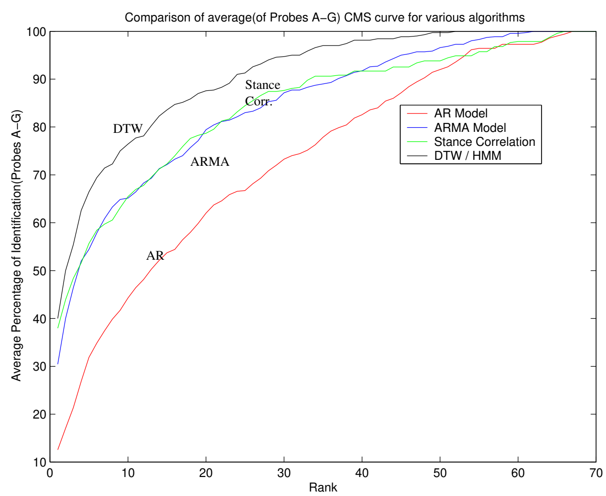


Fig. 4. Average(average of Probes A-G) CMS curves(Percentage of Recognition Vs Rank) using various methods.

- Both methods based on kinematics alone(Stance based AR and ARMA model) do not perform as well as the methods based on shape.
- The results support our belief that kinematics helps to boost recognition performance but is not sufficient as a stand-alone feature for person identification.
- The performance of the ARMA model is better than that of the Stance based AR model. This is because the observation matrix(C) encodes information about the features in the

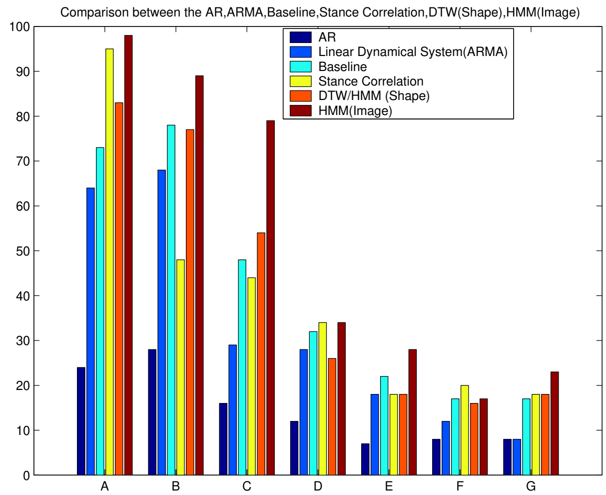


Fig. 5. Bar Diagram comparing the identification rate of various algorithms.

image, in addition to the dynamics encoded in the transition matrix(A).

- Similar conclusions may be obtained by looking at the CMS curves for the 7 experiments(Probes A-G) separately. We have shown the average CMS curve for simplicity.

Figure 5 shows a comparison of the identification rate(rank 1) of the various shape and kinematics based algorithms. It is clearly seen that shape based algorithms perform better than purely kinematics based algorithms. Note however, that a mere comparison of the identification rates will not lead to the conclusions above. For that we need to compare the average CMS curve of various methods(Figure 4). Also, as expected, using the images directly as the feature vector gives better results but at the cost of very high computational requirements.

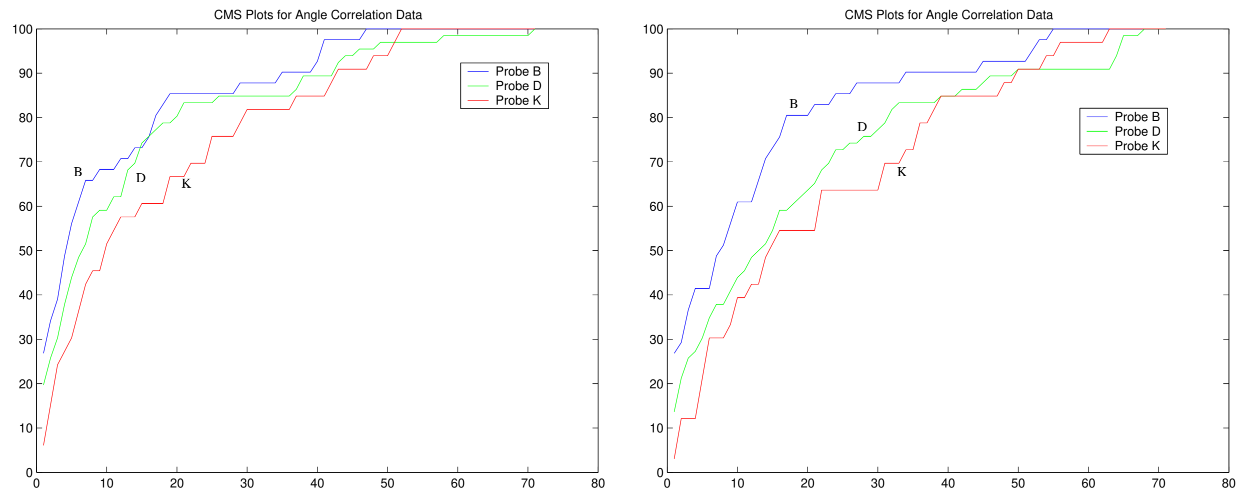
2) Results using Joint angles: In this section we describe experiments designed to verify the fact that our inference about the role of kinematics in gait recognition was not dependent on the feature that we chose for representation (Kendall's statistical shape). In order to test this, we performed some experiments on the actual physical parameters that are observable during gait i.e., the angles of various body parts. We used the manually segmented images provided in the USF dataset for these experiments. We inferred the angles (angle in the image plane) of eight joints (both shoulders, both Elbows, both Hips and both Knees) as the subjects walked frontoparallel to the camera. We used these angles (which are physically realizable parameters) as the features representing the kinematics of gait. We performed recognition experiments using the DTW directly on this feature. Figure 6(a) shows the CMS curves for three probes for which the manual segmented images were available. The recognition performance is comparable to purely kinematics based methods using our shape feature vector (refer to Figure 3(b)).

We also generated synthetic images of an individual walking using a Truncated Elliptic cone model for the human body and using the joint angles extracted from the manually segmented images. Figure 7 shows some sample images that were generated using this truncated elliptic cone model. We also ran recognition experiments on this simulated data using the DTW based shape sequence analysis method described in section 4.1. Figure 6(b) shows the CMS curves for this experiment. The results of these experiments are consistent with the experiments described earlier (3(b) and 6(a)), indicating that for the purposes of gait recognition, the amount of discriminability provided by the dynamics of the shape feature is similar to the discriminability provided by the dynamics of physical parameters like joint angles. This means that there is very little (if any) loss in using the dynamics of the shape feature instead of dynamics of the human body parts. Therefore, our inferences about the role of kinematics will most probably remain unaffected irrespective of the features used for representation.

The USF database does not contain any significant variation in terms of activity. Therefore, we cannot make any claims about the significance of kinematics and shape cues for activity modeling and recognition based on the experiments on the USF database. The CMU dataset enables this.

3) Results on the CMU Dataset: The CMU dataset has 25 subjects performing four different activities- fast walk, slow walk, walking with a ball and walking on an inclined plane. We perform an experiment on the recognition performance(i.e., identification rate) using the Stance Correlation(pure shape) method and compare our results with HMM based recognition results available at http://degas.umiacs.umd.edu/hid/cmu-eval.html.

The following conclusions are drawn from Table I:



(a) DTW on joint angles

(b) Shape sequence DTW on simulated data

Fig. 6. CMS curve using(a)DTW on joint angles and (b)Shape sequence DTW on simulated data

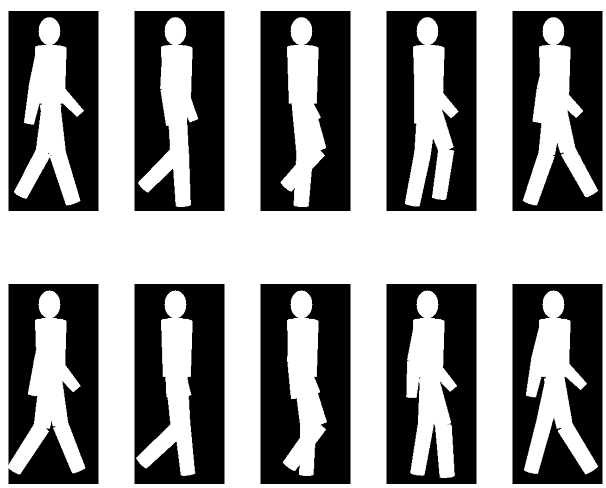


Fig. 7. Sequence of silhouettes simulated using joint angles and truncated elliptic cone human body model

- On a database of 25 people, the pure shape based method (Stance Correlation) provides almost 100% recognition when the Gallery and the Probe sets belong to the same activity. The improvement in performance over the USF dataset is because of higher quality of input video data.
- When we move across activities that change in shape also (eg. Slow Walk vs walking with a ball), we see that there is considerable degradation in recognition performance as expected.

Activity	Slow Walk	Fast Walk	Walk with Ball	Inclined plane
Slow Walk	100(72)	80(32)	48	48
Fast Walk	84	100(68)	48	28
Walk with Ball	68	48	92	12
Inclined plane	32	44	20	92

TABLE I

IDENTIFICATION RATES ON THE CMU DATA USING STANCE CORRELATION (BRACES DENOTE HMM IDENTIFICATION RATES)

- When we move across activities that differ only in their kinematics (eg.Slow Walk vs Fast Walk), we see that there is a slight degradation in recognition performance. The decrease in recognition performance of the purely shape based Stance Correlation method is not as drastic as is observed in the HMM method. This is because the HMM implicitly uses kinematics information for recognition. We can attribute the reduction in performance of the shape based method to the change in the shape of stances of the person due to a change in the walking speed [56].

4) Inferences about the role of shape in human movement analysis: The gait-based human recognition experiments on the USF database clearly indicate that given an activity (eg. gait), shape is more significant for person identification than kinematics. The experiment also indicates that kinematics does aid the task of recognition but pure kinematics is not enough for identification of an individual. The experiments on the CMU dataset indicate that when performing the same activity at differing speeds, a pure shape based approach tends to perform better than some other approaches that use kinematics also.

Activity	Slow Walk	Fast Walk	Walk with Ball
Slow Walk	100	72	64
Fast Walk	75	100	60
Walk with Ball	70	50	100

TABLE II

IDENTIFICATION RATES ON THE CMU DATA USING ARMA MODEL

C. Experiments on Activity Recognition

There are several scenarios where the manner in which the shape of on object changes provides clues about the nature of the activity being performed. Under these scenarios, we can use the methods we have proposed to perform activity recognition. We describe one such scenario in this section and also conduct experiments on activity recognition using the models we have built. The experiments on activity recognition are performed using the CMU and MOCAP datasets.

1) Results on the CMU dataset:

On the CMU dataset we did two experiments. First we conducted a recognition experiment using the ARMA model. In Figure 8 we have shown the similarity matrix that we obtained. The similarity matrix shown is a 75×75 matrix with the rows/columns numbered 1-25 representing different individuals performing slow walk, while rows/columns numbered 26-50 represent the corresponding individuals performing fast walk and rows/columns 51-75 representing the same individuals walking with a ball in their hand. The strong diagonal diagonal line indicates that identification performance for similar activities is very high. The four dark lines parallel to the diagonal indicate that identification is possible even when the activity performed is different. The actual identification rates are indicated in the table II.

Consider the three activities slow walk, walk with a ball and walk on an inclined plane. Considering the shape and kinematics of these three activities we expect that the ball alters the

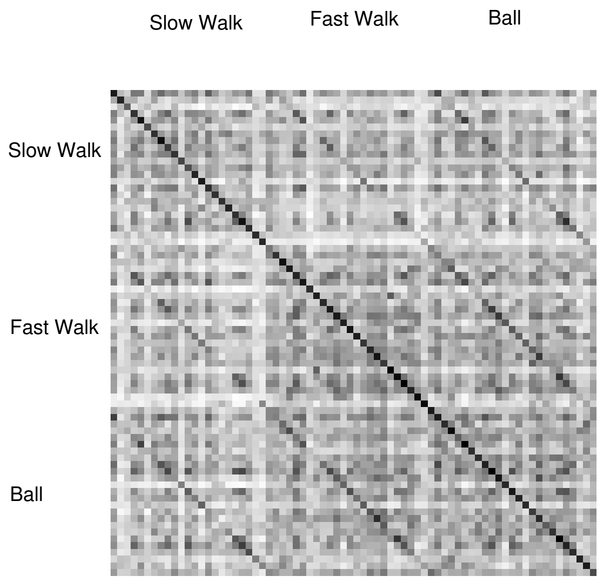
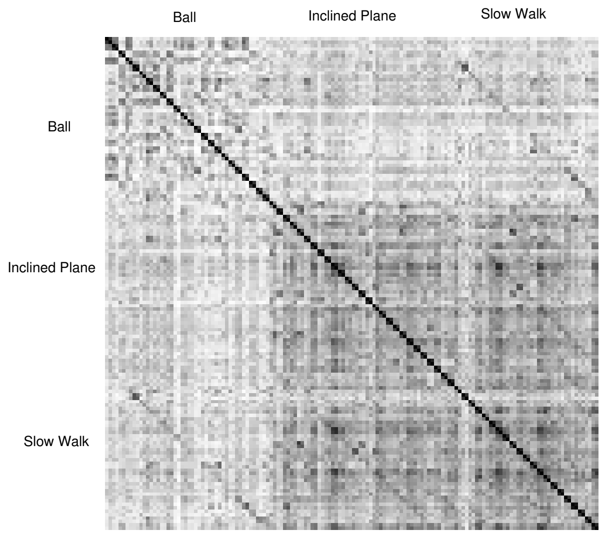


Fig. 8. Similarity matrix for the CMU data using the ARMA model

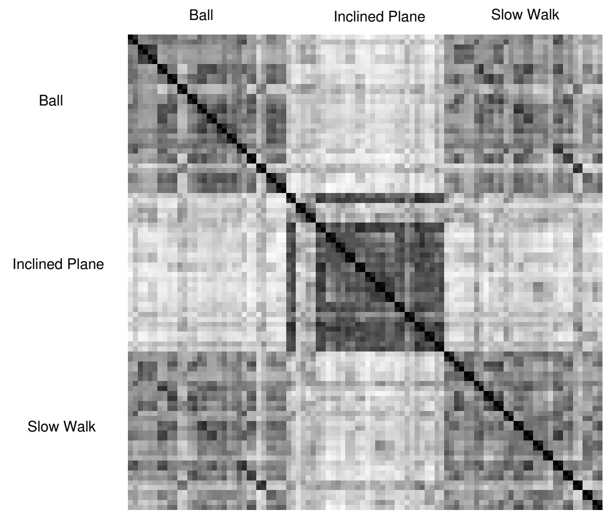
shape of the silhouette of the top half of the body, while the inclined plane alters the kinematics (and to a lesser extent shape) of the lower half of the body. In the first experiment we build a ARMA model for the shape of the top half of the body. Frobenius distance between the principal angles of the ARMA models is computed. Figure 9(a) shows the similarity matrix for the database of 25 people performing the three above mentioned activities when the model is built for the shape of the top half of the silhouette. Figure 9(b) shows a similar similarity matrix, when the model is built for the shape of the bottom half of the silhouette. The activity fast walk is distinctly different from all the other three activities in its kinematics (both in the top and the bottom half of silhouette) and therefore we did not use it in the current experiment.

The following conclusions may be drawn from Figure 9:

- From Figure 9(a), we see that walking with the ball is very dissimilar to both inclined plane and slow walk. Moreover both inclined plane and slow walk themselves are quite similar to each other since the inclined plane would significantly alter only the leg kinematics.
- From Figure 9(b), we see that walking on an inclined plane is very dissimilar to ball and slow walk. This indicates that a change in the kinematics of the lower half of the silhouette



(a)Top half of silhouette



(b) Bottom half of silhouette

Fig. 9. Similarity matrix using the ARMA model with (a)Top half of silhouette and (b)Bottom half of silhouette

affects the model. Moreover, we see that activities slow walk and ball remain quite similar to each other as expected.

2) Results on the MOCAP dataset: The MOCAP dataset consists of locations of 53 joints during a typical realization of several different activities. We use these joint locations to build an AR model and an ARMA model for each activity. The similarity matrix computed using both of these models, for the different activities[4] is shown in Figures 10 and 11. We notice that the discriminating power of a simple AR model (Fig.10) is not as good as that of the ARMA model (Fig.11). For example, we see that several different instances of walking are closer to each other in the ARMA model than in the AR model. This is because the ARMA model implicitly contains both shape and kinematics information. From the similarity matrix in Figure 11, we notice that the different kinds of walk are very similar to each other. The three different kinds of sitting poses are also very similar to each other. Moreover, walking as an activity is very different from sitting. As expected, jogging is very similar to walking while being dissimilar to sitting.

[4]walk1, walk2 and walk3 correspond to normal walking; walk4:exaggerated walking; walk5:walk with drooped shoulders and walk6:prowl walk

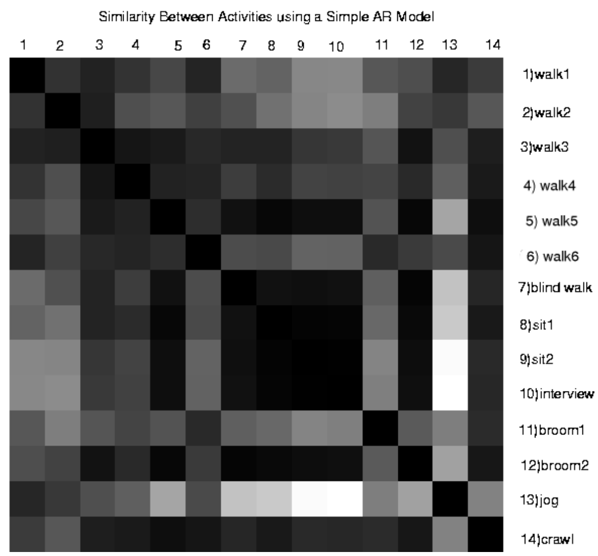


Fig. 10. Similarity matrix for the Mocap data using the AR model

These observations lead us to believe that the dynamical system contains enough information for activity classification.

3) Inferences about the role of kinematics in human movement analysis: The activity recognition based experiment on the CMU dataset indicates that a kinematics based approach does have the ability to differentiate activities that differ either in shape (slow walk vs ball) or in kinematics (slow walk vs inclined plane) because the system formulation(A, C, K) contains both shape information(C) and kinematics information(A). The ARMA model is also capable of performing person identification within a given activity when the number of subjects is small and the resolution of the image is high. The experiment on the MOCAP dataset reinforces our belief that the ARMA model can be used for activity recognition, even though its performance on person identification in the USF database(large number of subjects in outdoor environment) is not very good.

VI. Conclusions and Future Work

We have proposed novel methods for comparing two stationary shape sequences and shown their applicability to problems like gait recognition and activity recognition. The non-parametric

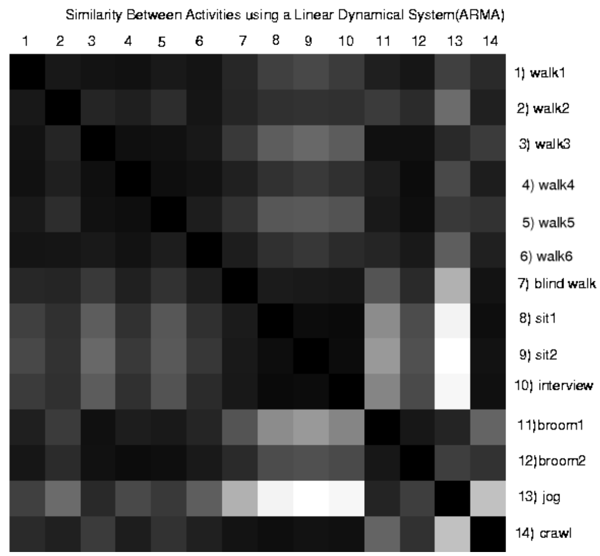


Fig. 11. Similarity matrix for the Mocap data using the ARMA model

method, using DTW is applicable to situations where there is very little domain knowledge and therefore parametric modeling of shape sequences is difficult. We have also used parametric AR and ARMA models on the tangent space projections of a shape sequence. The ability of these methods to serve as pattern classifiers for sequences of shapes has been shown by applying them to the problem of gait and activity recognition. We are currently working on building complex parametric models that capture more details about the appearance and motion of objects and models that can handle non-stationary shape sequences. We are also attempting to build models on the shape space instead of working with the tangent space projections.

Moreover, our experiments on gait recognition lead us to make an interesting observation about the role of shape and kinematics in human movement analysis from video. The experiments on gait recognition indicate that body shape is a significantly more important cue than kinematics for automated recognition, but using the kinematics of human body improves the person identification capability of shape based recognition systems.

REFERENCES

[1] S. Loncaric, "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31(8), pp. 983–1001, 1998.

[2] R. C. Veltkamp and M. Hagedoorn, "State of the art in shape matching," *Technical Report UU-CS-1999-27, Utrecht*, vol. 27, 1999.

[3] T. Pavlidis, "A review of algorithms for shape analysis," *Computer Graphics and Image Processing*, vol. 7, pp. 243–258, 1978.

[4] R. Kashyap and R. Chellappa, "Stochastic models for closed boundary analysis: Representation and reconstruction." *IEEE Trans. on Information Theory*, vol. 27, pp. 627–637, 1981.

[5] E. Persoon and K. Fu, "Shape discrimination using fourier descriptors," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 7(3), pp. 170–179, March 1977.

[6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, April 2002.

[7] H. Blum and R. Nagel, "Shape description using weighted symmetric axis features," *Pattern Recognition*, vol. 10, pp. 167–180, 1978.

[8] M. Hu, "Visual pattern recognition by moment invariants." *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.

[9] D. Kendall, "Shape manifolds, procrustean metrics and complex projective spaces," *Bulletin of London Mathematical society*, vol. 16, pp. 81–121, 1984.

[10] A. Khotanzad and Y. Hong, "Invariant image recognition by zernike moments." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12(5), pp. 489–497, 1990.

[11] C. C. Chen, "Improved moment invariants for shape discrimination," *Pattern Recognition*, vol. 26(5), pp. 683–686, 1993.

[12] A. Goshtasby, "Description and discrimination of planar shapes using shape matrices," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 738–743, 1985.

[13] S. Parui, E. Sarma, and D. Majumder, "How to discriminate shapes using the shape vector." *Pattern Recognition Letters*, vol. 4, pp. 201–204, 1986.

[14] H. Freeman, "On the encoding of arbitary geometric configurations," *IRE Transactions*, vol. 10, pp. 260–268, 1961.

[15] E. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell, "An efficiently computable metric for polygonal shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 209–216, 1986.

[16] A. Srivastava, W. Mio, E. Klassen, and S. Joshi, "Geometric analysis of continuous, planar shapes," *Proc. 4th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2003.

[17] F. Bookstein, "Size and shape spaces for landmark data in two dimensions," *Statistical Science*, vol. 1, pp. 181–242, 1986.

[18] I. Dryden and K. Mardia, *Statistical shape analysis*. John Wiley and sons, 1998.

[19] M. Prentice and K. Mardia, "Shape changes in the plane for landmark data," *The annals of statistics*, vol. 23-6, pp. 1960–1974, 1995.

[20] D. Geiger, T. Liu, and R. Kohn, "Representation and self-similarity of shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25-1, pp. 86–99, January 2003.

[21] R. Berthilsson, "A statistical theory of shape," *Statistical Pattern Recognition*, pp. 677–686, 1998.

[22] I. Dryden, “Statistical shape analysis in high level vision,” *IMA Workshop: Image analysis and High level vision*, 2000.

[23] A. Rangarajan, H. Chui, and F. Bookstein, “The softassign procrustes matching algorithm,” *Information Processing in medical imaging*, pp. 29–42, Springer 1997.

[24] A. Yezzi and S. Soatto, “Deformotion: Deforming motion, shape average and the joint registration and approximation of structure in images,” *International Journal of Computer Vision*, vol. 53(2), pp. 153–167, 2003.

[25] N. Vaswani, A. RoyChowdhury, and R. Chellappa, ““shape activities” : A continuous state hmm for moving/deforming shapes with application to abnormal activity detection,” *IEEE Trans. on Image Processing*, Accepted for Publication- 2004.

[26] C. Liu and N. Ahuja, “A model for dynamic shape and its applications,” *Conference on Computer Vision and Pattern Recognition*, 2004.

[27] S. D. Mowbray and M. S. Nixon, “Extraction and recognition of periodically deforming objects by continuous, spatio-temporal shape description.” *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 895–901, 2004.

[28] S. Niyogi and E. Adelson, “Analyzing and recognizing walking figures in xyt,” MIT Media Lab Vision and Modeling Group, Tech. Rep. 223, 1994.

[29] J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, “The gait identification challenge problem: Data sets and baseline algorithm,” *International Conference on Pattern Recognition*, August 2002.

[30] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *Workshop on Multimodal User Authentication (MMUA 2003)*, pp. 181–188, December 2003.

[31] L. Wang, H. Ning, W. Hu, and T. Tan, “Gait recognition based on Procrustes shape analysis,” *International Conference on Image Processing*, 2002.

[32] J. Foster, M. Nixon, and A. Prugel-Bennett, “Automatic gait recognition using area-based metrics.” *Pattern Recognition Letters*, vol. 24, pp. 2489–2497, 2003.

[33] A. Bobick and A. Johnson, “Gait recognition using static activity-specific parameters,” *Conference on Computer Vision and Pattern Recognition*, Dec. 2001.

[34] R. Collins, R. Gross, and J. Shi, “Silhoutte based human identification using body shape and gait,” *Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 351–356, 2002.

[35] A. Kale, A. Rajagopalan, Sundaresan.A., N. Cuntoor, A. Roy Cowdhury, V. Krueger, and R. Chellappa, “Identification of humans using gait,” *IEEE Trans. on Image Processing*, Sept. 2004.

[36] L. Lee, G. Dalley, and K. Tieu, “Learning pedestrian models for silhoutte refinement,” *International Conference on Computer Vision*, 2003.

[37] D. Tolliver and R. Collins, “Gait shape estimation for identification,” *4th Intl. Conf. on AVBPA*, June 2003.

[38] D. Cunado, M. Nash, S. Nixon, and N. Carter, “Gait extraction and description by evidence gathering,” *Proc. of the Intl. Conf. on AVBPA*, pp. 43–48, 1994.

[39] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto, “Recognition of human gaits,” *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 52–57, 2001.

[40] G. Johansson, “Visual perception of biological motion and a model for its analysis,” *PandP*, vol. 14, no. 2, pp. 201–211, 1973.

[41] E. Muybridge, *The Human Figure in Motion*. Dover Publications, 1901.

[42] D. Gavrilla, “The visual analysis of human movement: A survey,” *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, January 1999.

[43] E. Hoenkamp, “Perceptual cues that determine the labelling of human gait,” *Journal of Human Movement Studies*, vol. 4, pp. 59–69, 1978.

[44] M. Murray, A. Drought, and R. Kory, “Walking patterns of normal men,” *Journal of Bone and Joint surgery*, vol. 46-A(2), pp. 335–360, 1964.

[45] J. Cutting and L. Kozlowski, “Recognizing friends by their walk : Gait perception without familiarity cues,” *Bulletin of the Psychonomic Society*, vol. 9(5), pp. 353–356, 1977.

[46] J. Cutting and D. Proffitt, “Gait perception as an example of how we may perceive events,” *Intersensory perception and sensory integration*, 1981.

[47] G. Veres, L. Gordon, J. Carter, and M. Nixon, “What image information is important in silhouette based gait recognition?” *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 776–782, 2004.

[48] R. Tanawongsuwan and A. Bobick, “Modelling the effects of walking speed on appearance-based gait recognition,” *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 783–790, 2004.

[49] L. Rabiner and B. Juang, *Fundamentals of speech recognition.* Prentice Hall, 1993.

[50] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. Springer-Verlang, 1987.

[51] P. Overschee and B. Moor, “Subspace algorithms for the stochastic identification problem,” *Automatica*, vol. 29, pp. 649–660, 1993.

[52] S. Soatto, G. Doretto, and Y. Wu, “Dynamic textures,” *International Conference on Computer Vision*, vol. 2, pp. 439–446, 2001.

[53] G. Golub and C. Loan, *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1989.

[54] K. Cock and D. Moor, “Subspace angles and distances between ARMA models,” *Proc. of the Intl. Symp. of Math. Theory of networks and systems*, 2000.

[55] A. Veeraraghavan, A. RoyChowdhury, and R. Chellappa, “Role of shape and kinematics in human movement analysis,” *Conference on Computer Vision and Pattern Recognition*, 2004.

[56] A. Bobick and Tanawongsuwan, “Performance analysis of time-distance gait parameters under different speeds,” *4th Intl. Conf. on AVBPA*, June 2003.

Ashok Veeraraghavan Ashok Veeraraghavan received his B.Tech in Electrical Engineering from the Indian Institute of Technology, Madras in 2002 and M.S from the Department of Electrical and Computer Engineering at the University of Maryland, College Park in 2004. He is currently a Doctoral student in the Department of Electrical and Computer Engineering at the University of Maryland at College Park. His research interests are in signal, image and video processing, computer vision and pattern recognition.

AmitRoy Chowdhury Amit K. Roy-Chowdhury received the M.S. degree in Systems Science and Automation from the Indian Institute of Science, Bangalore in 1997 and Ph.D. from the Dept. of Electrical and Computer Engineering, University of Maryland, College Park in 2002. His PhD thesis was on statistical error characterization of 3D modeling from monocular video sequences. He is an Assistant Professor in the Dept. of Electrical Engineering, University of California, Riverside. He was previously with the Center for Automation Research, University of Maryland as a Research Associate, where he worked in projects related to face, gait and activity recognition. He is presently coauthoring a research monograph on recognition of humans and their activities from video. His broad research interests are in signal, image and video processing, computer vision and pattern recognition.

Rama Chellappa Rama Chellappa received the B.E. (Hons.) degree from the University of Madras, India, in 1975 and the M.E. (Distinction) degree from the Indian Institute of Science, Bangalore, in 1977. He received the M.S.E.E. and Ph.D. Degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981 respectively.

Since 1991, he has been a Professor of electrical engineering and an affiliate Professor of computer science at the University of Maryland, College Park. He is also affiliated with the Center for Automation Research (Director) and the Institute for Advanced Computer Studies (Permanent member). Prior to joining the University of Maryland, he was an Assistant (1981-1986) and Associate Professor (1986-1991) and Director of the Signal and Image Processing Institute (1988-1990) with the University of Southern California, Los Angeles. Over the last 23 years, he has published numerous book chapters, peer-reviewed journal and conference papers. He has edited a collection of Papers on Digital Image Processing (published by IEEE Computer Society Press), co-authored a research monograph on Artificial Neural Networks for Computer Vision (With Y.T. Zhou) published by Springer-Verlag, and co-edited a book on Markov Random fields (with A.K. Jain) published by Academic Press. His current research interests are face and gait analysis, 3D modeling from video, automatic target recognition from stationary and moving platforms, surveillance and monitoring, hyper spectral processing, image understanding, and commercial applications of image processing and understanding.

Dr. Chellappa has served as an associate editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IMAGE PROCESSING, and NEURAL NETWORKS. He was co-Editor-in-Chief of Graphical models and Image Processing and served as the Editor-in-Chief of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE during 2001-2004. He also served as a member of the IEEE Signal Processing Society Board of Governors during 1996-1999 and was the Vice President of Awards and Membership during 2002-2004. He has received several awards, including NSF Presidential Young Investigator Award, an IBM Faculty Development Award, the 1990 Excellence in Teaching Award from School of Engineering at USC, the 1992 Best Industry Related Paper Award from the International Association of Pattern Recognition (with Q. Zheng), the 2000 Technical Achievement Award from the IEEE Signal Processing Society. He was elected as a Distinguished Faculty Research Fellow (1996-1998) at the University of Maryland. He is a Fellow of the International Association for Pattern Recognition. He has served as a General the Technical Program Chair for several IEEE international and national conferences and workshops.