

419. Role of shape and kinematics in human movement analysis

Abstract

Human gait and activity analysis from video is presently attracting a lot of attention in the computer vision community. In this paper, we analyze the role of two of the most important cues in human motion- shape and kinematics. We present an experimental framework whereby it is possible to evaluate the relative importance of these two cues in computer vision based recognition algorithms. In the process, we propose a new gait recognition algorithm by computing the distance between two sequences of shapes that lie on a spherical manifold. In our experiments, shape is represented using Kendall's definition of shape. Kinematics is represented using a Linear Dynamical system. We place particular emphasis on human gait. Our conclusions show that shape plays a role which is more significant than kinematics in human identification using gait. As a natural extension we study the role of shape and kinematics in activity recognition. Our experiments indicate that we require models that contain both shape and kinematics in order to perform accurate activity classification. These conclusions also allow us to explain the relative performance of many existing methods in computer-based human activity modeling.

1. Introduction

Human activities are characterized by a sequence of poses or configurations of the human body. There has been considerable work in the psychophysics community on whether body shape or kinematics plays a greater role in our understanding of human activities[1][2]. In this paper we analyze the relative importance of shape and kinematics in human motion analysis from the point of view of existing computer algorithms. Two areas of focus are person identification using gait as a biometric and human activity analysis. Several algorithms have been suggested for the task of person identification at a distance using gait. Most of these algorithms extract certain features from the silhouettes of the person and attempt to identify individuals based on these features and the temporal variation of these features. While the features characterize the shape of the silhouette in various stances, the temporal variation of these features characterize the kinematics of these stances. Although intuition suggests that both shape and kinematics are important for recognition, there has been very little study on the impor-

tance of each of these cues in recognition performance. We attempt to address this issue. We propose a new gait recognition method based on shape and compare recognition results with other shape and kinematics based algorithms. We also look at the problem of activity recognition and provide some insights regarding the role of shape and kinematics for this task.

1.1 Motivation and Prior work

Considerable effort has been made in the computer vision, artificial intelligence and image processing communities to the understanding of human activities. A survey of work in human motion analysis can be found in [3]. Recently, significant effort has been devoted to the study of human gait. The motivation for using gait as a biometric for person identification comes from the observation that people are able to recognize others by simply observing their gait. Several studies have been done on the various cues that humans use for gait recognition[1][4]. Cutting and Kozlowski suggest that kinematic cues like speed, bounciness and rhythm are more important for human recognition than static cues like height. Cutting and Proffitt [2] 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 kinematics was crucial to gender discrimination using gait. Therefore, it is intuitive to expect that kinematics plays a role in person identification though shape information might also be equally important. In the light of such developments, we wish to explore the importance of shape and kinematics in human movement analysis from the perspective of computer vision algorithms and analyze their role in existing gait recognition methodologies.

There are several shape or kinematics based methods for activity based person identification. Collins et al. build a silhouette based nearest neighbour classifier [5] to do recognition. Several researchers have used Hidden Markov Models for the task of gait based identification [6][7][8]. Another shape based method for identifying individuals from noisy silhouettes is provided in [9].

Apart from these image based approaches Cunado et al. [10] model the movement of thighs as articulated pendulums and extract a gait signature using this model. In another kinematics based approach [11], trajectories of the various parameters of a kinematic model of the human body are used to classify different kinds of human activities. Thus

there have been several static image based approaches and some kinematics based approaches to do recognition. Some of these methods(like HMM) try to take advantage of both these cues. Yet the relative importance of shape and kinematics in human motion has not been understood.

In this paper, we propose a new gait recognition algorithm using shape by computing the distances between two shape sequences. The shapes lie on a spherical manifold. Dynamic time warping(DTW) is used to compute the distance between two paths on this manifold. We also propose a kinematics based recognition algorithm by modeling gait as a linear dynamical system. We learn a dynamical model for each individual from the video sequence of his walking. We compute the distance between two dynamical models and use this distance measure to perform recognition. We then present results of an experimental study in order to understand the numerical significance of shape and kinematics separately in state of the art gait recognition algorithms. As a natural extension, we also study their role in activity recognition.

2. Shape Analysis

“Shape is all the geometric information that remains when location, scale and rotational effects are filtered out from the object.”[12] In this paper we use the shape as defined in [12]. In this section we provide a brief 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.

2.1 Characterization of Shape

The binarized silhouette of a walking person is obtained. A shape feature is extracted from the binarized silhouette. This feature vector must be invariant to translation because identity should be independent of where the individual is located. It should also be invariant under scaling since identity should not depend on the distance of the subject from the camera. This yields the pre-shape of the walking person 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

$$Z_c = \frac{CX}{\|CX\|}, \text{ where } C = I_k - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T, \quad (1)$$

where I_k is a $k \times k$ identity matrix and $\mathbf{1}_k$ is a vector of k ones.

2.2 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 [12]. Consider two complex configurations X and Y with corresponding preshapes α and β . The Full Procrustes fit between X and Y is chosen so as to minimize

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

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

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

Since the preshapes are obtained after filtering out effects of translation and scale, 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.

$$d_P(X, Y) = \inf_{\Gamma \in SO(m)} \| \beta - \alpha \Gamma \| . \quad (4)$$

The Procrustes distance $\rho(X, Y)$ is the closest great circle distance between α and β on the preshape sphere. The three distance measures defined above are all trigonometrically related as given below:

$$d_F(X, Y) = \sin \rho \quad d_P(X, Y) = 2 \sin\left(\frac{\rho}{2}\right). \quad (5)$$

When the shapes are very close to each other there is very little difference between the various shape distances. Our experiments show that the choice of shape-distance does not alter recognition performance significantly since the shapes of a single individual lie close to each other. We show the results corresponding to the partial Procrustes distance in all our plots in this paper.

2.3 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.,

$$\mu = \arg \inf_{\mu} \sum d_F^2(Y_i, \mu). \quad (6)$$

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

$$v(\alpha, \mu) = \alpha \alpha^* \mu - \mu |\alpha^* \mu|^2. \quad (7)$$

where μ is the Procrustes mean shape of the data.

2.4 Shape based methods for Recognition

In order to study the importance of shape for recognition we use the Kendall's shape vector as the image feature. Recognition performance is obtained using three independent systems, a stance correlation based method, DTW [7] and another method based on the generic Hidden Markov Model(HMM) framework suggested in [6]. The Stance Correlation method uses only shape cues for recognition. The DTW and HMM attempt to exploit the kinematics in gait, although the kinematics is not explicitly modelled or captured.

2.4.1 Stance Correlation

Given a sequence of frames of a person walking, each frame is classified as belonging to one of six stances by first locating the cycle boundaries and then locating the stance boundaries. An exemplar is obtained for each stance as an average of all the frames belonging to that particular stance. The correlation between the corresponding exemplars of two sequences is used as the similarity score between these two sequences. Recognition is performed based on these similarity scores.

2.4.2 Dynamic Time Warping in shape space

We develop a new shape-based recognition algorithm using DTW on the spherical shape manifold. Given a sequence of a person walking, the shape of the silhouette is extracted for every frame. We use DTW to obtain the similarity score between two such shape sequences. The DTW algorithm enforces end point constraints, computes the local and cumulative error and then uses backtracking to obtain the warping

path. One major difference from the approach described in [7] is that we do the DTW between shapes on a spherical manifold instead of using Euclidean distances. The Procrustes shape distances are used as the local distance measure and end point constraints are satisfied by aligning the rest stances of two sequences. The cumulative error at the end of the best warping path is computed as the distance between two sequences of shapes that lie on the spherical manifold. This cumulative error is minimized when the trajectories corresponding to these shape sequences lie close to each other on the spherical shape space. Therefore the similarity between two sequences is high when this cumulative error is a minimum. Recognition is performed based on these similarity scores.

2.4.3 Hidden Markov Model with shape cues

Using a HMM for modeling gait has been studied in the last few years[6][8]. The postures that an individual adopts can be regarded as the states of the HMM and these states are distinct for each individual. The HMM is used in a probabilistic framework to perform recognition. The shape of the silhouette of the walking person is extracted for each frame and is used as the feature vector. We follow the HMM framework reported in [6] by replacing their image cues with our shape feature vector.

3. Dynamical Models

Gait is an activity with structured kinematics associated with it. Therefore, a purely shape based method is not an optimal approach to gait recognition. DTW and HMM attempt to capture the kinematics of gait, though not explicitly. In this section, we consider two kinematical models for gait and provide recognition results using these kinematical models. The first method uses an autoregressive(AR) model while the second approach uses an autoregressive and moving average model(ARMA).

3.1 Stance Based AR Model

In order to study the role that kinematics plays in gait recognition, we need to come up with a model for gait kinematics. This model must be insensitive to shape, i.e., if two people with vastly different shapes display the same kinematics, then the model inferred for these two individuals must be the same. This will ensure that we examine the effect of kinematics alone on recognition performance.

The video sequence of a person walking is divided into N distinct stances. Within each stance, we learn the dynamics of the shape vector. The time series of the tangent space projections of the pre-shape vector of each stance is modelled as a Gauss Markov process, i.e.,

$$\underline{\alpha}_j(t) = A_j \underline{\alpha}_j(t-1) + w(t) \quad (8)$$

where, w is a zero mean white Gaussian noise process and A_j is the transition matrix corresponding to the j^{th} stance. For convenience and simplicity A_j is assumed to be a diagonal matrix. Note that A_j is computed for each stance separately. Since the gait signal is periodic, we would need at least a second order model. However, for each stance there is no periodicity; hence the first order model is valid. We could, of course, use both the average shape of each stance(exemplar) and the transition matrices for each stance as the model of gait and use them for recognition. But, the exemplars contain shape information. Therefore, we use only the transition matrices for recognition in order to understand the role of kinematics.

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

$$D(A, B) = \sum_{j=1}^{j=N} \|A_j - B_j\|_F, \quad (9)$$

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

3.2 ARMA Model

The problem of gait recognition is transformed to one of learning a dynamical model from the observations and computing the distances between the dynamical models thus learnt. 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 models is non-trivial. Let us assume that the time-series of shapes is given by $\alpha(t), t = 1, 2, \dots, \tau$. Then an ARMA model is defined as[11]

$$\alpha(t) = Cx(t) + w(t); w(t) \sim N(0, R) \quad (10)$$

$$x(t+1) = Ax(t) + v(t); v(t) \sim N(0, Q). \quad (11)$$

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’’[13] which is unique.

3.2.1 Learning the 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 [13] and [14]. Given observations $\alpha(1), \alpha(2), \dots, \alpha(\tau)$, 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[13]). Note that in the innovation representation, the state covariance matrix $\lim_{t \rightarrow \infty} E[x(t)x^T(t)]$ is asymptotically diagonal. Let $[\alpha(1)\alpha(2)\alpha(3)\dots\alpha(\tau)] = U\Sigma V^T$ be the Singular value decomposition of the data. Then

$$\hat{C}(\tau) = U \quad (12)$$

$$\hat{A} = \Sigma V^T D_1 V (V^T D_2 V)^{-1} \Sigma^{-1} \quad (13)$$

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

3.2.2 Distance between ARMA Models

Subspace angles [15] between two ARMA models are defined as the principal angles ($\theta_i, i = 1, 2, \dots, n$) between the column spaces generated by the observability spaces of the two models extended with the observability matrices of the inverse models [16]. 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 [16]. Using these subspace angles $\theta_i, i = 1, 2, \dots, n$, three distances, Martin distance(d_M), gap distance(d_g) and Frobenius distance(d_F) between the ARMA models are defined as follows:

$$d_M^2 = \ln \prod_{i=1}^n \frac{1}{\cos^2(\theta_i)}, \quad (14)$$

$$d_g = \sin \theta_{max}, \quad (15)$$

$$d_F^2 = 2 \sum_{i=1}^n \sin^2 \theta_i. \quad (16)$$

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

4. Experimental Results

The various algorithms described above were used to obtain recognition results on two standard databases. The USF database [17] consists of 71 people in the Gallery¹. Various covariates like camera position, shoe type, surface and time were varied in a controlled manner to design a set of challenge experiments²[17]. The results are evaluated using cumulative match scores³(CMS) curves and the identification rate. The CMU database [5] 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

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.

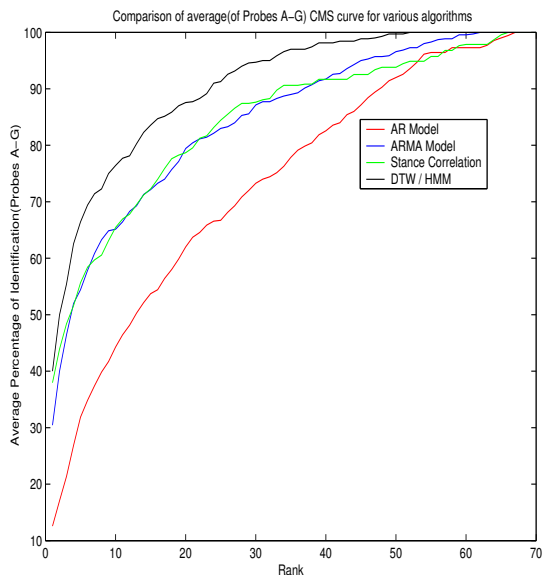


Figure 1: Average(average of Probes A-G) CMS curves(Percentage of Recognition Vs Rank) using the various shape and kinematics based methods.

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.

4.1 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.

4.2 Experiments on Gait Recognition

4.2.1 Results on the USF Database

On the USF database we conducted experiments on recognition performance using all the five methods- Stance Correlation,DTW on shape space,HMM using shape,Stance based AR and ARMA model. The CMS curves for challenge experiments A-G were obtained. Figure 1 shows the average CMS curves(average of the 7 Challenge experi-

ments:Probes A-G) for the various shape and kinematics based methods. An analysis of the average CMS curve is provided below.

Conclusions from Figure 1:

- 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 [17]).
- 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.
- The performance of the HMM with shape as the feature vector is very similar to that of the DTW.
- Both methods based on kinematics(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 image, in addition to the dynamics encoded in the transition matrix(A).
- Exactly similar conclusions may be obtained by looking at the CMS curves for the 7 experiments(Probes A-G) separately. But we have shown the average CMS curve for simplicity.

Figure 2 shows a comparison of the identification rate(rank 1) of the various shape and kinematics based algorithms⁴. 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 the various methods(Figure 1). Also, as expected using the images directly as the feature vector gives better results but at the cost of very high computational requirements.

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

⁴The results for HMM using Image features is from [6].

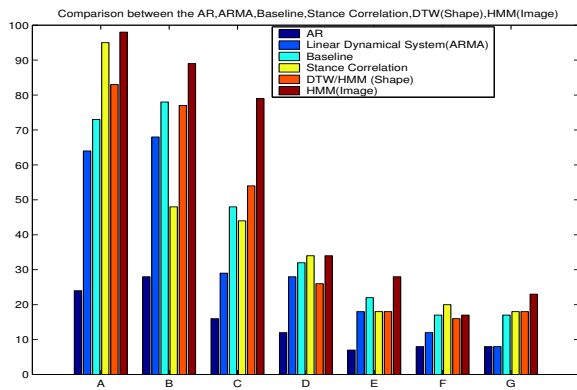


Figure 2: Bar Diagram comparing the identification rate of various algorithms.

for activity modeling and recognition based on the experiments on the USF database. The CMU dataset enables this.

4.2.2 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.umiaccs.umd.edu/hid/cmu-eval.html>.

Conclusions from Table 1:

- On a database of 25 people pure shape based method (Stance Correlation) provides almost 100% recognition when the Gallery and the Probe 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 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 our purely shape based Stance Correlation method is not as drastic as is observed in the HMM. This is because HMM implicitly uses kinematics information for recognition. We can attribute the reduction in performance of our shape based method to the change in shape of stances of the person due to a change in the walking speed [18].
- When we move across activities that change in shape also (eg. Slow Walk Vs Ball), we see that there is considerable degradation in recognition performance as expected.

Table 1: Identification rates on the CMU Dataset using Stance Correlation method (Figures in Braces denote HMM identification rates)

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

4.2.3 Inferences

The gait 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.

4.3 Experiments on Activity Recognition

In this section we discuss some experiments on activity recognition using the models we have built. The experiments on activity recognition are performed on the CMU and MOCAP datasets.

4.3.1 Results on the CMU dataset

On the CMU dataset we did two experiments to show the effect of kinematics and shape on activity recognition using our ARMA model. 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 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 3(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 3(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 our current experiment.

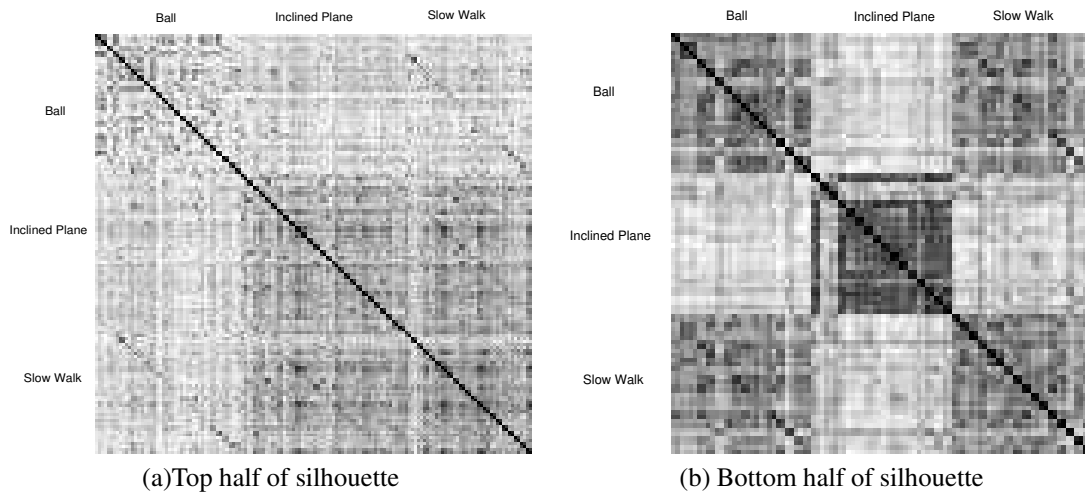


Figure 3: Similarity matrix using an ARMA model with (a)Top half of silhouette and (b)Bottom half of silhouette

The following conclusions may be drawn from Figure 3:

- From Figure 3(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 3(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 affects the model. Moreover, we see that activities slow walk and ball remain quite similar to each other as expected.

4.3.2 Results on the MOCAP dataset

The Mocap dataset consists of the locations of 53 joints during a typical realization of several different activities. We use these joint locations to build an AR model (without any stance distinctions) and an ARMA model for each activity. The similarity matrix computed using both these models, for the different activities is shown in Figure 4. We notice that the discriminating power of a simple AR model (Fig.4(a)) is not as good as that of the ARMA model (Fig.4(b)). 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 4(b) 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. These

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

4.3.3 Inferences

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) was not very good.

5. Conclusions and Future Work

The experiments conducted in this paper clearly indicate the following⁵:

1. Given an activity, shape is more important for person identification than kinematics. Moreover, methods based on shape are invariant to speed of the activity and therefore perform better when speed changes, provided the speed change does not alter the shape significantly.
2. Kinematics can help to enhance the performance of shape based person identification methods.
3. For the task of activity recognition we require models that encode both shape and kinematics (Refer to

⁵Conclusions may vary with technique used.

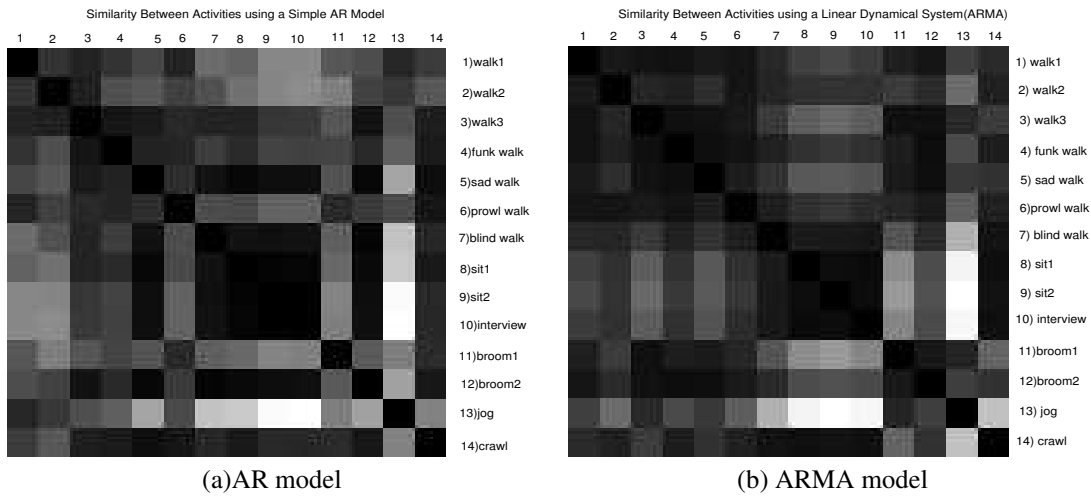


Figure 4: Similarity matrix using (a)AR model (b)ARMA model.

[19] for an activity classification technique based on Kendall's shape).

4. An ARMA dynamical system that encodes shape information also, has greater discriminating power for activities than a simple AR kinematical system.
5. The person identification capability of the ARMA model based system is significantly poorer than shape based methods(eg. DTW/HMM on shape space). However, with dynamical system models like ARMA, we can possibly perform person identification across activities that are not significantly different.

In future, we wish to build more advanced kinematical models of human motion and study the relation of shape and kinematics in such models.

References

- [1] 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.
- [2] J.E. Cutting and D.R. Proffitt, "Gait perception as an example of how we may perceive events," *Intersensory perception and sensory integration*, 1981.
- [3] D.M. Gavrilla, "The visual analysis of human movement: A survey," *CVIU*, vol. 73, no. 1, pp. 82–98, January 1999.
- [4] E. Hoenkamp, "Perceptual cues that determine the labelling of human gait," *Journal of Human Movement Studies*, vol. 4, pp. 59–69, 1978.
- [5] R.T. Collins, R. Gross, and J. Shi, "Silhouette based human identification using body shape and gait," *Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 351–356, 2002.
- [6] A. Sundaresan, A. Roy Chowdhury, and R. Chellappa, "A hidden markov model based framework for recognition of humans from gait sequences," *ICIP*, 2003.
- [7] A. Kale, A.N. Rajagopalan, Sundaresan.A., N. Cuntoor, A. Roy Chowdhury, V. Krueger, and R. Chellappa, "Identification of humans using gait," *IEEE Transactions on Image Processing*, Accepted.
- [8] L. Lee, G. Dalley, and K. Tieu, "Learning pedestrian models for silhouette refinement," *ICCV*, 2003.
- [9] D. Tolliver and R.T. Collins, "Gait shape estimation for identification," *4th Intl. Conf. on AVBPA*, June 2003.
- [10] D. Cunado, M.J. Nash, S.M. Nixon, and N.J. Carter, "Gait extraction and description by evidence gathering," *Proc. of the Intl. Conf. on AVBPA*, pp. 43–48, 1994.
- [11] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto, "Recognition of human gaits," *CVPR*, vol. 2, pp. 52–57, 2001.
- [12] I.L. Dryden and K.V. Mardia, *Statistical shape analysis*, John Wiley and sons, 1998.
- [13] P.V. Overschee and B.D. Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, pp. 649–660, 1993.
- [14] S. Soatto, G. Doretto, and Y.N. Wu, "Dynamic textures," *Intl. Conf. on Computer Vision*, vol. 2, pp. 439–446, 2001.
- [15] G. Golub and C.V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1989.
- [16] K.D. Cock and D.B. Moor, "Subspace angles and distances between ARMA models," *Proc. of the Intl. Symp. of Math. Theory of networks and systems*, 2000.
- [17] J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K.W. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," *ICPR*, August 2002.
- [18] A. Bobick and Tanawongsuwan, "Performance analysis of time-distance gait parameters under different speeds," *4th Intl. Conf. on AVBPA*, June 2003.
- [19] N. Vaswani, A. Roy Chowdhury, and R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," *CVPR*, 2003.