A Measure of Deformability of Shapes, with Applications to Human Motion Analysis

Amit K. Roy-Chowdhury

Dept. of Electrical Engineering,
University of California, Riverside, CA 92507, USA
amitrc@ee.ucr.edu

Abstract

In this paper we develop a theory for characterizing how deformable a shape is. We define a term called "deformability index" for shapes. The deformability index is computed from the tracked positions of a sequence of deformable shapes, using a scaled orthographic camera projection model. Our method assumes that a deformable shape sequence can be represented by a linear combination of basis shapes, where the weights assigned to each basis shape changes with time. The tracked points obtained from the shape sequence is transformed to a 3D shape space. Using statistical models to separate out the "true" deformations from those induced by noise in the trajectories, the dimension of this shape space is estimated using spectral analysis methods. The dimension of this shape space determines the number of basis shapes needed to represent the shape sequence, which, in turn, determines the deformability index. Rigid 3D transformations of the shape are taken into account in estimating the deformability index; however, the method does not require estimation of 3D structure or motion. Experimental results are shown using motion capture data as well as real imagery of different human activities. The results show that the deformability index is in accordance with our intuitive judgement and corroborates certain hypotheses in human movement analysis studies.

1 Introduction

Shape analysis is an active area of research in computer vision with applications in object recognition. Most of the current work in this area has focused on comparing the similarity of two or more shapes extracted from images. A detailed review of existing work in shape analysis is provided in Section 2. Recently, there has been some work on shape sequence analysis, including an understanding of the underlying dynamics. Applications of shape dynamics in event recognition and image formation have being stud-

ied in [6, 18, 19, 10]. In modeling the dynamics of shape evolution, it is important to separate out the "global" motion of the shape (i.e. the translation and rotation) from its "deformation", an issue that was analyzed in [12]. While there are well-defined measures for the global motion of an object, quantitative measures of its deformations are less well known. This paper addresses the issue of quantifying the deformation of a shape sequence by defining a "deformability index". This index measures how deformable a shape is. For a rigid shape (i.e., the shape does not change from one image frame to the next), the deformability index is one. We show how to derive this index in shape space using tools from spectral analysis. Experiments on real-life data of human activities are carried out and the results are in accordance with our intuitive judgment of the deformation involved in those activities and corroborate certain experimental findings in human gait analysis.

Consider the set of images of a walking person in Figure 1 (obtained from the USF database for the Gait Challenge problem [8]). The binary representation is used to clearly show the change in shape of the body for one complete walk cycle. Contrast this with the set of images that would be produced if the person was standing still for the same period of time. In the first case, the shape of the body deforms as the person walks, while in the latter case, the shape remains almost the same. If we consider a different example, like a dance sequence, the shape of the body changes and most viewers would agree that, in this case, the shape usually deforms more than in the case of a walk. Thus the shape of the human body while dancing is more deformable than that when walking, which is more deformable than when standing still. Our aim is to quantify this intuitive idea of deformability of a shape. Since it is possible for the human observer to obtain an idea of deformability based on the contents of the video sequence, the information about how deformable a shape is must be contained in the sequence itself. Therefore, it should be possible to quantify the deformability of the shape from a set of tracked points on the object.
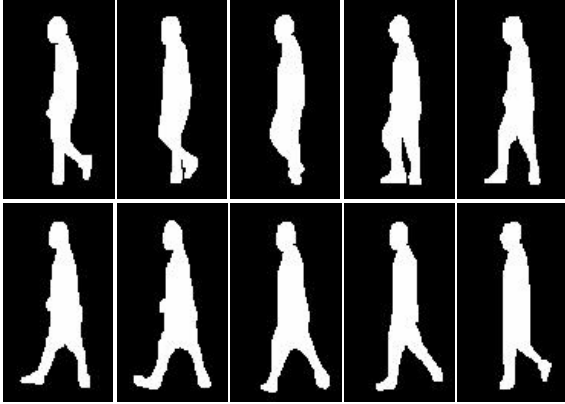
Figure 1. Sequence of shapes representing a human walk.

The aim of this paper is to derive a quantitative measure of shape deformability and justify it through experimental evaluation. The basic input for this derivation is the set of tracked feature points (landmark points) on the object (shape) over all the frames. This set of points is transformed to a 3D shape space using the ideas of deformable shape modeling in [17]. Using statistical models to separate out the "true" deformations from those induced by noise in the trajectories, the dimension of this shape space is estimated using tools from spectral analysis [13]. Experimental evaluation on simulation data, as well as real imagery, is presented.

The paper is organized as follows. We start with a review of relevant work in shape analysis in the next section. We then proceed to outline the theory for deriving the "deformability index" in Section 3. We first revisit the techniques for estimation of 3D deformable models and then present how they can be extended to compute a measure of deformability. Experimental evaluation is done on motion capture data and real-life video of various human activities. We conclude by outlining future extensions to this work.

2   Related Work

Some of the commonly used representations of shape are Fourier descriptors [20], extended Gaussian images [3], splines and deformable snakes [9], all of which model the shape of continuous curves. Active shape models [1] and Kendall's statistical shape theory [5, 2] have considered the shape of a discrete set of points. Methods for deformation of one shape into another and for comparing the similarity of two shapes have been proposed in [2, 7, 11]. However, there has been very little work on *shape sequence* processing. Some recent work in this area involves shape-

dynamical models for activity [18] and human motion analysis [19] and for image synthesis [6]. In the domain of 3D shape estimation from 2D images, the factorization theorem is one well-known approach, though it is usually applied under the assumption of a scaled orthographic camera projection model [16]. Its extension to modeling deformable shapes was proposed in [17], by approximating a non-rigid object by a composition of basis shapes, thus limiting the rank of the measurement matrix of the entire image sequence. In this paper, we propose a method for estimating the deformability of shapes by analyzing a sequence of such shapes from video. Though the calculation of the deformability index is carried out purely on the 2D shapes, it involves estimation of the dimension of the 3D shape space. Thus the method is capable of taking into account general 3D rigid transformations of the shape.

3   Estimating the Deformability Index

3.1   Intuitive Explanation

As a shape deforms, the position of the set of points defining the shape changes from one image frame to the next. The change in the position of this sequence of points determines how much the shape is changing, e.g. whether it is being squeezed or expanded or remaining the same. Defining a deformability index depends on the ability to obtain a mathematical description of this shape change. A shape sequence that deforms over time can be thought of to be composed of a number of basis shapes, where the weight given to each basis shape changes with time, thus leading to deformations in the original shape. This idea has been used by a number of authors and applied to studies in human motion and activity analysis [17, 4, 10]. It is usually the case that more deformable a shape is, more is the number of basis shapes required to represent it. However, there is no well-defined criterion for estimating the number of basis shapes. Presently, this is done heuristically by analyzing the physics of the underlying deforming process. At a minimum, a rigid shape would require only one basis shape, while there is no theoretical upper limit. Therefore we need a method to estimate the number of basis shapes from the point sequence.

The theoretical derivation which follows does precisely this. It proceeds by transforming the point sequence to a shape space and estimating the dimensionality of this shape space. Spectral analysis provides a method for achieving this purpose. The intuitive notion is that spatial deformation of the shape must be reflected in the frequency domain. The dimensionality of the shape space determines the deformability index. The noise in the sequence of feature positions is taken into account in order to correctly estimate the deformability index. Since the noise can randomly al-

ter the positions of the points, it can give a false notion of increased variability in the shape sequence, leading to a higher dimensionality of the shape space. Also, rigid 3D transformations of the shape can provide the impression of deformation. This will be factored out in estimating the deformability index. However, estimation of 3D structure will not be required for this purpose.

3.2 From Trajectories to Shape Space

We hypothesize that each shape sequence can be represented by a linear combination of 3D basis shapes. Mathematically, if we consider the trajectories of P points representing the shape (e.g. landmark points), then the overall configuration of the P points is represented as a linear combination of the basis shapes as

$$S = \sum_{i=1}^{K} l_i S_i, \qquad S, S_i \in \Re^{3 \times P}, l \in \Re. \qquad (1)$$

The choice of K determines the deformability of the shape sequence and is the focus of the derivation that follows. We will assume a scaled orthographic projection model for the camera.

A number of methods exist in the computer vision literature for estimating the basis shapes. In [16], the authors considered P points tracked across F frames in order to obtain two $F \times P$ matrices $\mathbf{U}$ and $\mathbf{V}$. Each row of $\mathbf{U}$ contains the x-displacements of all the P points for a specific time frame, and each row of $\mathbf{V}$ contains the corresponding y-displacements. It was shown in [16], that for 3D rigid motion under orthographic camera model, the rank, r, of $\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$ has an upper bound of 3. The rank constraint is derived from the fact that $\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$ can be factored into two matrices $\mathbf{M}_{2F \times r}$ and $\mathbf{S}_{r \times P}$, corresponding to the pose and 3D structure of the scene, respectively. In [17], it was shown that for non-rigid motion, the above method could be extended to obtain a similar rank constraint, but one that is higher than the bound for the rigid case. We will adopt the last mentioned method for computing the basis shapes. We will outline the basic steps of their approach in order to clarify the notation for the remainder of the paper.

Given F frames of a video sequence with P moving points, we can obtain the trajectories of all these points over the entire video sequence. These P points can be repre-

sented in a measurement matrix as

$$\mathbf{W}_{2F \times P} = \begin{bmatrix} u_{1,1} & \cdots & u_{1,P} \\ v_{1,1} & \cdots & v_{1,P} \\ \vdots & \vdots & \vdots \\ u_{F,1} & \cdots & u_{F,P} \\ v_{F,1} & \cdots & v_{F,P} \end{bmatrix}, \qquad (2)$$

where $u_{f,p}$ represents the x-position of the pth point in the fth frame and $v_{f,p}$ represents the y-position of the same point. Under weak perspective projection, the P points of a configuration in a frame f, are projected onto 2D image points $(u_{f,.}, v_{f,.})$ as

$$\begin{bmatrix} u_{f,1} & \cdots & u_{f,P} \\ v_{f,1} & \cdots & v_{f,P} \end{bmatrix} = \mathbf{R}_f \left(\sum_{i=1}^{K} l_{f,i} S_i \right) + \mathbf{T}_f, \quad (3)$$

where,

$$\mathbf{R}_f = \begin{bmatrix} r_{f,1} & r_{f,2} & r_{f,3} \\ r_{f,4} & r_{f,5} & r_{f,6} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{R}_f^{(1)} \\ \mathbf{R}_f^{(2)} \end{bmatrix}. \qquad (4)$$

$\mathbf{R}_f$ represents the first two rows of the full 3D camera rotation matrix and $\mathbf{T}_f$ is the camera translation. The translation component can be eliminated by subtracting out the mean of all the 2D points, as in [16]. We now form the measurement matrix $\mathbf{W}$, which was represented in (2), with the means of each of the rows subtracted. The weak perspective scaling factor is implicitly coded in the configuration weights, $\{l_{f,i}\}$.

Using (2) and (3), it is easy to show that

$$
\begin{aligned}
\mathbf{W} &= \begin{bmatrix} l_{1,1}\mathbf{R}_1 & \cdots & l_{1,K}\mathbf{R}_1 \\ l_{2,1}\mathbf{R}_2 & \cdots & l_{2,K}\mathbf{R}_2 \\ \vdots & \vdots & \vdots \\ l_{F,1}\mathbf{R}_F & \cdots & l_{F,K}\mathbf{R}_F \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_K \end{bmatrix} \qquad (5) \\
&= \mathbf{Q}_{2F \times 3K} . \mathbf{B}_{3K \times P}, \qquad (6)
\end{aligned}
$$

which is of rank $3K$. The matrix $\mathbf{Q}$ contains the pose for each frame of the video sequence and the weights $l_1, ..., l_K$. The matrix $\mathbf{B}$ contains the basis shapes corresponding to each of the activities. In [17], it was shown that $\mathbf{Q}$ and $\mathbf{B}$ can be obtained using singular value decomposition (SVD), and retaining the top $3K$ singular values, as $\mathbf{W}_{2M \times P} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ and $\mathbf{Q} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{B} = \mathbf{D}^{\frac{1}{2}}\mathbf{V}^T$.

3.3 Estimating Deformability Index in Shape Space

The above mentioned rank constraint requires knowledge of K in order to estimate the shape and motion parameters. This is usually determined heuristically from the physics of the object whose structure is being estimated.

We now provide a theoretical method for estimating K by reinterpreting the above equations in stochastic framework. In turn, it leads to a definition of deformability of a shape sequence.

Consider the set of coordinates representing the shape of the deformable object in a particular frame of a video sequence to be the realization of a random process. The sequence of frames depicts the deformation of the shape, along with the effects of the 3D translation and rotation. Represent the x and y coordinates of the sampled points in a *single* frame as a vector $\mathbf{y} = [u_1, ..., u_P, v_1, ..., v_P]^T$. Then, from (6), it is easy to show that for K basis shapes (K is unknown)

$$
\mathbf{y}^T = \left[l_1\mathbf{R}^{(1)}, ..., l_K\mathbf{R}^{(1)}, l_1\mathbf{R}^{(2)}, ..., l_K\mathbf{R}^{(2)} \right] *
$$
$$
\begin{bmatrix} S_1 & & \\ \vdots & & 0 \\ S_k & & \\ & & S_1 \\ 0 & & \vdots \\ & & S_k \end{bmatrix} + \eta^T, \quad (7)
$$

$$
i.e., \quad \mathbf{y} = (\mathbf{q}_{1\times 6K}\mathbf{b}_{6K\times 2P})^T + \eta
$$
$$
= \mathbf{b}^T\mathbf{q}^T + \eta, \quad (8)
$$

where η represents the noise in the sequence of tracked points and is assumed to be a zero-mean random process. The vector $\mathbf{q}$ is obtained by juxtaposing two consecutive rows of $\mathbf{Q}$, corresponding to the same image frame, in equation (6). The matrix $\mathbf{b}$, which is constant across all the frames, is obtained by duplicating $\mathbf{B}$ in equation (6), as shown in equation (7).

Assuming that the coordinates of the points representing the shape in all the F frames can be considered to be realizations of the same random process (which is a reasonable assumption since they represent the same shape), with possibly different noise statistics, we can compute the correlation matrix of $\mathbf{y}$. Let $\mathbf{R_y} = E[\mathbf{yy}^T]$ be the correlation matrix of $\mathbf{y}$ and $\mathbf{C}_\eta$ the covariance matrix of η. Hence,

$$
\mathbf{R_y} = \mathbf{b}^T E[\mathbf{q}^T\mathbf{q}]\mathbf{b} + \mathbf{C}_\eta. \quad (9)
$$

The correlation matrix, $\mathbf{R_y}$, is of size $2P \times 2P$ and can be estimated from the sequence of points representing the shapes as $\mathbf{R_y} = \frac{1}{F}\sum_{f=1}^{F} \mathbf{y}_f\mathbf{y}_f^T$, where $\mathbf{y}_f$ is the vector $\mathbf{y}$ (defined above) in the frame f. The expectation on the right hand side of equation (9) can be computed similarly as $E[\mathbf{q}^T\mathbf{q}] = \frac{1}{F}\sum_{f=1}^{F} \mathbf{q}_f^T\mathbf{q}_f$, where $\mathbf{q}_f$ is the vector $\mathbf{q}$ (defined above) for frame f and is obtained from the matrix $\mathbf{Q}$ in equation (6).

The noise covariance matrix, C_η, represents the accuracy with which the feature points are tracked and needs to be estimated from the image frames. Since η need not be an independent and identically distributed (IID) noise process, C_η will not necessarily have a diagonal structure (but it is symmetric and positive semi-definite). For the purposes of setting a precise threshold (which will become clear soon), it is desirable that $\mathbf{C}_\eta$ be a diagonal matrix.

Consider the diagonalization of $\mathbf{C}_\eta = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, where $\mathbf{\Lambda} = \text{diag}[\mathbf{\Lambda}_s, 0]$ and $\mathbf{\Lambda}_s$ is an $L \times L$ matrix of non-zero singular values of $\mathbf{\Lambda}$. Let $\mathbf{P}_s$ denote the orthonormal columns of $\mathbf{P}$ corresponding to the non-zero singular values. Therefore,

$$
\mathbf{C}_\eta = \mathbf{P}_s\mathbf{\Lambda}_s\mathbf{P}_s^T. \quad (10)
$$

Premultiplying equation (8) by $(\mathbf{P}_s\mathbf{\Lambda}_s^{\frac{1}{2}})^{-1}$, we see that (8) becomes

$$
\tilde{\mathbf{y}} = \tilde{\mathbf{b}}^T\mathbf{q}^T + \tilde{\eta}, \quad (11)
$$

where $\tilde{\mathbf{y}} = \mathbf{\Lambda}_s^{-\frac{1}{2}}\mathbf{P}_s^T\mathbf{y}$ is a $L \times 1$ vector, $\tilde{\mathbf{b}}^T = \mathbf{\Lambda}_s^{-\frac{1}{2}}\mathbf{P}_s^T\mathbf{b}^T$ is a $L \times 6K$ matrix and $\tilde{\eta} = \mathbf{\Lambda}_s^{-\frac{1}{2}}\mathbf{P}_s^T\eta$. It can be easily verified that the covariance of $\tilde{\eta}$ is an identity matrix $\mathbf{I}_{L\times L}$. This is known as the process of "whitening", whereby the noise process is transformed to be IID [13].

Representing by $\mathbf{R}_{\tilde{\mathbf{y}}}$ the correlation matrix of $\tilde{\mathbf{y}}$, it can be seen that

$$
\mathbf{R}_{\tilde{\mathbf{y}}} = \tilde{\mathbf{b}}^T E[\mathbf{q}^T\mathbf{q}]\tilde{\mathbf{b}} + \mathbf{I} = \mathbf{\Delta} + \mathbf{I}, \quad (12)
$$

where, for simplicity, $\mathbf{\Delta} \triangleq \tilde{\mathbf{b}}^T E[\mathbf{q}^T\mathbf{q}]\tilde{\mathbf{b}}$. Now, $\mathbf{R}_{\tilde{\mathbf{y}}}$ is of dimension $L \times L$, $\tilde{\mathbf{b}}^T$ is of size $L \times 6K$ and $E[\mathbf{q}^T\mathbf{q}]$ is of size $6K \times 6K$. Thus, $\mathbf{\Delta}$ has maximum rank $6K$, where K is the number of basis shapes (assuming $L > 6K$). This is based on the fact that if $\mathbf{A}_{m\times n} = \mathbf{F}_{m\times r}\mathbf{G}_{r\times n}$, then the Rank$(\mathbf{A}) \leq r$. For a general 3D scene undergoing translation and rotation, the rank will be $6K$, which is the case we will consider below. Representing by $\mu_i(\mathbf{A})$ the ith eigenvalue of the matrix $\mathbf{A}$, we see that

$$
\begin{aligned}
\mu_i(\mathbf{R}_{\tilde{\mathbf{y}}}) &= \mu_i(\mathbf{\Delta}) + 1, \quad \text{for} \quad i = 1, ..., 6K, \quad \text{and} \\
\mu_i(\mathbf{R}_{\tilde{\mathbf{y}}}) &= 1, \quad \text{for} \quad i = 6K+1, ..., L. \quad (13)
\end{aligned}
$$

Hence, there are $6K$ eigenvalues above 1. By counting the number of eigenvalues that are greater than 1 and dividing it by 6, we can obtain an estimate of K, which is the dimensionality of the shape space represented by the sequence of deforming points. Since K denotes the number of basis shapes that can model the feature point sequence, it provides a measure of the deformability of the shape sequence. The more the number of basis shapes required to model a shape sequence, the more deformable it is. Thus, for a general 3D scene undergoing translation and rotation, we have

$$
\text{Deformability Index} = \frac{\#\text{eigenvalues of } \mathbf{R}_{\tilde{\mathbf{y}}} > 1}{6}. \quad (14)
$$

3.4 Properties of the Deformability Index

- For the case of a 3D rigid body, the deformability index is 1. In this case, the only variation in the values of the vector $\mathbf{y}$ from one image frame to the next is due to the global rigid translation and rotation of the object. The rank of the matrix $\boldsymbol{\Delta}$ will be 6 [17, 16] and the deformability index will be 1.

- For the special case of a planar scene, the corresponding rank of $\boldsymbol{\Delta}$ would be $4K$, and thus the deformability index should be calculated by dividing the number of eigenvalues over 1 by 4.

- Estimation of the deformability index does not require explicit computation of the 3D structure and motion in equation (6), since we need to compute the eigenvalues of the covariance matrix of the 2D feature positions. In fact, for estimating the shape and rotation matrices in equation (6) it is essential to know the value of K. Thus the method outlined in Section 3.3 should precede computation of the shape in Section 3.2. Using our method, it is possible to obtain an algorithm for deformable shape estimation without having to guess the value of K.

- The computation of the deformability index takes into account any rigid 3D translation and rotation of the object (as recoverable under a scaled orthographic camera projection model), even though it has the simplicity of working only with the covariance matrix of the 2D projections. Thus it is more general than a method that considers purely 2D image plane motion.

- The "whitening" procedure described above enables us to choose a *fixed* threshold of one for comparing the eigenvalues.

- The proposed algorithm in non-iterative. This is unlike other approaches (e.g. [4, 17]) where the sum of squared differences is minimized to determine K.

4 Experimental Results

Experimental evaluation of the above theory was carried out on real life imagery. The experiments were in the domain of human motion analysis, where the deformation of the human body during various activities was estimated using the deformability index. We performed two sets of experiments. In the first one, motion capture data for various activities was used as the input. We computed the deformability index of the human body for each of these activities, and found them to be very consistent with what would be expected intuitively by a human observer. Next, we applied our theory to walking sequences of humans as available in

the USF Gait Challenge Database [8]. Here we found that our deformability estimates are in accordance with some of the results on shape representation reported in the gait recognition literature. In both the experiments, the shapes were centered in the image frames, scaled and aligned so as to make the human body upright.

4.1 Experiments With Motion Capture Data

We used the motion-capture data available from Credo Interactive Inc. and Carnegie Mellon University in the Bio-Vision Hierarchy and Acclaim formats. It has a number of examples of different activities and is thus a rich dataset for studying shape sequences. [1] The combined dataset included a number of subjects performing various activities, like walking, jogging, sitting, crawling, brooming, etc. For each of these activities, we had multiple video sequences. Also, many of the activities contained video from different viewpoints.

Using the video sequences and the theory outlined in Section 3.2, we computed the 3D basis shapes and their combination coefficients (see equation (1)). The first basis shapes are shown in Figure 2 for six different activities.

For the different activities in this database, we computed the deformability index from equation (14). The deformability index, computed for each of these sequences, is shown in Table 1. Since this value denotes the number of basis shapes required to represent the video sequences, we resynthesized the original sequences using the basis shapes and combination coefficients obtained from equation (6). Equation (3) was used for the synthesis and the value of K was determined by the procedure in Section 3. In all the cases, the error at none of the feature points was more than 1 pixel.

From Table 1, a number of interesting observations can be made. For the walk sequences, the deformability index was between 5 and 6. This matches the hypotheses in papers on gait recognition where it is mentioned that about five exemplars are necessary to represent a full cycle of gait [4]. The number of basis shapes increases for fast walk, as expected from some of the results in [15]. When the person walks doing some other things (like moving head or hands or a blind person's walk), the number of basis shapes needed to represent it (i.e. the deformability index) increases from that of normal walk. The result that might seem surprising initially is the high deformability index for sitting sequences. On closer examination though, it was found that the person, while sitting, was making all kinds of random gestures as in talking to someone else. That increased the deformability index for these sequences. The deformability index, thus calculated, is used to estimate the 3D shape, some of which are shown in Figure 2.

---

[1] While there are a number of standard datasets for shapes, we could not find many for the study of shape sequences.
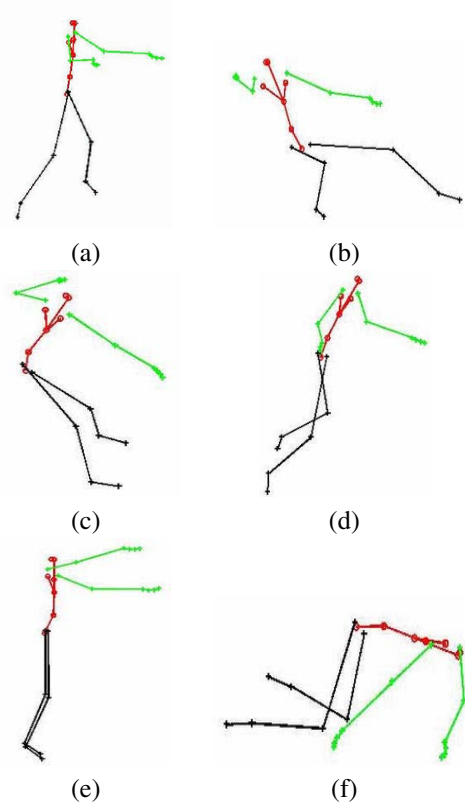
Figure 2. (a) - (f): Plots of the first basis shape, S_1, for walk, sit, broom, jog, blind walk and crawl sequences, respectively.

4.2 Experiments on Gait Dataset

The USF Gait Challenge Dataset [8] was used for our experiments because of two reasons. It has a number of examples of different people walking under different conditions. Thus it would allow us to test the consistency of the estimates for the deformability index. Secondly, a number of researchers have reported results on this dataset and thus we would be able to corroborate our conclusions with their results.

We used the background subtracted images of the walking person, when the person is presenting a side view to the camera, as shown in Figure 1. The outer boundary of the person was sampled in order to obtain the shape vector. The method described in [14] was adopted to estimate the variance of the noise in the feature positions from the original images. The method uses the inverse of the Hessian matrix of the second-order partial derivatives of the intensity along the horizontal and vertical axes. By using the same number of sample points in each frame, an approximate correspondence was maintained between the feature points in the different frames. We experimented with 10 subjects walk-

Table 1. Deformability Index for Human Activities Using Motion Capture Data

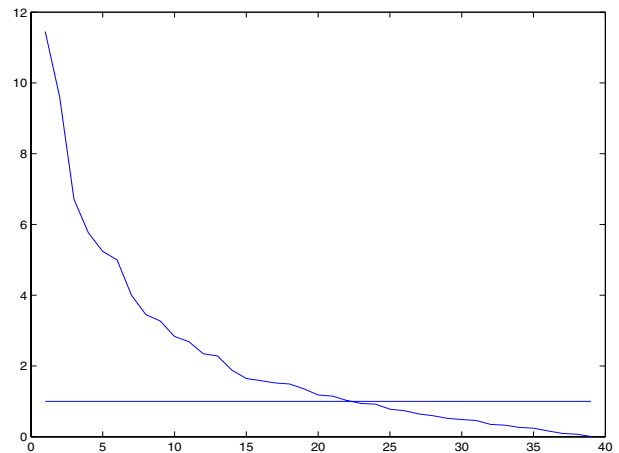| Activity	Deformability Index
Walk (Seq. 1)	5.8
Walk (Seq. 2)	4.7
Fast Walk	8.0
Walk while throwing hands around	6.8
Walk with drooping head	8.8
Blind walk	8.8
Female Walk	7.0
Slow Dance	8.0
Jog	5.0
Broom (Seq. 1)	7.5
Broom (Seq. 2)	8.8
Broom in a circle	9.0
Crawl	8.0
Sit (Seq. 1)	8.0
Sit (Seq. 2)	8.2
Sit (Seq. 3)	8.2

Figure 3. Plot of the eigenvalues, in decreasing order of magnitude, for a typical walking sequence in the USF database.

ing on grass and concrete surfaces and wearing different types of shoes. For all the cases, the deformability index ranged from 3.8 to 5.2. Figure 3 shows a typical plot of the eigenvalues arranged in descending order of magnitude along with the threshold of one. It has been noted in [4] that four to five exemplars are needed to represent a complete cycle of gait. Reasonably good recognition results were obtained using their algorithm on the USF database. Our analysis, presented in this paper, provides a theoretical justification for the choice of the number of exemplars. The results show that the deformability index for human walk is indeed between 4 and 5.

5 Conclusions

In this paper, we have a presented a method for estimating the deformability of shape sequences obtained from a set of video frames. We assumed a scaled orthographic camera projection model and that a deformable shape can be represented using a linear combination of basis shapes. The theory relied on estimating the number of 3D basis shapes from the 2D feature positions representing the shapes. The deformability index is directly related to the number of basis shapes. The computation of the deformability index can handle rigid 3D transformations of the shape, though it does not require prior estimation of the 3D structure or motion. We presented experimental results in human movement analysis using motion capture and real-life video images. The estimates of the deformability index are in accordance with what would be expected intuitively by a human observer and corroborate certain hypotheses in the existing literature on human motion analysis.

References

[1] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[2] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.

[3] B. Horn. Extended gaussian images. *Proceedings of the IEEE*, 72:1671–1686, 1984.

[4] A. Kale, A. Rajagopalan, A. Sundaresan, N. Cuntoor, A. Roy-Chowdhury, A. Krueger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. on Image Processing*, pages 1163–1173, September 2004.

[5] D. Kendall, D. Barden, T. Carne, and H. Le. *Shape and Shape Theory*. John Wiley and Sons, 1999.

[6] C. Liu and N. Ahuja. A model for dynamic shape and its applications. In *Computer Vision and Pattern Recognition*, pages II: 129–134, 2004.

[7] W. Mio and A. Srivastava. Elastic-string models for representation and analysis of planar shapes. In *Computer Vision and Pattern Recognition*, 2004.

[8] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. *Proc of the International Conference on Pattern Recognition*, 2002.

[9] D. Rogers and J. Adams. *Mathematical Elements of Computer Graphics*. WCB/McGraw-Hill, 1990.

[10] A. Roy-Chowdhury and R. Chellappa. A factorization approach for event recognition. In *CVPR Event Mining Workshop*, 2003.

[11] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *PAMI*, 26(5):550–571, May 2004.

[12] S. Soatto and A. Yezzi. Deformotion: Deforming motion, shape average and the joint registration and segmentation of images. In *European Conference on Computer Vision*, page III: 32 ff., 2002.

[13] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[14] Z. Sun, V. Ramesh, and A. Tekalp. Error characterization of the factorization method. *Computer Vision and Image Understanding*, 82(2):110–137, May 2001.

[15] R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. In *Computer Vision and Pattern Recognition*, pages II:783–790, 2002.

[16] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9:137–154, November 1992.

[17] L. Torresani and C. Bregler. Space-time tracking. In *European Conference on Computer Vision*, 2002.

[18] N. Vaswani, A. Roy-Chowdhury, and R. Chellappa. Activity recognition using the dynamics of the configuration of interacting objects. In *Computer Vision and Pattern Recognition*, 2003.

[19] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *Computer Vision and Pattern Recognition*, 2004.

[20] C. Zahn and R. Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, C-21:269–281, 1972.