CAPTURING HUMAN ACTIVITY BY A CURVE

Sheng Yi, Hamid Krim

North Carolina State University
{syi,ahk}@ncsu.edu

ABSTRACT

One of the main challenges of human behavior analysis is the high dimensionality of the representation space. In shape representation, however, a specific human behavior may naturally be described by a 1D path which lies in shape space. According to Whitney Embedding Theorem, such a 1D manifold may be embedded in $\mathbb{R}^3$. Motivated by the potential of reducing the dimensionality of behavior representation, we construct an embedding to map the path of evolution of the silhouette in shape space to a representational curve in $\mathbb{R}^3$. In contrast to other behavioral embedding, where each point of the path in shape space is projected to lower dimension, we embed the homotopy function of the whole path to be a planar curve function. The proposed embedding utilizes sampling theory to provide computational efficiency and simple reconstruction from the embedding space. Upon validating such a representation, we proceed to model different activities by an AR model of the representative curve. Experiments are provided to illustrate our technique and to demonstrate its viability.

***Index Terms*—** embedding, homotopy, curve, behavior modeling, auto regression

1. INTRODUCTION

Human behavior recognition is a very important problem in many vision based applications like, human identification, video surveillance, video content indexing. In recognition problems, the geometry of the representational space is crucial to determining a distance measure for subsequent classification applications. In this paper, we focus on constructing a lower dimensional behavior representation that is generically more amenable to statistical analysis.

For representational purposes, research in behavior analysis has described any activity as a flow of interesting points or as a sub-volume in the spatio-temporal intensity volume [1] [6] [7]. Many parametric template-based representations have been proposed including, stick figure, 2D polygons and volumetric models. The main challenge of many of these representations is the computational burden of parameter estimation they entail in a high dimensional space. More recently, the shape dynamic representation of human behavior was shown to have better performance in human identification [8] [2] [5]. In shape space, the human motion in video may be viewed as a path connecting a starting human silhouette and an ending human silhouette. The advantage of shape-based representation is the existing preponderance of results on shape analysis which can naturally be applied to behavioral modeling and analysis. In [5], for example, the distance of two behaviors is equivalent to a DTW(dynamic time wrapping) distance between two paths on the shape sphere. In the same paper, the authors also show that the tangent space of shape sphere provides a valid framework for AR modeling of human behavior.

The major limitation of the above behavior representation is its high dimensionality. To overcome the so-called curse of dimensionality, many dimension reduction techniques [10] [9] have been advocated for behavioral data to yield a non-parametric representation in lower dimension.

In this paper, we combine dimension reduction and shape dynamics to achieve a plane curve representation of the behavior data. In the preprocessing of raw data, we assume a perfect segmentation of human silhouettes from the raw video data. The shape path is obtained by the silhouettes sequence on Kendall shape manifold. From the shape sequence ,a corresponding homotopy function describing the path in shape space is first constructed. To reduce the dimension of the homotopy function, we next propose an embedding by sampling the high dimensional function with the guidance of a 1D curve in its coordinate domain. The 1D curve is generated in adaption to the geometry of the homotopy, using the Nyquist frequency. Finally, with the proposed embedding, the shape space homotopy function is mapped to a plane curve function.

In contrast to the shape sequence representation in [5], we view the whole sequence as a unit which is described by a homotopy function. Consequently, our embedding is a mapping between two functional spaces, which is different from the current embedding technique in shape based human behavior analysis. For example, in [10] [9] the embedding separately maps every high dimensional point representing a shape or silhouette to a lower dimensional point.

Since our embedding is constructed in a sampling frame work, the computation of mapping and inverse mapping exploits the FFT algorithm effectively. In comparison with other lower dimensional representations, the proposed representation admits a simpler and more effective projection and reconstruction. In addition, according to Frenet theorem, the resulting plane curve is uniquely defined by curvature, which is of much lower dimension description than the path in high dimension shape space. Moreover, our representation is also more convenient for statistical analysis. For example, a pth order AR model for shape sequence have p parameters, each of which is $N \times N$ matrix (N is the number of sample points on each shape). Our representation, on the other hand, will require p scalar parameters for an AR model of the same order.

The rest of the paper is organized as follows: In Section 2, we derive a curve representation of human behavior as a dimension reduction of the behavior path in shape space; in Section 3, in combination with its AR modelling, the proposed representation is validated in a behavioral recognition problem. In Section 4, we provide some concluding remarks theoretically justifying the overall approach, along with a numerical illustration and substantiation.

2. CURVE REPRESENTATION OF HUMAN BEHAVIOR

In this section, we focus on how to construct an embedding that maps a homotopy function of a path in shape space (as shown in figure 1) to a plane curve function. Firstly, we construct a unique coordinate expression for the homotopy in shape space, as shown in figure 2. In the coordinate space (domain) of the homotopy function, a curve is generated according to Nyquist sampling frequency. Using this generated curve in the coordinate space, the function is sampled to be a plane curve. In figure 3, we give a spatial visualization of the sampling.

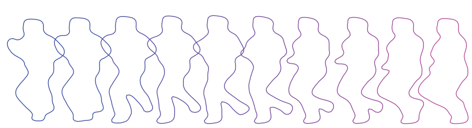


Fig. 1. the shape sequence extracted from video

2.1. Coordinate Expression of Path in Shape Space

We define a shape as $f : S^1 \to \mathbb{R}^2$. Let the shape manifold be M. $\forall f_1, f_2 \in M$, the path connecting f_1, f_2 is described by the homotopy H

$$H : (s,t) \in S^1 \times I \to (x,y) \in \Re^2, \textit{where } t \in [0,T]$$

with,

$$H(\cdot,0) = f_1, H(\cdot,T) = f_2, H(s+2n\pi,\cdot) = H(s,\cdot)$$

In practice, however, the coordinate expression of H is neither known nor accessible. It is because, for a given sequence of human silhouettes extracted from a video sequence, there exists a set of equivalent coordinate expressions with different choices of initial points as shown below,

$$\{H(s+\Delta s(t),t)\}_{\Delta s(t)}, \textit{where } \forall t, \Delta s(t) \in [0,2\pi] \quad (1)$$

In addition to not having a unique expression, another problem is that some choice of $\Delta s(t)$ may yield a highly discontinuous H. Any singularity will adversely affect our sampling of H which we will discuss in detail in the next section. In order to have a unique smooth coordinate representation for a certain shape sequence, we align shapes in different frames by optimizing the smoothness of H as follows:

To analyze the real data, we consider the time sampling of video camera. Let the time sampling rate of video camera be r_t, the down sampled version of H is noted as $H(s, nr_t)$, where $n \in [0, \lfloor \frac{T}{r_t} \rfloor]$. The shift operator $\Delta s(t)$ in equation 1 is written as,

$$\Delta s(nr_t) = ds(nr_t) + c$$

where c is a constant offset, $ds(0) = 0$. Clearly, c determine the initial point at the very first frame. According to c, function $ds(nr_t)$ describes the initial point positions of the other frames in terms of the relative difference. The reason to write $\Delta s(nr_t)$ in such form is that the constant c does not affect the smoothness of H. So we first estimate $ds(nr_t)$ by smoothness constraint on H for an fixed c . Then based on estimation of $ds(nr_t)$, c is determined by the geometric feature of $H(s + ds(nr_t) + c, t)|_{n=0}$.

- Step 1: For arbitrary c, $\forall n > 0$, $ds(n)$ is determined according to,

  $$ds(n) = \arg min_{ds(nr_t)} \parallel H(s + ds(nr_t) + c, nr_t) -$$

  $$H(s + ds((n-1)r_t) + c, (n-1)r_t) \parallel_{L_2}$$

  so that for a given constant c, $ds(nr_t)$ is sequentially estimated for $n = 1, 2, 3...$

- Step 2: c is determined as,

  $$c = \arg min_c \sum_n (H(ds(nr_t) + c, nr_t) -$$

  $$H(ds((n-1)r_t) + c, (n-1)r_t))^2$$

  with a resulting coordinate expression for H chosen to be $H(s + \Delta s(nr_t), nr_t), \Delta s(nr_t) = ds(nr_t) + c$.

In the remainder of this paper, all the H functions used throughout, should be perceived as the result of the above optimization algorithm.

2.2. Dimension Reduction

In this section, we generate a curve (s, sk) in the coordinate space (s,t) of H. Then by sampling H along (s, sk) with the proper parameter k, a high dimensional homotopy function H is mapped to a plane curve function $h(s)$ defined as:

$$h : I \to \Re^2$$

$$h(s) = (x(s, sk), y(s, sk))$$

To ease the visualization of the vector valued 2D homotopy function, H is visualized as a graph in $\mathbb{R}^3$: $(x(s,t), y(s,t), t) \in \Re^3$. Similarly, h is visualized as $(x(s, sk), y(s, sk), sk) \in \mathbb{R}^3$.

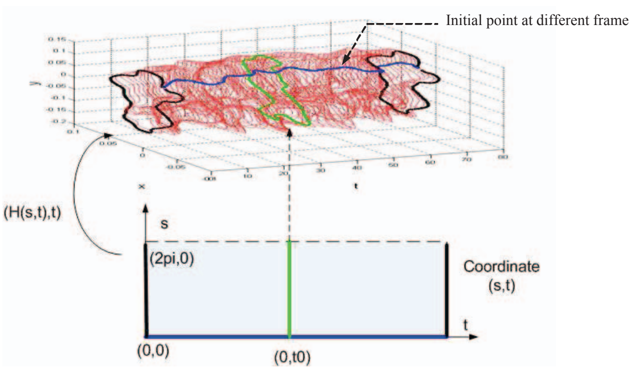


Fig. 2. Coordinate expression of Homotopy function, H is visualized as a mapping $(s,t) \in \Re_2 \to (x(s,t), y(s,t), t) \in \Re^3$. The initial point of each shape in different frames is determined using the optimization algorithm in Section 2.1

The parameter k of the proposed embedding is determined by the spectral characteristics of H. Let $\hat{H}$ be the Fourier Transform of H over t.

$$\hat{H}(s, \omega) = \int_I H(s,t) e^{j\omega t} dt$$

Then the Nyquist frequency $NF(s)$ of H for different s is computed as follows,

$$NF(s) = \inf\{\varpi : \frac{\| \int_{-\varpi}^{+\varpi} \widehat{H}(s,\omega) d\omega \|}{\| \int_{-\infty}^{+\infty} \widehat{H}(s,\omega) d\omega \|} \geq \eta\}$$

From $NF(s)$ the parameter k for curve (s, sk) is determined,

$$k = \frac{1}{\max NF(s)}$$

The form of h is then defined as a sampling of H along (s, sk),

$$h(s) = (H(s, sk), s \in [0, \lfloor \frac{T}{2\pi k} \rfloor]$$

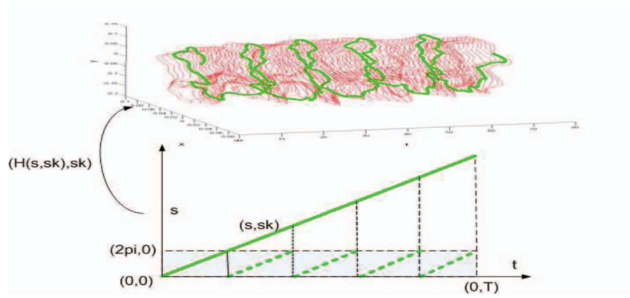


Fig. 3. sampling on $H(s,t)$ with curve $(s, sk) \in (s,t)$

In a sampling framework, there exists a natural inverse mapping from h to H as up sampling. To reconstruct the 2D function H from the 1D dimension function h, we first rewrite h as a 2D function $h(s + 2\pi n)$. In fact, we have

$$h(s + 2\pi n) = H(s, sk + 2\pi nk) \qquad (2)$$

Then for a fixed s, $H(s,t)$ can be reconstructed by up sampling $h(s + 2\pi n)$ as shown below,

$$\hat{H}(s,\omega) = \begin{cases} \sum h(s + 2\pi n) e^{j(\omega 2\pi k)n - \phi(s)}, & \omega \in [-\frac{1}{k}, \frac{1}{k}]; \\ 0, & \text{else.} \end{cases}$$

$$where, \phi(s) = (\omega 2\pi k) \cdot sk$$

$$H(s,t) = \int \hat{H}(s,\omega) e^{j\omega t} d\omega$$

Note that the phase correction $\phi(s)$ is a result of spatially shifting $H(s, sk + 2\pi nk)$ shown in Equation 2.

3. AR MODEL FOR THE EMBEDDING CURVE

In this section, the representative curve is modeled by an AR model for behavior recognition. In our experiment, we use the human behavior database provided by Irani's group as appeared in [2]. In figure 4, we illustrate three human behavior image sequences from the database. Based on the proposed behavior representation, the comparison of two behaviors is reduced to that of two plane curves. Since each of these is uniquely identified by its curvature function k up to rigid motion (translation, rotation and reflection), we proceed to construct an AR(p) model on the curvature k as,

$$k(t) = a_1 k(t-1) + a_2 k(t-2) + \cdots + a_p k(t-p) + \omega(t)$$

The parameter $a_i, i \in [1, p]$ is estimated by standard Yule-Walker equations. Let vector $A = (a_1, a_2, \cdots, a_p)$ and A_1, A_2 be the parameter vector of two behaviors. Then the distance between two behaviors is defined as the L_2 distance between A_1 and A_2 in a vector space.

$$D(A_1, A_2) = \|A_1 - A_2\|_{L_2}$$

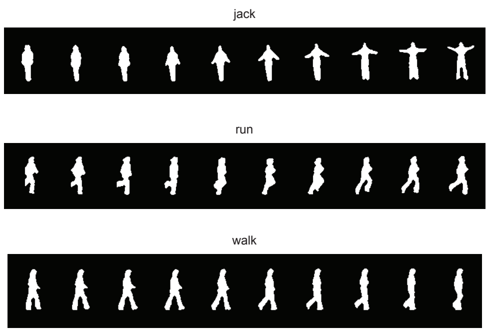
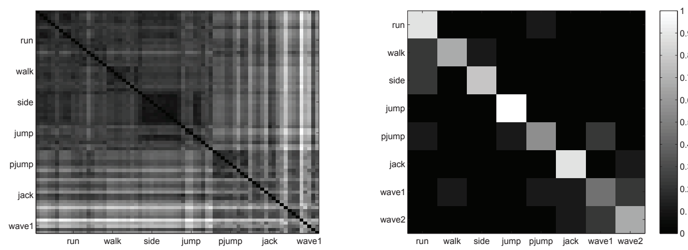


Fig. 4. Example of image sequences for behavior: run,side,skip



(a) (b)

Fig. 5. (a) Distance matrix computed with AR(30) model. (b) Probability(estimate=category B | input=category A),using the nearest neighborhood rule according to the distance matrix shown in (a)

In figure 5, we show the distance matrix of 8 categories of behaviors. Each category comprises 9 observations for 9 different people. The AR model order is set to be $p = 30$. Based on the distance matrix shown in figure $5(a)$, we perform a classification similar to the "leave-one-out" experiment in [2]. For a certain category of behavior A, we calculate probability of A being classified to category B, denoted by $P(B|A)$,using the nearest neighborhood classifier. In figure 5 (b), the result is shown as a matrix where A is a row index and B is the column index.

4. CONCLUSION

We have proposed a lower dimensional representation of human behavior as a plane curve function to greatly simplify the comparison of two behaviors. Analytically, we show that such a curve representation is the result of an embedding of a path in the high dimensional shape space. In addition, we provide an explicit reconstruction from the representative plane curve to the original shape path. The experiment results of behavior recognition illustrate that our representation captures well the features of human behavior.

5. REFERENCES

[1] E. Shechtman, M. Irani, "Space-Time Behavior Based Correlation", *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, no. 6, pp. 405-412, 2005.

[2] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, "Actions as Space-Time Shapes", *IEEE International Conf. on Computer Vision*, 2005.

[3] A. J. Yezzi, S. Soatto, "Deformotion Deforming Motion, Shape Average and the Joint Registration and Approximation of Structures in Images", *International Journal of Computer Vision*, vol. 53, no. 2, pp.153-167 , 2003.

[4] C. Liu, N. Ahuja, "A Model for Dynamic Shape and Its Applications", *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[5] A. Veeraraghavan, A.K. Roy-Chowdhury, R. Chellappa, "Matching Shape Sequences in Video with Applications in Human Movement Analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 27, no. 12, pp. 1896-1909, 2005.

[6] S.A. Niyogi, E.H. Adelson, "Analyzing and recognizing walking figures in XYT ", *IEEE Conf. on Computer Vision and Pattern Recognition*,pp. 469, 1994.

[7] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, "Behavior Recognition via Sparse Spatio-Temporal Features", *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance* ,pp. 65-72, 2005.

[8] A. Veeraraghavan, A. Chowdhury, R. Chellappa, "Role of Shape and Kinematics in Human Movement Analysis", *IEEE Conf. on Computer Vision and Pattern Recognition*,pp. 65-72, vo.l 1, pp. 730-737, 2004.

[9] C.S. Lee and A. Elgammal, "Modeling View and Posture Manifolds for Tracking", *IEEE International Conf. on Computer Vision*, 2007

[10] L. Wang, D. Suter, "Analyzing Human Movements from Silhouettes Using Manifold Learning ", *IEEE International Con. on Video and Signal Based Surveillance*, 2006