Activity Recognition from Silhouettes using Linear Systems and Model (In)validation Techniques*

Roberto Lublinerman[1] Necmiye Özay[2] Dimitrios Zarpalas[2] Octavia Camps[1,2]
[1] *Department of Computer Science and Engineering* [2] *Department of Electrical Engineering*
The Pennsylvania State University
Univeristy Park, PA 16802
{*rluble, necmiye, dzarpalas, camps*}*@psu.edu*

Abstract

In this work we propose a model (in)validation approach to gait recognition, using a system that tries to discriminate specific activities of people. The recognition process departs from an abstraction obtained from video image sequences for different activities performed by different people, by first using a suitable representation for each frame and for each frame sequence. For each frame two commonly used models for describing silhouettes are employed: Fourier Descriptors and vectors of widths. Then each sequence is modeled as a linear time invariant (LTI) system that captures the dynamics of the evolution of the frame description vectors in time. Finally a standard classification tool, SVM, is used to recognize activities using similarity measures obtained through model (in)validation. The main contribution of this work is the provision of an activity recognition model and the performance evaluation of this model using two different feature spaces.

1. Introduction

The ability to identify people only based on video image sequences (possibly taken from a distance) can have great impact in the field of Biometrics; and the ability to detect different patterns of behavior can also have several interesting applications, for example in the field of physical security. These two problems have received a significant amount of attention from the Computer Vision community in the later years. Moreover several gait sequences databases (USF Gait Challenge[1], CMU MoBo[2], UMD[3]) have been made available to provide a common test bed for researchers to advance this field of work.

2. Previous Work

Activity [9], gait [11, 1, 5] and gesture recognition are very active research areas in computer vision.

Liu and Ahuja [7] suggest using Fourier descriptors (the ones corresponding to the lowest n frequencies) to represent each contour, and a autoregressive linear model to describe the dynamics. Although they formulate the problem by using a generic p-order AR model, in their experiments they state that the AR(1) model is enough to represent the dynamics of fire. Moreover they also suggest that the dynamics for each descriptor is independent.

Bissacco et al. [2] propose the use of joint angles to abstract the body position at a certain point in time and an ARMA model to represent the dynamics of the gait. Subspace identification methods are used subsequently to determine the parameters of the model and also a distance metric is introduced for clustering and classifying the gaits.

Kale et al. [6] represent the silhouettes by using a vector of widths and use an HMM representation to account for the dynamics.

Sarkar et al. [12] propose a simple method, the Baseline algorithm and a dataset to establish a minimum model for performance evaluation. Their method involves finding the maximum correlation of the binarized silhouettes as a similarity measure and a very simple approach to determine the gait period.

Collins et al. [4] and Liu et al. [8], utilize various measures from the width and height to determine the gait period and propose the normalized crosscorrelation to perform alignment and compute the similarity score.

3. Proposed method

The organization of the paper is as follows. In section 3.1 two different methods for representing the contours will

*This work was supported in part by AFOSR under grant FA9550–05–1–0437 and NSF, under grants IIS–0117387, ECS–0221562 and ITR–0312558.

[1]http://figment.csee.usf.edu/GaitBaseline/
[2]http://hid.ri.cmu.edu/HidEval/
[3]http://degas.umiacs.umd.edu/hid

be described. Section 3.2 will explain the system that will be used to model the activities. The classification method will be presented in section 3.4 and finally the experimental results will be given in section 4.

3.1. Contour Representation

Two different contour representations are used in order to represent the contours in a suitable way for our model. In particular it is interesting to explore how sensitive is our method to different choices of representation:

- **Silhouette Width**: Given the contour of a human's image, the silhouette width is the vector that contains for each x the difference $max\{y^x\} - min\{y^x\}$, where x denotes the horizontal coordinate and y^x is the vertical one from the pair (x, y), first proposed in [4].
- **Fourier Descriptors**: Depicting (x, y) as a complex number $z = x + iy$, the Fourier coefficients of the z sequence are computed. Then the ones with the high amplitude are finally stored.

In doing so the information that the initial sequences carry is transformed in an compact way preserving important description of the activities, and the individuals. These low dimensional vectors will be used in the next step in order to model the dynamics of these activities.

3.2. Activity Modeling

Variations in the shape of the silhouette carry some distinctive information about the activity being performed by an individual. Given a time series of feature vectors representing the silhouette over a sequence of frames, temporal changes in these feature vectors can be used for gait recognition and classification purposes. To this effect, the feature vector series are modeled as the output of a causal discrete linear shift invariant system. Further, observing the quasi periodic nature of human gait, the identification and model reduction method for neutrally stable systems proposed by Sznaier et al. [13] can be applied directly.

3.2.1. Gait as Linear Time Invariant Neutrally Stable System

The sequence of feature vector trajectories $\{y_k\}$ is assumed to be the unit impulse response of an unknown causal discrete-time linear shift invariant system. It is a well-known fact that from an input-output point of view such a system can be represented with its minimal state-space realization:

$$G : \left\{ \begin{array}{rcl} x_{k+1} & = & Ax_k + Bu_k \\ y_k & = & Cx_k + Du_k \end{array} \right. \tag{1}$$

Since human gait can be regarded as a repetitive motion an additional constraint can be imposed, $A^{T_o} = I$, where T_o is the period found using the method described in section 3.2.2 and I is the identity matrix of appropriate size.

The next task is to determine the system matrices of this neutrally stable system and if possible to reduce the order of the system in order to obtain a compact representation for each model. The algorithm proposed in [13] is adopted for this purpose. For the sake of completeness a brief summary of the algorithm is presented next.

Let $\{y_k\}_{k=1}^{T_o}$ be the sequence of feature vectors over one whole period. The following block matrix is formed:

$$H_{T_o} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{T_o} \\ y_2 & y_3 & \cdots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ y_{T_o} & y_1 & \cdots & y_{T_o-1} \end{bmatrix} \tag{2}$$

Singular value decomposition is performed:

$$H_{T_o} = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ V_\perp^T \end{bmatrix} \tag{3}$$

$$S = diag(\sigma_1, \dots, \sigma_{T_o}), \sigma_i \geq \sigma_j, i \geq j \tag{4}$$

Assuming $\sigma_r > \sigma_{r+1}$ the system matrices for the reduced order model with r states (r was set to 15 for our experiments) are:

$$\mathbf{S}_r = diag(\sigma_1, \dots, \sigma_r) \tag{5}$$

$$\mathbf{A}_r = \mathbf{S}_r^{-1/2}\mathbf{U}_r^T\mathbf{P}\mathbf{U}_r\mathbf{S}_r^{1/2} \tag{6}$$

$$\mathbf{B}_r = \mathbf{S}_r^{1/2}\mathbf{V}_r^{(1)} \tag{7}$$

$$\mathbf{C}_r = \mathbf{U}_r^{(1)}\mathbf{S}_r^{1/2} \tag{8}$$

where,

$$\mathbf{P} = \begin{bmatrix} 0 & \mathbf{I}_p & 0 & \dots & 0 \\ 0 & 0 & \mathbf{I}_p & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_p & 0 & 0 & \dots & 0 \end{bmatrix} \tag{9}$$

$\mathbf{U}_r(\mathbf{V}_r)$ denotes the submatrix formed by the first r columns (rows) of $\mathbf{U}(\mathbf{V}^T)$ and where $\mathbf{U}_r^{(1)}$ and $\mathbf{V}_r^{(1)}$ denote the first $p \times r$ block of $\mathbf{U}_r$ and $r \times 1$ block of $\mathbf{V}_r^T$, respectively.

A model associated with each sample sequence is built using this algorithm. These models have some further useful properties. Every new sequence corresponding to a specific gait, can be reconstructed by driving the system representing that gait with an appropriate input. Finding the input is an infinite dimensional search. However, due to periodicity the effect at frame k of a signal applied at $k - s$ is the same as the effect of the same signal applied at $k - s - (m * T_o)$, where m is a positive integer. As shown in [13], this allows to assume that there is no input for $k \geq 0$. Under this assumption, the problem reduces to finding an initial condition x_o for which

$$\begin{array}{rcl} x_{k+1} & = & Ax_k \\ y_k & = & Cx_k \end{array} \tag{10}$$

is satisfied. The above equation is equivalent to $Y = \mathcal{O}x_o$, where

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T_o-1} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{T_o-1} \end{bmatrix} \qquad (11)$$

Since a minimal realization is observable, the observability matrix $\mathcal{O}$ has full column rank by construction and least square solution to above equation is given by $x_o = \left(\mathcal{O}^T\mathcal{O}\right)^{-1}\mathcal{O}^T\mathbf{Y}$.

3.2.2. Determining the Period of each Gait Sequence

One of the key problems in identification of a system with periodic impulse response is to determine the period T_o. This problem can be solved using the Hankel rank approach described by Sznaier and Camps [14]. A naive probabilistic approach was employed that gives sufficient results for the specific case under exploration.

The period of an activity should be the least common multiple of the periods of individual features (i.e. an entry in the feature vector). Each feature is considered as a one dimensional discrete signal and the Fourier Transform is calculated for each of these signals. A new vector is produced by summing the normalized Fourier coefficients of all these signals for each frequency, to get the frequency which is dominant in all of the features. When Fourier descriptors are used to represent the silhouette, only the features that correspond to low frequency coefficients are taken into account. Silhouette widths seem to be more reliable to determine the frequency since they are less sensitive to noise of the silhouette contour and preserve periodicity. However experiments show that gait periods are not affected in a meaningful way by our choice of descriptors.

3.3. Computing Similarity Measures using Model (In)validation

Once the nominal dynamic models associated with each activity are obtained, recognition can be recasted as a robust model (in)validation problem [10, 13]. Here, instead of finding a nominal model for each activity all models built using the training set are considered as a sample for the activity they are associated with and a classifier is used to determine the class membership of any testing sample. The statement given in the end of section 3.2.1 leads to the setup in figure 1. The minimum size uncertainty block is used as a measure of dissimilarity between a training sample and the new sequence. For the system in 1, there exists an operator $\Delta \in \mathcal{BL}(l_2)(\gamma)$ such that $\mathbf{v} = \Delta\mathbf{z}$ if and only if $\|v\|_2^2 \leq \gamma^2 \|z\|_2^2$. With straight forward calculations the following inequality is derived:

$$\mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathcal{O}_N\mathbf{x}_o - \mathbf{x}_o^T\mathcal{O}_N^T\mathbf{Y} + (1-\gamma^2)\mathbf{x}_o^T\mathcal{O}_N^T\mathcal{O}_N\mathbf{x}_o \leq 0 \qquad (12)$$

which can be converted to an LMI optimization problem (see Chapter 2 of [3]) in x_o and $\alpha = (1-\gamma^2)^{-1}$:

$$\min \gamma \text{ subject to} \\ \begin{bmatrix} \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathcal{O}_N\mathbf{x}_0 - (\mathcal{O}_N\mathbf{x}_0)^T\mathbf{Y} & (\mathcal{O}_N\mathbf{x}_0)^T \\ \mathcal{O}_N\mathbf{x}_0 & (\gamma^2-1)^{-1} \end{bmatrix} \leq 0 \qquad (13)$$

where $\mathcal{O}_N^T = [\mathbf{C}^T \ \ldots \ (\mathbf{CA}^{N-1})^T]$. For computational reasons in this paper instead of solving the LMI, we solve for the minimum γ with the least square initial condition $x_o = \left(\mathcal{O}_N^T\mathcal{O}_N\right)^{-1}\mathcal{O}_N^T\mathbf{Y}$.

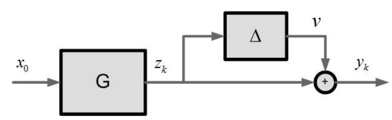

Figure 1. System's Block Diagram

3.4. Classifying Activities

The features for representing each gait in this work are the LTI models obtained using the formulation in subsection 3.2.1. These models lie in a very high dimensional space defined by parameters of the LTI model (Equation 1), i.e. the A, B, C and D matrices. Classification in this space using roughly the components of these matrices as feature vectors and any distance measure based on a matrix or vector norm definition does not seem meaningful, especially considering that different matrices can be realizations of the same LTI system (some authors, e.g. [9], use this approach). Instead we propose, as noted previously, computing the minimum size block uncertainty as a similarity measure to derive a RBF kernel and use kernel based classification methods.

For our experiments in particular we chose SVM [15] using the kernel matrix K to compute the maximum margin classifier in the induced space. The kernel matrix K used for classification in this work is defined as follows

$$K(i,j) = e^{-\frac{(\gamma_{i,j}+\gamma_{j,i})}{2}\frac{1}{2\sigma^2}} \qquad (14)$$

where the average between minimum block uncertainty (γ) when trying to fit series i to model j and the converse to obtain one similarity measure from two one-way relations. The matrix K obtained in this way is a Gram Matrix. σ, the standard scale parameter, was chosen appropriately.

4. Experimental Results

For experimental validation of our proposed method the MoBo dataset is used. From this dataset side view outline samples for the four activities (slow walk, fast walk, walk on incline and walk with ball) performed several times for twenty five different subjects were used. The experiments performed on this dataset consist on two multi class classification and five two class classification. The multiclass experiments are (1) classify the activity performed among all the activities available and (2) discriminate between walking (fast or slow), incline walking and walking with a ball. The two-class experiments are: (1) identify incline walking, (2) identify walking with ball, (3) identify fast walking, (4) identify slow walking, (5) identify walking (fast or slow).

All the experiments were run for the two choices of frame representation, Fourier Descriptors and Silhouette Widths, using thirty fold cross validation. The mean error and standard deviation across the thirty fold runs for these experiments are depicted in Table 1.

	Descriptors		Widths	
	Mean Error	Error Deviation	Mean Error	Error Deviation
All four activities	0.289	0.057	0.256	0.05
walk vs. incline vs. ball	0.156	0.038	0.044	0.021
incline vs. rest	0.033	0.019	0	0
ball vs. rest	0.178	0.038	0.067	0.025
fast walk vs. rest	0.211	0.044	0.256	0.05
slow walk vs. rest	0.2	0.044	0.244	0.045
fast or slow walk vs. rest	0.178	0.041	0.022	0.015

Table 1. Classification results

Table 1 shows that the choice for shape representation plays an important role in activity recognition. It is interesting to note that some activities seem to be easier classified than others, e.g. slow walking and fast walking are the most difficult activities to tell apart, but walking as a single activity is easier to tell apart from the rest. Judging from the results obtained our model captures the similarity of the same activities as performed by different people, considering fast and slow walk a single activity (walk vs. incline vs. ball line in table 1). The difference in performance for the choices of representation can be attributed to sensitivity to noise for contour extraction.

5. Conclusions and Future Work

These preliminary experimental results show that identification using gait information extracted only from silhouette is possible without having to rely on the correct segmentation of specific marker point that are required for, for example, methods based on the determination of joint angles. At the same time they raise attention to several important issues about the influence in the choice of representation for individual frames and for the frame sequence that represents a gait instance.

On the other hand having good gait image database is crucial in advancing this specific area and much work has been done in building them, although we feel that for this specific problem there is a need for comprehensive databases of real-world sequences.

The model proposed could also be suitable for discriminating between people and thus use gait as a biometric. This topic is being explored as a continuation of this work.

Acknowledgments

We would like to thank Dr. Mario Sznaier for sharing his ideas and enriching us through motivating discussions.

References

[1] C. B. Abdelkader. Motion-based recognition of people in eigengait space. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 267, Washington, DC, USA, 2002. IEEE Computer Society.

[2] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *CVPR (2)*, pages 52–57. IEEE Computer Society, 2001.

[3] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Studies in Applied Mathematics 15, 1994.

[4] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 366, Washington, DC, USA, 2002. IEEE Computer Society.

[5] A. M. Elgammal and C.-S. Lee. Separating style and content on a nonlinear manifold. In *CVPR (1)*, pages 478–485, 2004.

[6] A. A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. R. Chowdhury, V. Krüger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004.

[7] C.-B. Liu and N. Ahuja. A model for dynamic shape and its applications. In *CVPR (2)*, pages 129–134, 2004.

[8] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *Proceedings of the 7th European Conference on Computer Vision (ECCV'02)*, May 2002.

[9] C. Lu and N. J. Ferrier. Repetitive motion analysis: Segmentation and event classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):258–263, 2004.

[10] C. Mazzaro, M. Sznaier, O. I. Camps, S. Soatto, and A. Bissacco. A model (in)validation approach to gait recognition. In *3DPVT*, pages 700–703. IEEE Computer Society, 2002.

[11] S. D. Mowbray and M. S. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. *Lecture Notes in Computer Science*, 2688:566–573, 2003.

[12] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2):162–177, 2005.

[13] M. Sznaier, O. Camps, and M. C. Mazzaro. Finite horizon model reduction of a class of neutrally stable systems with applications to texture synthesis and recognition. In *IEEE CDC '04*, pages 3068–3073.

[14] M. Sznaier and O. I. Camps. A hankel operator approach to texture modelling and inpainting. In *IEEE Texture 2005 in conjunction with ICCV*, 2005.

[15] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.