Individual Recognition by Kinematic-based Gait Analysis

Bir Bhanu and Ju Han
Center for Research in Intelligent Systems
University of California, Riverside, California 92521, USA
{bhanu,jhan}@cris.ucr.edu

Abstract

Current gait recognition approaches only consider individuals walking frontoparallel to the image plane. This makes them inapplicable for recognizing individuals walking from different angles with respect to the image plane. In this paper, we propose a kinematic-based approach to recognize individuals by gait. The proposed approach estimates 3D human walking parameters by performing a least squares fit of the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. A genetic algorithm is used for feature selection from the estimated parameters, and the individuals are then recognized from the feature vectors using a nearest neighbor method. Experimental results show that the proposed approach achieves good performance in recognizing individuals walking from different angles with respect to the image plane.

1. Introduction

Current human recognition methods, such as fingerprints, face or iris biometrics, generally require a cooperative subject, views from certain aspects and physical contact or close proximity. These methods can not reliably recognize non-cooperating individuals at a distance in real-world changing environmental conditions. Moreover, in various applications of personal identification, many established biometrics can be obscured. Gait, which concerns recognizing individuals by the way they walk, has been used as a important biometric without the above-mentioned disadvantages.

In recent years, some approaches have already been employed in automatic gait recognition. Niyogi and Adelson [6] make an initial attempt in a spatiotemporal (XYT) volume. They first find the bounding contours of the walker, and then fit a simplified stick model on them. A characteristic gait pattern in XYT is generated from the model parameters for recognition. Little and Boyd [4] propose a model-free approach making no attempt to recover a structural model of human motion. Instead they describe the shape of the motion with a set of features derived from moments of a dense flow distribution. Similarly, He and Debrunner's [1] approach detects a sequence of feature vectors based on Hu's moments of motion segmentation in each frame, and the individual is recognized from the feature vector sequence using hidden Markov models. To avoid feature extraction process which may reduce the reliability, Murase and Sakai [5] propose a template matching method to calculate the spatio-temporal correlation in a parametric eigenspace representation for gait recognition. Huang et al. [2] extend this approach by combining canonical space transformation (CST) based on canonical analysis, with eigenspace transformation (EST) for feature selection.

However, existing gait recognition approaches only consider individuals walking frontoparallel to the image plane. In this paper, we propose a kinematic-based approach to recognize individuals by gait. The proposed approach estimates 3D human walking parameters by performing a least squares fit of the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. Our approach eliminates the assumption of individuals walking frontoparallel to the image plane, which is desirable in many gait recognition applications.

2. 3D Human Modeling

2.1. Human Kinematic Model

A human body is considered as an articulated object, consisting of a number of body parts. The body model adopted here is shown in Figure 1, where a circle represents a joint and a rectangle represent a body part (N: neck, S: shoulder, E: elbow, W: waist, H: hip, K: knee, and A: ankle).

Most joints and body part ends can be represented as spheres, and most body parts can be represented as cones. The whole human kinematic model is represented as a set of cones connected by spheres [3]. Figure 2 shows that most of the body parts can be approximated well in this manner. However, the head is approximated only crudely by a sphere and the torso is approximated by a cylinder with two spheroid ends.

2.2. Matching between 3D Model and 2D Silhouette

The matching procedure determines a parameter vector **x** so that the proposed 3D model fits the given

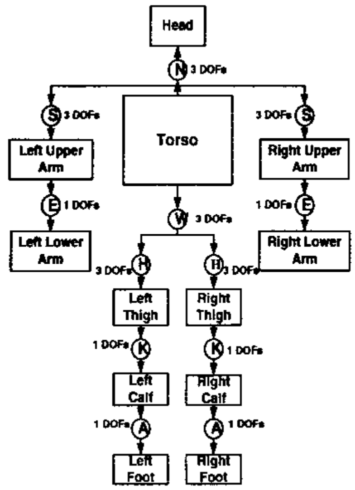

Figure 1. 3D Human Kinematic Model.

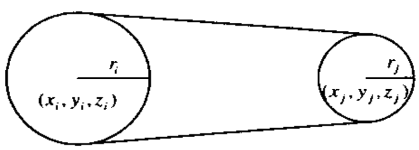

Figure 2. Body part geometric representation.

2D silhouette as well as possible. For that purpose, two chained transformations transform human body local coordinates (x, y, z) into image coordinates (x', y')[7]: the first transformation transforms local coordinates into camera coordinates; while the second transformation projects camera coordinates into image coordinates.

Each 3D human body part is modeled by a cone with two spheres $\mathbf{s}_i$ and $\mathbf{s}_j$ at its ends, as shown in Figure 2 [3]. Each sphere $\mathbf{s}_i$ is fully defined by 4 scalar values, (x_i, y_i, z_i, r_i), which define its location and size. Given these values for two spheroid ends (x_i, y_i, z_i, r_i) and (x_j, y_j, z_j, r_j) of a 3D human body part model, its projection $P_{(ij)}$ onto the image plane is the convex hull of the two circles defined by (x'_i, y'_i, r'_i) and (x'_j, y'_j, r'_j).

If the 2D human silhouette is known, we may find the relative 3D body parts locations and orientations with the knowledge of camera parameters. We propose a method to perform a least squares fit of the 3D human model to the 2D human silhouette. That is, to estimate the set of sphere parameters $\mathbf{x} = \{\mathbf{x}_i : (x_i, y_i, z_i, r_i)\}$ by choosing $\mathbf{x}$ to minimize

$$error(\mathbf{x}; I) = \sum_{x', y' \in I} (P_{\mathbf{x}}(x', y') - I(x', y'))^2, \qquad (1)$$

where I is the silhouette binary image, $P_{\mathbf{x}}$ is the binary projection of the 3D human model to image plane, and x', y' are image plane coordinates.

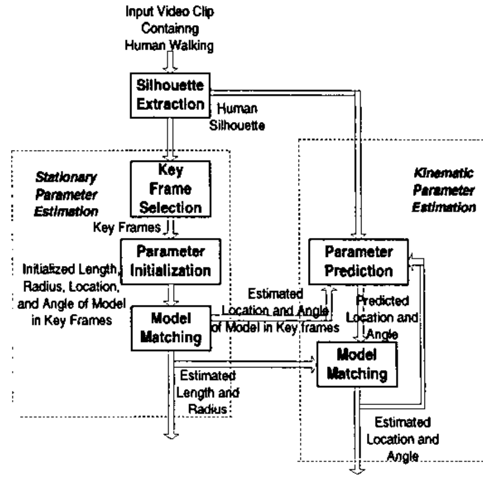

Figure 3. Diagram of the proposed approach for human gait analysis.

2.3. Model Parameter Selection

Human motion is very complex due to so many DOFs. To simplify the matching procedure, we use the following reasonable assumptions: (1) the camera is stationary; (2) people are walking before the camera at a distance; (3) people are moving in a constant direction; (4) the swing direction of arms and legs parallels to the moving direction.

According to all the above-mentioned assumptions, we do not need to consider the waist joint, and only need to consider 1 DOF for each other joint. Therefore, the elements of the parameter vector of the 3D human kinematic model is defined as follows:

- Stationary parameters – Radius r_i (11): torso(3), shoulder, elbow, hand, hip, knee, ankle, toe, and head; Length l_i (9): torso, inter-shoulder, inter-hip, upper arm, forearm, thigh, calf, foot, and neck.
- Kinematic parameters – Location (x, y) (2); Angle θ_i (11): neck, left upperarm, left forearm, right upperarm, right forearm, left thigh, left calf, left foot, right thigh, right calf, and right foot.

With these 33 parameters, the projection of the human model can be completely determined.

3. Model Parameter Estimation

The realization of our proposed approach is shown in Figure 3, and the major processing steps are detailed in the following subsections.

3.1. Silhouette Extraction

Assuming that people are the only moving objects in the scene, they can be extracted by a simple background subtraction method. Notice that an area cast into shadow often results in a significant change in intensity without much change in chromaticity. Given a video sequence containing moving people and the corresponding background image, for each frame I_i in the sequence, the color value difference $\Delta \mathbf{p}_i(x,y) = \|\mathbf{p}_i(x,y) - \mathbf{p}_b(x,y)\|$ is computed for each pixel, where $\mathbf{p}_i(x,y)$ and $\mathbf{p}_b(x,y)$ are RGB color values of the pixel at (x,y) in the ith frame and background image, respectively. The chromaticity is computed as

$$\begin{aligned} r_c(x,y) &= r(x,y)/(r(x,y)+g(x,y)+b(x,y)) \\ g_c(x,y) &= g(x,y)/(r(x,y)+g(x,y)+b(x,y)). \end{aligned}$$

We have

$$\begin{aligned} \Delta r_{ci}(x,y) &= |r_{ci}(x,y) - r_{cb}(x,y)| \\ \Delta g_{ci}(x,y) &= |g_{ci}(x,y) - g_{cb}(x,y)|. \end{aligned}$$

Given thresholds t_1 and t_2, if

$$(\Delta \mathbf{p}_i(x,y) > t_1) \wedge ((\Delta r_{ci}(x,y) > t_2) \vee (\Delta g_{ci}(x,y) > t_2))$$

the pixel at (x,y) is determined to be part of the moving objects; otherwise, it is part of the background.

After the silhouette has been cleaned by a preprocessing procedure, its height, width and centroid can be easily extracted for motion analysis. In addition, the moving direction of the walking person is determined as follows

$$\theta = \begin{cases} tan^{-1}\frac{f(h_1-h_N)}{h_1 y_N - h_N y_1}, & \text{if } y_1 > y_N; \\ tan^{-1}\frac{f(h_1-h_N)}{h_1 y_N - h_N y_1} + \pi, & \text{otherwise.} \end{cases} \tag{2}$$

where f is the camera focus length, y_1 and y_N are the horizontal centroid of the silhouette in the first and Nth frame, and h_1 and h_N are the height of the silhouette in the first and Nth frame.

3.2. Stationary Parameter Estimation

The stationary parameters include body part length parameters and joint radius parameters. Notice that human walking is a cyclic motion, so a video sequence can be divided into motion cycles and studied separately. In each walking cycle, the silhouette with minimum width means that people stand straight in that frame and that means the most occlusion; the silhouette with maximum width means the least occlusion and, therefore, it is more reliable.

To estimate the stationary parameters, we first select some key frames (4 frames in our experiments) which contain more reliable silhouettes, and then perform matching procedure on the key frames as a whole. The corresponding feature vector thus includes 20 common stationary parameters and 13*4 individual kinematic parameters. Next, we first initialize these parameters according to the human statistical information. Then, the set of parameters is estimated from this initial parameters by choosing a parameter vector $\mathbf{x}$ to minimize the least square error in equation (1) with respect to the same kinematic constraints.

After the matching algorithm is converged, the estimated stationary parameters are obtained and will be used for kinematic parameter estimation of other frames. At the same time, the estimated kinematic parameters of key frames will be used for prediction.

3.3. Kinematic Parameter Estimation

To reduce the search space and make our matching algorithm converge faster, we use the predicted parameters from the previous frames as the initialization of the current frame:

$$\begin{aligned} \theta^{(i)} &= \theta^{(i-1)} + (\theta^{(i-1)} - \theta^{(i-2)}) \\ y^{(i)} &= y^{(i-1)} + (y^{(i-1)} - y^{(i-2)}) \\ x^{(i)} &= x^{(i-1)} \end{aligned} \tag{3}$$

After the matching algorithm is converged, the estimated kinematic parameters are obtained for each frame.

4. Feature Selection and Gait Recognition

In this paper, we only use angle features for gait recognition, including: (1) average head angle, (2) average leading upperarm angle, (3) average lagging upperarm angle, (4) average leading forearm angle, (5) average lagging forearm angle, (6) average leading thigh angle, (7) average lagging thigh angle, (8) average leading calf angle, and (9) average lagging calf angle. In our approach, foot angles are not considered here because the foot is too short to be well matched in our approach; while average angles are used because they are less sensitive to noise than maximum or minimum angles. The distance between feature vectors is measured in a weighted Euclidean space $D(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n} w_i(x_i - y_i)^2}$, where n is the dimension of each feature vector. In this equation, the weight w_i indicates the importance of the ith feature for gait recognition. Considering these weights depend on data in the given gait database, we aquire them by genetic algorithm whose fitness function is defined as the retrieval accuracy in the training dataset.

5. Experimental Results

The video data used in our experiment are real human walking data recorded in outdoor environment. In these video data, there is only one walking person at the same time. Six different people walk along different directions (within $[-\pi/4, \pi/4]$ along the image plane). The size of image frames is 180×240. In our experiments, we first manually divide video data into single-cycle sequences, and then select 15 sequences from each person: 10 sequences for training and 5 sequences for testing. Figure 4 shows some sample sequences in our

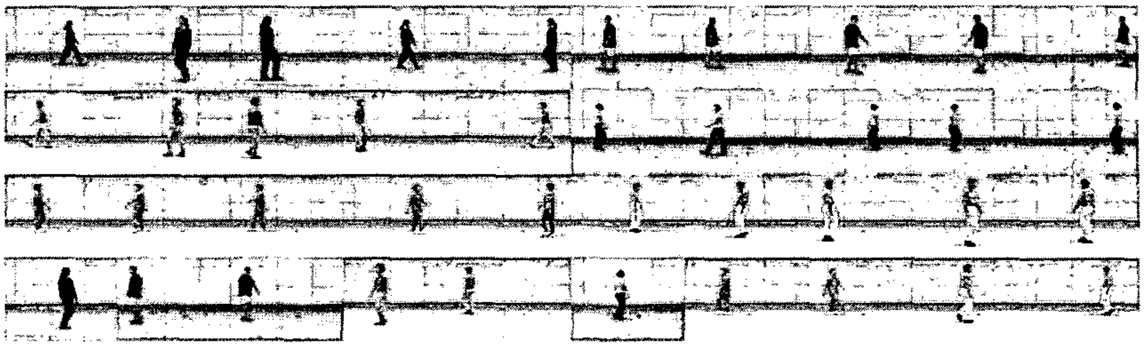

Figure 4. Sample sequences in our gait database.

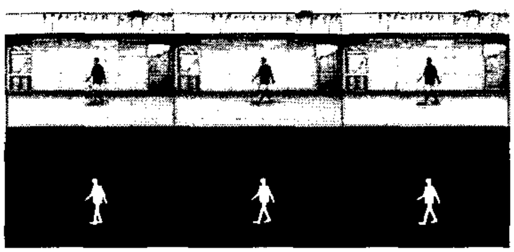

Figure 5. An example of matching results.

gait database (The first 3 rows from training dataset and the last row from testing dataset).

The proposed approach was implemented in Matlab. The least square matching is done using the *constr* function in the Optimization Toolbox. This function uses a Sequential Quadratic Programming (SQP) method. Figure 5 shows an example of matching results.

In our experiments, we use Genetic algorithm to obtain the weights of the features on the training dataset. The weight vector so obtained is $[3, 2, 2, 2, 1, 3, 0, 3, 0]$ which means that lagging thigh and calf angles play an unimportant role for recognition in the training set. Using these weights, our approach achieves 88% recognition on the training dataset using a 5 nearest neighbor (5-NN) Leave-One-Out method (7 errors out of 60 training sequences). The performance on the testing data is 77% recognition (7 errors out of 30 testing sequences). After examining the experimental results, we find that most of the errors occur primarily due to silhouette segmentation errors. The matching algorithm during the stationary parameter estimation phase can converge by applying constraints on all parameters. For kinematic parameter estimation, an inaccurate segmentation error causes errors. We are investigating various error recovering method to overcome this problem. Walking at different speed does not affect the performance since we can change the sampling rate from the original video.

6. Conclusions

In this paper, we proposed an approach to estimate 3D human motion from a monocular image sequence for automatic gait recognition. The proposed approach performs a least squares fit of a complex 3D human model to the 2D human silhouette. To reduce the search space and make our matching algorithm converge faster, the proposed approach also includes a prediction procedure. Experimental results show that the proposed approach achieves good performance using automatic gait recognition to recognize individuals walking from different angles with respect to the image plane.

Acknowledgment

This work was supported in part by grants F49620-97-1-0184 and DAAD19-01-0357; the contents and information do not necessarily reflect the position or policy of U.S. Government.

References

[1] Q. He and C. Debrunner. Individual recognition from periodic activity using hidden markov models. in *Proc. IEEE Workshop on Human Motion*, pages 47–52, 2000.

[2] P. Huang, C. Harris, and M. Nixon. Recognizing humans by gait via parameteric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[3] M. Lin. Tracking articulated objects in real-time range image sequences. in *Proc. International Conference on Computer Vision*, pages 648–653, 1999.

[4] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):1–32, 1998.

[5] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17(2):155–62, 1996.

[6] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[7] S. Wachter and H.-H. Nagel. Tracking of persons in monocular image sequences. in *Proc. IEEE Workshop on Nonrigid and Articulated Motion*, pages 2-9, 1997.