Human Recognition on Combining Kinematic and Stationary Features

Bir Bhanu and Ju Han

Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA
{bhanu,jhan}@cris.ucr.edu

Abstract. Both the human motion characteristics and body part measurement are important cues for human recognition at a distance. The former can be viewed as kinematic measurement while the latter is stationary measurement. In this paper, we propose a kinematic-based approach to extract both kinematic and stationary features for human recognition. The proposed approach first estimates 3D human walking parameters by fitting the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. Kinematic and stationary features are then extracted from the kinematic and stationary parameters, respectively, and used for human recognition separately. Next, we discuss different strategies for combining kinematic and stationary features to make a decision. Experimental results show a comparison of these combination strategies and demonstrate the improvement in performance for human recognition.

1 Introduction

In many applications of personnel identification, established biometrics, such as fingerprints, face or iris, may be obscured. Gait, which concerns recognizing individuals by the way they walk, can be used as a biometric to recognize people under these situations. However, most existing gait recognition approaches [1–4] only consider human walking frontoparallel to the image plane. In this paper, we propose a kinematic-based approach to recognize human by gait which relaxes this condition. The proposed approach estimates 3D human walking parameters by fitting the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. Since both the human motion characteristics and body part measurement are important cues for human recognition at a distance, kinematic and stationary features are extracted from the estimated parameters, and used for human recognition separately. Moreover, we combine the classifiers based on stationary and kinematic features to increase the accuracy of human recognition. Experimental results show a comparison of different combination strategies and demonstrate the improvement in performance for human recognition.

2 Technical Approach

In our approach, we first build a 3D human kinematic model for regular human walking. The model parameters are then estimated by fitting the 3D human

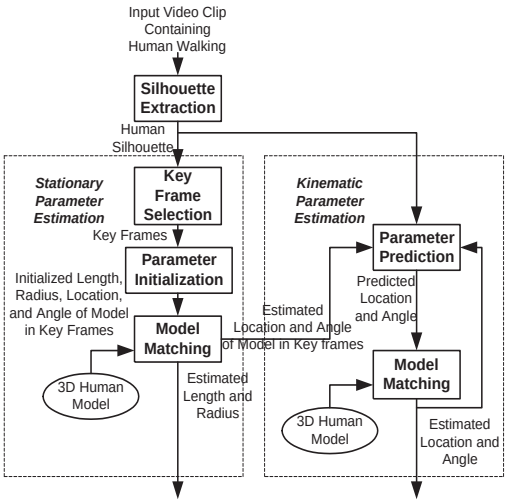

Fig. 1. Diagram of the proposed approach for human gait analysis.

kinematic model to the extracted 2D human silhouette. Finally, stationary and kinematic parameters are extracted from these parameters for human recognition. The realization of our proposed approach is shown in Figure 1.

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 2(a), 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 [5]. Figure 2(b) shows that 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.

Matching between 3D Model and 2D Silhouette: The matching procedure determines a parameter vector $\mathcal{X}$ so that the proposed 3D model fits the given 2D silhouette as well as possible. 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(b) [5]. 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 prior knowledge of camera parameters. We propose a method to perform a least squares fit of the 3D human model

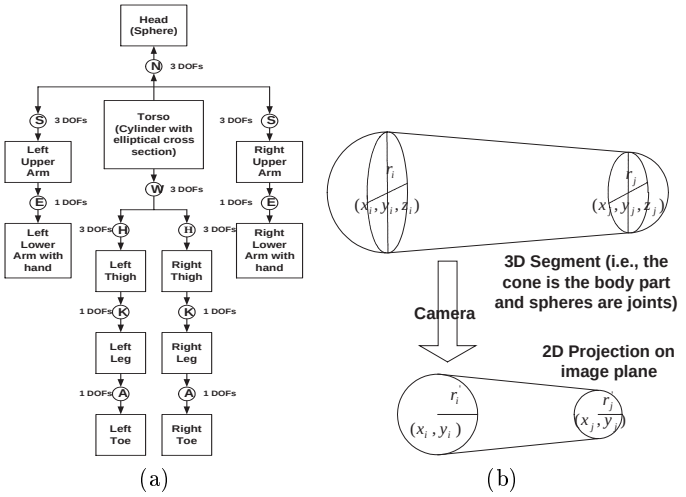

Fig. 2. (a) 3D Human Kinematic Model; (b) Body part geometric representation.

to the 2D human silhouette. That is, to estimate the set of sphere parameters $\mathcal{X} = \{\mathcal{X}_i : (x_i, y_i, z_i, r_i)\}$ by choosing $\mathcal{X}$ to minimize

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

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

Model Parameter Selection: Human motion is very complex due to so many degrees of freedom (DOFs). To simplify the parameter estimation 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 these assumptions, we do not need to consider the waist joint, and only need to consider one DOF for each other joint. Therefore, the elements of the parameter vector $\mathcal{X}$ of the 3D human kinematic model are defined as: (a) 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; (b) Location $(x, y)(2)$; Angle $\theta_i(11)$: neck, left upper arm, left forearm, right upper arm, right forearm, left thigh, left calf, left foot, right thigh, right calf, and right foot. With 33 stationary and kinematic parameters, the projection of the human model can be completely determined.

2.2 Model Parameter Estimation

Assuming that people are the only moving objects in the scene, their silhouette can be extracted by a simple background subtraction method [7]. After the silhouette has been cleaned by a pre-processing procedure, its height, width and centroid are easily extracted for motion analysis. The human moving direction is estimated through the silhouette width variation in the video sequence [7].

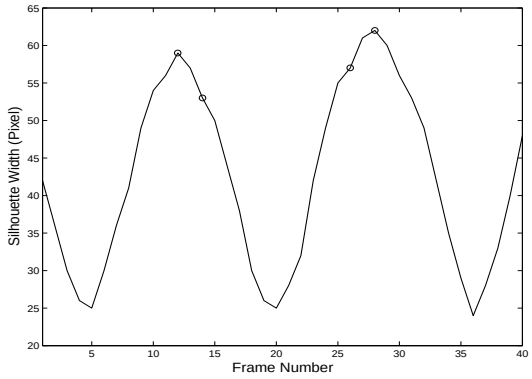

Fig. 3. Human silhouette width variation in a video sequence (Circles represent frames selected as key frames for stationary parameter estimation).

Stationary Parameter Estimation: The stationary parameters include body part length and joint radius. Human walking is a cyclic motion, so a video sequence can be divided into motion cycles and studied separately. The walking cycle can be detected by exploiting the silhouette width variation in a sequence as shown in Figure 3. In each walking cycle, the silhouette with minimum width means the most occlusion; the silhouette with maximum width means the least occlusion and is more reliable for model parameter estimation.

To estimate the stationary parameters, we first select 4 key frames (see Figure 3) from one walking cycle, and then perform matching procedure on these frames as a whole because the human silhouette from a single frame might not be reliable due to noise. The corresponding feature vector thus includes 20 common stationary parameters and 13*4 individual kinematic parameters. Then, the set of parameters is estimated from these initial parameters by choosing a parameter vector $\mathcal{X}$ to minimize the least square error in equation (1) with respect to the same kinematic constraints. The parameters are initialized according to the human statistical information. After the matching algorithm is converged, the estimated stationary parameters are obtained.

Kinematic Parameter Estimation: To reduce the search space and make our matching algorithm converge faster, we use the linear prediction of parameters from the previous frames as the initialization of the current frame. After the matching algorithm is converged, the estimated kinematic parameters are obtained for each frame.

2.3 Kinematic and Stationary Feature Classifiers

In our approach, kinematic features are the mean and standard deviation values extracted from the kinematic parameters of each frame in the whole image sequence containing one human walking cycle. Assuming that human walking is symmetric, that is, the motion of the left body part is the same as or similar to the right body parts, the kinematic feature vector $\mathbf{x_k}$ selected for human recog-

nition includes 10 elements: the mean and standard deviation of angles of neck, upper arm, forearm, thigh, and leg.

Stationary features are directly selected from the estimated stationary parameters of each sequence containing human walking. Among those model stationary parameters, joint radius depends on human clothing, and inter-shoulder and inter-hip length is hardly estimated due to the camera view (human walking within small angle along the frontparallel direction). Assuming the body part length is symmetric for left and right body parts, the stationary feature vector $\mathbf{x_s}$ selected for human recognition includes 7 elements: neck length, torso length, upper arm length, forearm length, thigh length, calf length, and foot length.

After the kinematic and stationary features are extracted, they are used to classify different people separately. For simplicity, we assuming the feature vector $\mathbf{x}$ ($\mathbf{x}$ could be $\mathbf{x_s}$ or $\mathbf{x_k}$) for a person ω_i is normally distributed in the feature space, and each of the independent features have Gaussian distribution with the same standard deviation value. Under this assumption, minimum distance classifier is established: $\mathbf{x}$ is assigned to the class whose mean vector has the smallest Euclidean distance with respect to $\mathbf{x}$.

2.4 Classifier Combination Strategies

To increase the efficiency and accuracy of human recognition, we need to combine the two classifiers in some way. Kittler et al. [8] demonstrate that the commonly used classifier combination schemes can be derived from a uniform Baysian framework under different assumptions and using different approximations. We use these derived strategies to combine the two classifiers in our experiments.

In our human recognition problem with M people in the database, two classifiers with feature vector $\mathbf{x_s}$ and $\mathbf{x_k}$, respectively, are combined to make a decision on assigning each sample to one of the M people $(\omega_1, ..., \omega_M)$. The feature space distribution of each class ω_i is modeled by the probability density function $p(\mathbf{x_s}|\omega_i)$ and $p(\mathbf{x_k}|\omega_i)$, and its a priori probability of occurrence is $P(\omega_i)$. Under the assumption of equal priors, the classifier combination strategies are described as follows:

- Product rule
  $\{\mathbf{x_s}, \mathbf{x_k}\} \in \omega_i$, if $p(\mathbf{x_s}|\omega_i)p(\mathbf{x_k}|\omega_i) = \max_{k=1}^{M} p(\mathbf{x_s}|\omega_k)p(\mathbf{x_k}|\omega_k)$
- Sum rule
  $\{\mathbf{x_s}, \mathbf{x_k}\} \in \omega_i$, if $p(\mathbf{x_s}|\omega_i) + p(\mathbf{x_k}|\omega_i) = \max_{k=1}^{M} (p(\mathbf{x_s}|\omega_k) + p(\mathbf{x_k}|\omega_k))$
- Max rule
  $\{\mathbf{x_s}, \mathbf{x_k}\} \in \omega_i$, if $\max\{p(\mathbf{x_s}|\omega_i), p(\mathbf{x_k}|\omega_i)\} = \max_{k=1}^{M} \max\{p(\mathbf{x_s}|\omega_k), p(\mathbf{x_k}|\omega_k)\}$
- Min rule
  $\{\mathbf{x_s}, \mathbf{x_k}\} \in \omega_i$, if $\min\{p(\mathbf{x_s}|\omega_i), p(\mathbf{x_k}|\omega_i)\} = \max_{k=1}^{M} \min\{p(\mathbf{x_s}|\omega_k), p(\mathbf{x_k}|\omega_k)\}$

In our application, the estimate of a posteriori probability is computed as follows:

$$P(\omega_i|\mathbf{x}) = \frac{\exp\{-||\mathbf{x} - \mu_\mathbf{i}||^2\}}{\sum_{k=1}^{M} \exp\{-||\mathbf{x} - \mu_\mathbf{k}||^2\}}, \tag{2}$$

where $\mathbf{x}$ is the input of the classifier, and $\mu_\mathbf{i}$ is the ith class center.

Fig. 4. Sample human walking sequences in our database.

3 Experimental Results

The video data used in our experiment are real human walking data recorded in outdoor environment. Eight different people walk within [-45°,45°] with respect to frontparallel direction. We manually divide video data into single-cycle sequences with an average of 16 frames. In each sequence, only one person walks along the same direction. There are a total of 110 single-cycle sequences in our database, and the number of sequences per person ranges from 11 to 16. The image size is 180×240.Figure 4 shows some sample sequences in our database.

We use Genetic algorithm for model parameter estimation. Each of the extracted kinematic and stationary features is normalized by $\frac{x-\mu}{\sigma}$, where x is the specific feature value, μ and σ are the mean and standard deviation of the specific feature over the entire database. Recognition results in our experiments are obtained using Leave-One-Out method.

Feature Size	Stationary Features	Recognition Rate
1	neck	31%
2	neck, torso	32%
3	neck, torso, upper arm	45%
4	neck, torso, upper arm, forearm	50%
5	neck, torso, upper arm, forearm, thigh	55%
6	neck, torso, upper arm, forearm, thigh, calf	59%
7	neck, torso, upper arm, forearm, thigh, calf, foot	62%

Table 1. Comparison of performance using different number of stationary features.

Performance of Stationary Feature Classifier: The recognition rate with all the 7 stationary features is 62%. Table 1 shows the human recognition performance using different number of stationary features. From this table, we can see that the recognition rate increases when feature number increases. Therefore, each of these features has its own contribution to the overall recognition performance using stationary features. On the other hand, the contribution varies among different features. For example, adding torso length into the feature vector with neck length makes 1% improvement, while adding upper arm length into the feature vector with torso and neck length makes 13% improvement. As a result, better recognition performance might be achieved by using weighted

Euclidean distance instead of regular Euclidean distance. This requires a training procedure. However, due to the high feature space dimension (7) and small class number (8) in the database, overfitting becomes a big problem under this situation, i.e., training results achieve high recognition rate on training data and low recognition rate on testing data. Therefore, we do not carry out weight training in this paper. We expect such a procedure to be carried out when a large database with a large number of classes (people) becomes available.

Feature Size	Kinematic Features	Recognition Rate
5	Mean	50%
5	Standard Deviation	49%
10	Mean and Standard Deviation	72%

Table 2. Comparison of performance using mean and standard deviation features computed from each body part angle variation sequences over a single-cycle sequence.

Feature Size	Kinematic Features	Recognition Rate
2	neck	34%
4	neck, upper arm	51%
6	neck, upper arm, forearm	57%
8	neck, upper arm, forearm, thigh	63%
10	neck, upper arm, forearm, thigh, leg	72%

Table 3. Comparison of performance using different number of kinematic features.

Performance of Kinematic Feature Classifier: The recognition rate with all the 10 kinematic features is 72%. In Table 2, it is shown that the mean and standard deviation features computed from each body part angle variation sequences over a single-cycle sequence achieve similar recognition rate, 50% and 49%, respectively. Table 3 shows the human recognition performance using different number of kinematic features. Similar to stationary features, the recognition rate increases when feature number increases. We also expect a weight training procedure carried out on a large human walking database in the future.

Combination Rule	Recognition Rate
Product Rule	83%
Sum Rule	80%
Max Rule	73%
Min Rule	75%

Table 4. Comparison of performance using different combination strategies.

Performance with Classifier Combination: Table 4 shows the human recognition performance on classifier combination with different strategies. Considering the recognition rate on stationary and kinematic classifiers are 62% and 72%, respectively, all the four rules achieve better recognition performance on human recognition. Among the combination strategies, product rule achieves the best recognition rate of 83%. Sum rule also achieves a better recognition rate of 80%. The recognition rates achieved by max and min rules are only slightly better than that of kinematic classifier (72%). Sum rule has been mathematically proved to be robust to errors by Kittler et al. [8]. We believe that the main reason for the good performance achieved by product rule is the holding of the

conditional independence assumption (the features used in different classifiers are conditionally statistically independent) in product rule for our application. The poor performance of max and min rules may come from their order statistics and sequential sensitivity to noise. Similar results are found in Shakhnarovich and Darrell's work on combining face and gait features [9].

4 Conclusions

In this paper, we propose a kinematic-based approach for human recognition. The proposed approach estimates 3D human walking parameters by fitting the kinematic model to the 2D silhouette extracted from a monocular image sequence. The kinematic and stationary features are extracted from the estimated parameters, and used for human recognition separately. Next, we use different strategies to combine the two classifiers to increase the accuracy of human recognition. Experimental results show that our proposed approach achieves the highest 83% recognition rate by using product rule on combining classifiers of stationary features and kinematic features. Note that this performance is achieved under the situation of people walking from $-45°$ to $45°$ with respect to the frontparallel direction, and the low resolution of human walking sequences. With higher resolution human walking sequences and a weight training procedure for weighted Euclidean distance, we expect a better recognition performance in our future work.

Acknowledgment

This work was supported in part by grants F49620-97-1-0184, F49620-02-1-0315 and DAAD19-01-0357; the contents and information do not necessarily reflect the position or policy of U.S. Government.

References

1. S.A. Niyogi, E.H. Adelson. Analyzing and recognizing walking figures in XYT. in *Proc. IEEE Conference on CVPR*, pp. 469-474, 1994.
2. J.J. Little, J.E. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):469-474, 1998.
3. H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17(2):155-62, 1996.
4. P.S. Huang and C.J. Harris and M.S. Nixon. Recognizing humans by gait via parameteric canonical space. *Artificial Intelligence in Engineering*, 13:359-366, 1999.
5. M.H. Lin. Tracking articulated objects in real-time range image sequences. in *Proc. ICCV*, pp. 648-653, 1999.
6. S. Wachter and H.-H. Nagel. Tracking of persons in monocular image sequences. in *Proc. IEEE Workshop on Nonrigid and Articulated Motion*, pp. 2-9, 1997.
7. B. Bhanu and J. Han. Individual recognition by kinematic-based gait analysis. in *Proc. International Conference on Pattern Recognition*, (3):343-346, 2002.
8. J. Kittler, M. Hatef, R. Duin, and J. Matas. On Combining Classifiers. *IEEE Trans. PAMI*, 20(3):226-239, 2001.
9. G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. in *Proc. IEEE International Conferenceon Automatic Face and Gesture Recognition*, pp. 169-174, 2002.