Fusion of Static and Dynamic Body Biometrics for Gait Recognition

Liang Wang, Huazhong Ning, Tieniu Tan, Weiming Hu
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080
{lwang,hzning,tnt,wmhu}@nlpr.ia.ac.cn

Abstract

Human identification at a distance has recently gained growing interest from computer vision researchers. This paper aims to propose a visual recognition algorithm based upon fusion of static and dynamic body biometrics. For each sequence involving a walking figure, pose changes of the segmented moving silhouettes are represented as an associated sequence of complex vector configurations, and are then analyzed using the Procrustes shape analysis method to obtain a compact appearance representation, called static information of body. Also, a model-based approach is presented under a Condensation framework to track the walker and to recover joint-angle trajectories of lower limbs, called dynamic information of gait. Both static and dynamic cues are respectively used for recognition using the nearest exemplar classifier. They are also effectively fused on decision level using different combination rules to improve the performance of both identification and verification. Experimental results on a dataset including 20 subjects demonstrate the validity of the proposed algorithm.

1. Introduction

Vision-based human identification has recently attracted much attention, e.g., the Human ID program of DARPA [20]. This strong interest is driven by a variety of potential applications such as visual surveillance, covert security and access control. As a newly emergent biometric feature, gait has the advantage of being non-invasive, and it is probably the only perceivable modality from a great distance. Gait recognition aims essentially to discriminate individuals by the way they walk.

Current gait recognition approaches may be explicitly classified into two major categories, namely model-based methods [2,4,10] and motion-based methods [7,8,11,12]. As model-based examples, Johnson and Bobick [4] used activity-specific static body parameters for gait recognition without directly analyzing gait dynamics, and Yam *et al*. [10] tried the running action of gait to recognize people as well as walking and explored the relationship between walking and running that was expressed as a mapping based on the idea of phase modulation. Most existing approaches are motion-based. BenAbdelkader *et al*. [7] used image self-similarity plots of a moving person to recognize people, and Phillips *et al*. [11] described a silhouette correlation based algorithm for the gait identification problem. These approaches further provide clear supports for the view that it is feasible to recognize people by gait.

For obtaining optimal performance, an automatic person identification system should integrate as many informative cues as available. There are various properties of gait that might serve as recognition features. We categorize them as static features and dynamic features. The former usually reflects geometry-based measurements such as body-height, stride and build, while the latter means joint-angle trajectories of lower limbs. Intuitively, recognizing people by gait depends greatly on how the static silhouette shape changes over time. So previous work on gait recognition mainly adopted low-level information such as silhouette [7,8,11,12]. Due to the difficulties of parameter recovery from video, few methods except [2,10] used higher-level information, e.g., temporal features of joint angles reflecting the dynamics of gait motion sufficiently. Based on the idea that body biometrics includes both the appearance of human body and the dynamics of gait motion measured during walking, here we attempt to fuse the two completely different sources of information available from walking video for personal recognition.

The proposed method is shown in Figure 1. For each image sequence, background subtraction is used to extract moving silhouettes of the walker. Static pose changes of these silhouettes over time are represented as an associated sequence of complex vector configurations in a common coordinate, and are then analyzed using the Procrustes shape analysis method to obtain an eigen-shape for reflecting the body appearance, i.e., static information. Also, a model-based approach under a Condensation framework together with human body model, motion model and constraints is presented to track the walker in

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

image sequences. From the tracking results, we can calculate joint-angle trajectories of main lower limbs, i.e., dynamics of gait. Both static and dynamic information may be independently used for recognition using the nearest exemplar pattern classifier. They are also combined on decision level to improve the final performance.

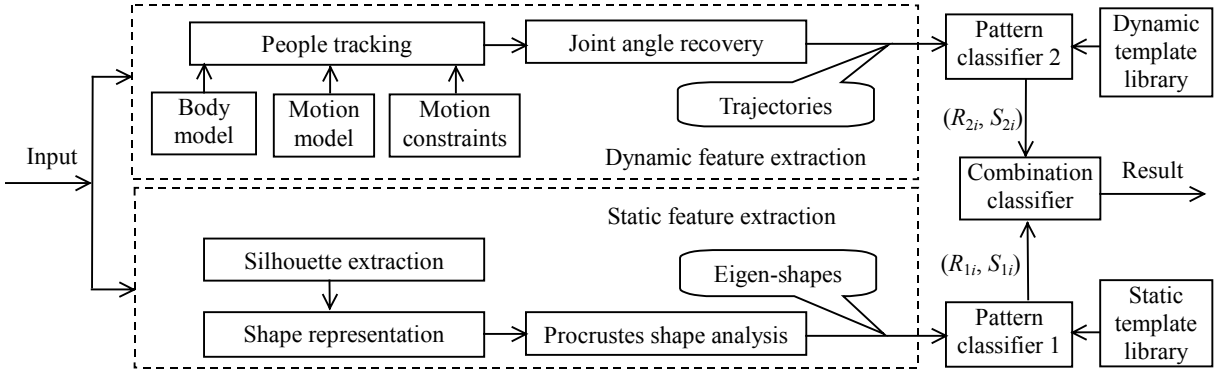


Figure 1. Overview of the proposed algorithm

2. Static feature extraction

2.1. Silhouette extraction and representation

A background subtraction procedure [19] is used to extract a single-connectivity moving region of the walker in each image. An important cue in determining underlying motion of a walking figure is his or her silhouette shape changes over time. For the sake of reducing redundancy, we only need to analyze spatial contours. The boundary can be obtained using a border following algorithm based on connectivity. Then, we compute its shape centroid (x_c, y_c). Let the centroid be the origin of a 2D shape space. We can unwrap the boundary as a set of pixel points (x_i, y_i) along outer-contour anticlockwise in a complex coordinate. That is, each shape can be described as a vector consisting of complex numbers with N_b boundary elements $\boldsymbol{z}=[z_1, z_2, \ldots, z_i, \ldots z_{Nb}]^T$, where $z_i=x_i+j*y_i$. Therefore, each gait sequence will be transformed into a sequence of such 2D shape configurations accordingly.

2.2. Procrustes shape analysis

We need one method that allows us to compare a set of static pose shapes in gait pattern and is robust to position, scale and slight rotation changes. A mathematically elegant way for aligning point sets is Procrustes shape analysis. A good brief review can be found in [1].

Procrustes shape analysis is intended to cope with 2D shapes. A shape in 2D space can be described by a vector of k complex numbers $\boldsymbol{z}=[z_1, z_2, \ldots, z_k]^T$, called a configuration. It is convenient to center shapes by defining the centered configuration $\boldsymbol{u}=[u_1, u_2, \ldots, u_k]^T, u_i = z_i - \bar{z}$, $\bar{z} = \sum_{i=1}^{k} z_i / k$. The full Procrustes distance between two configurations can be defined as

$$d_F(\boldsymbol{u}_1, \boldsymbol{u}_2) = 1 - \frac{\left|\boldsymbol{u}_1^* \boldsymbol{u}_2\right|^2}{\left\|\boldsymbol{u}_1\right\|^2 \left\|\boldsymbol{u}_2\right\|^2} \quad (1)$$

where the superscript * represents the complex conjugation transpose. Given a set of n shapes, we can find their mean by finding $\boldsymbol{u}$ that minimizes the objective function

$$\min_{\alpha_j, \beta_j} \sum_{j=1}^{n} \left\| \boldsymbol{u} - \alpha_j \boldsymbol{1}_k - \beta_j \boldsymbol{u}_j \right\|^2 \quad (2)$$

To find $\boldsymbol{u}$, we compute the following matrix

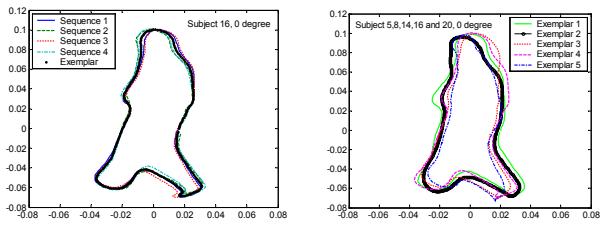
$$\boldsymbol{S}_u = \sum_{i=1}^{n} (\boldsymbol{u}_i \boldsymbol{u}_i^*) / (\boldsymbol{u}_i^* \boldsymbol{u}_i) \quad (3)$$

The Procrustes mean shape $\hat{\boldsymbol{u}}$ is the dominant eigenvector of $\boldsymbol{S}_u$, i.e., the eigenvector that corresponds to the greatest eigenvalue of $\boldsymbol{S}_u$ [1].

2.3. Static signature acquisition

Our approach uses these single shape representations from a gait sequence to find their mean shape as static signatures that can represent the appearance of body shape. The following summarizes the major steps in determining the Procrustes mean shape of a gait pattern.

1. Select a set of k points from the boundary to represent a 2D shape as a vector configuration $\boldsymbol{z}_j$ as discussed in Section 2.1;
2. Set the centered configuration. When we represent the silhouette shape, we use the shape centroid as the origin of 2D shape space to move all shapes to a common center. So we can directly set $\boldsymbol{u}_j=\boldsymbol{z}_j$, $j=1, 2, \ldots, n$;
3. Compute the matrix $\boldsymbol{S}_u$ using Eqn. (3). Then, compute the eigenvalues and the associated eigenvectors of $\boldsymbol{S}_u$;
4. Set the Procrustes mean shape $\hat{\boldsymbol{u}}$ as the eigenvector that corresponds to the maximum eigenvalue, and this mean shape is used as static signatures.



(a) (b)

Figure 2. Plots of mean shapes and the exemplars

For multiple mean shapes from multiple sequences of the same subject, we may acquire an exemplar by averaging them as a static template for that class to avoid selecting a random reference sample. Figure 2 (a) shows plots of mean shapes of four sequences of a subject and their exemplar, and Figure 2 (b) shows plots of multiple exemplars from different subjects. From Figure 2 we can see that the intra-subject changes in eigenshapes are very small, while the inter-subject changes are more significant. Such result implies that the mean shapes have considerable discriminating power. Further details on static feature extraction may be found in [19].

3. Dynamic feature extraction

For extracting dynamic features of gait motion, we present a new model-based approach to tracking the walking figure under the Condensation framework [6].

3.1. Prior knowledge

Our model knowledge includes three parts: human body model, motion model and motion constraints [18]. The human body model used in this paper, similar to [5], is composed of 14 rigid body parts, each of which is represented by a truncated cone except for the head represented by a sphere. With the constraint that people are walking parallel to the image plane, the state space can be naturally described by a 12-dimensional vector $P=\{x, y, \theta_1, \theta_2, \cdots, \theta_{10}\}$, where (*x, y*) is the global position of human body and θ_i (*i*=1~10) is joint angles. Our motion model is empirically described with Gaussian functions $G_{k,t}\left(u_{k,t}, {\sigma_{k,t}}^2\right)$ for each joint *k* (*k* = 1…10) at any phase *t* (*t* = 1…*T*) in the walking cycle. We also use conditional distributions to model the motion constraints of the dependencies of neighboring joint angles.

3.2. Tracking

Tracking is equivalent to relate the image data to the pose vector P . Since the articulated body model may be naturally formulated as a tree-like structure, a hierarchical estimation, i.e., locating the global position and tracking each limb separately, is suitable here.

Given the above considerations, we first predict the global position from the centroid of the detected moving human and then refine it by searching the neighborhood of the predicted position. Each limb is tracked under the Condensation framework that uses learnt dynamical models, together with visual observations, to propagate the random sample set.

The dynamic model needs to be designed carefully to improve the efficiency of factored sampling. Here, the learnt motion model serving as prior is integrated into the dynamic model. With the assumption that the Gaussian distributions at different phases in the motion model are independent, at time instant *t* the *i*th motion parameter $\theta_{i,t}$ satisfies the dynamic model

$$p(\theta_{i,t} \mid \theta_{i,t-1}) = G\left(\alpha u_{i,t} + \beta u_{i,t-1} + \gamma \theta_{i,t-1},\ \lambda\left((\alpha \sigma_{i,t})^2 + (\beta \sigma_{i,t-1})^2\right)\right)$$

where $\alpha+\beta+\gamma=1$ makes the drifting of $\theta_{i,t}$ not only from the tracking history $\theta_{i,t-1}$ but also from the motion model, and λ is a scalar that is often set to 1. This dynamic model is generally sufficient for all motion parameters, but motion constraints can further concentrate the samples for parameters of elbow, knee and ankle joints. For instance, after the shoulder joint $\theta_{s,t}$ is sampled, sample positions generated from the conditional distribution $p(\theta_{e,t} \mid \theta_{s,t})$ for the elbow joint $\theta_{e,t}$ also contain much information. So a mixed-state Condensation [17] can be included in the factored sampling scheme by choosing with a probability *q* to generate samples from the dynamic model and with a probability 1-*q* to generate samples from the conditional distribution, i.e., $\theta_{e,t}$ satisfies

$$p(\theta_{e,t} \mid \theta_{e,t-1}, \theta_{s,t}) = qG\left(\alpha u_{e,t} + \beta u_{e,t-1} + \gamma \theta_{e,t-1},\ \lambda\left((\alpha \sigma_{i,t})^2 + (\beta \sigma_{i,t-1})^2\right)\right) + (1-q)\, p(\theta_{e,t} \mid \theta_{s,t})$$

where $\alpha, \beta, \gamma, \lambda$ are defined as above.

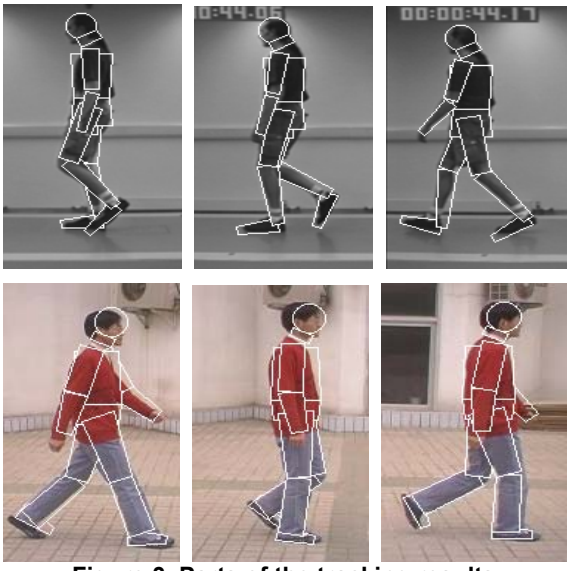


Figure 3. Parts of the tracking results

The PEF (Pose Evaluation Function) reveals the observation density $p(z_t \mid x_t)$ of an image z_t given that the human model has the posture x_t at time t. In general, boundary information improves the localization, whereas region information stabilizes the tracking. Therefore, we combine both of them in the PEF by computing the boundary matching error E_b and the region matching error E_r to achieve both accuracy and robustness. Here the tracking results of 2 sequences are showed in Figure 3. Due to space constraint, only the human areas clipped from the original images are given. More details on tracking may be found in [18].

3.3. Dynamic signature acquisition

Estimating an underlying skeleton from the tracking results enables us to measure joint-angle trajectories. Figure 4 shows signals of four joints: left and right hips, left and right knees from a walking instance, where the smoothed curves are the results after the median filtering. It is variations in the joint signals that we wish to consider as dynamic information of body biometrics, i.e., dynamics of gait.

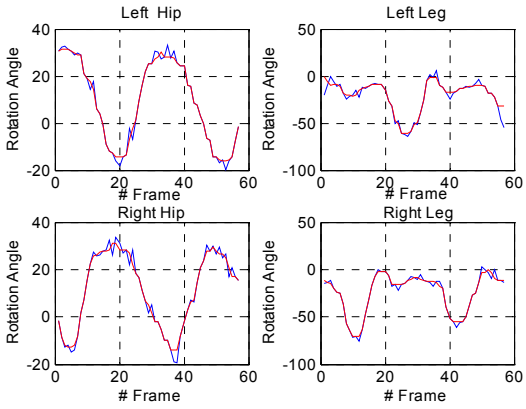
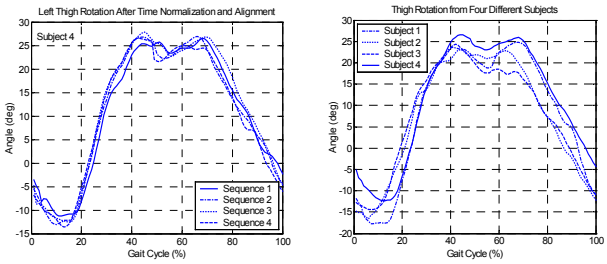


Figure 4. Joint-angle trajectories of lower limbs

Differences in body structure and dynamics naturally cause joint-angle trajectories to vary in both magnitude and time. To analyze these signals for identity recognition, we need to normalize them. We select only one walking cycle from each sequence. Without directly using the joint-angle trajectories, we carry out variance normalization by subtracting the mean of each signal and then dividing by the estimated standard deviation to reduce the effect of noise. DTW (Dynamic Time Warping) is applied to temporally align the signals to a fixed reference phase. Figure 5 shows the results of time-normalized signals of thigh rotation, from which we find that there are little variations among sequences from the same subject, whereas there are apparent variations among different subjects. We choose four normalized signals from left and right hips and knees to constitute a dynamic feature vector. Similarly, we also use multiple vectors from the same subject to obtain the exemplar by averaging them, which is regarded as a dynamic template for that class.



(a) (b)

Figure 5. Time-normalized signals of joint angles

4. Pattern classifiers and fusion rules

Gait recognition is a traditional pattern classification problem. Here we try the nearest neighbor classifier with class exemplar (*ENN*). No doubt, a more sophisticated classifier could be employed, but the interest here is to evaluate the discriminatory ability of the extracted features.

To measure similarity, we use both the Procrustes mean shape distance defined in Eqn. (1) for static features and the Euclidean distance for dynamic features respectively. The smaller the above distance measures are, the more similar the two gaits are.

Main reasons for combining classifiers are efficiency and accuracy. A variety of fusion approaches for biometric recognition are available, a few of which are mentioned here [13-16]. Here, we investigate several different approaches to classifier combination.

It should be noted that having obtained the score for each modality given the observation, one generally cannot directly combine these scores in a statistically meaningful way because they are not direct estimates of the posterior, but rather measures of the distance between the test example and the reference example [3]. These scores, with quite different ranges and distributions, must therefore be transformed to be comparable before fusion (the logistic function $e^{(\alpha+\beta x)}/(1+e^{(\alpha+\beta x)})$ is used in this paper).

First, we respectively investigate rank-summation-based and score-summation-based approaches described in [16]. Following the theoretical framework presented in [15], we also compare the max, min, mean, and product rules for combining classifier outputs.

To statistically justify the above rules, a monotonic transformation function over scores S needs to be applied to reflect the posterior probability. We use the similar approach proposed in [3]. That is, we may estimate a probability distribution over the scores assigned to the correct labels by a mapping function T from scores to the empirical distribution and treat $T(S)$ as the estimate of the posterior.

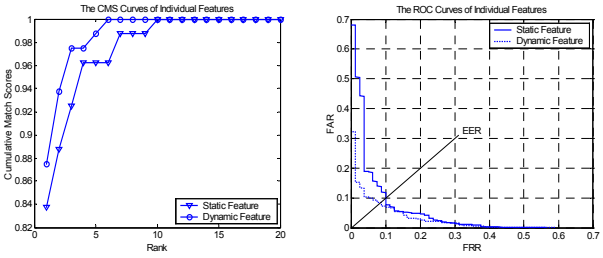
5. Experiments

We collected 80 sequences from 20 subjects and four sequences per subject for our experiments. Each sequence includes a walking figure and the walker moves naturally in the field of view without occlusion laterally with respect to the image plane. All image sequences are captured by a stationary digital camera at a rate of 25 frames per second.

5.1. Experimental results

For each image sequence, we first extract static features in the manner described in Section 2. Further, we perform model-based tracking and recover dynamic features in the manner described in Section 3. It should be noted that self-occlusion of body parts, shadow under the feet, the arm and the torso having the same color, and low quality of the images all bring challenges to our tracking method. For a small portion of failed tracking sequences, we manually obtain the motion parameters as the focus of this paper is not on tracking per se but on gait recognition using the tracking data as dynamic features.

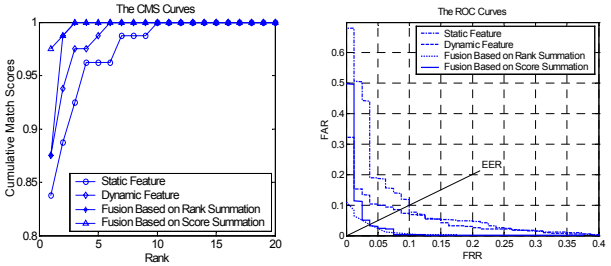
Due to a small number of examples, we hope to compute an unbiased estimate of the true recognition rate using a leave-one-out cross-validation method. That is, we first leave one example out, train on the rest, and then classify or verify the omitted element according to its differences with respect to the rest examples.



(a) Identification (b) Verification

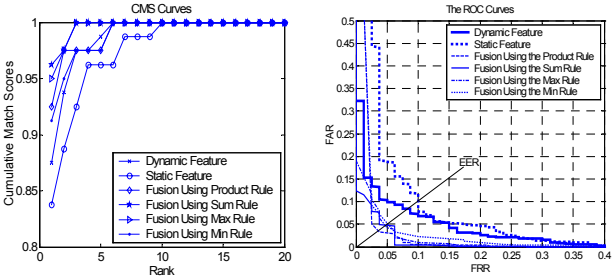
Figure 6. The results using a single modality

First, we separately use static features and dynamic features obtained from walking video for recognition. In identification mode, the classifier determines which class a given measurement belongs to. One useful measure of classification performance that is more general than classification error is CMS (Cumulative Match Scores) [9] which is firstly introduced in the FETET protocol for the evaluations of face recognition algorithms. It indicates the probability that the correct match is included in the top n matches. For completeness, we also use the ROC (Receiver Operating Characteristic) curves to report verification results. In verification mode, the classifier is asked to verify whether a new measurement really belongs to certain claimed class. ROC curves give plots of various pairs of FAR (False Acceptance Rate) and FRR (False Rejection Rate) under different decision threshold values for the acceptance. Figure 6 (a) and (b) respectively show performance of identification (for ranks up to 20) and verification using a single modality. It should be mentioned that the CCR (Correct Classification Rate) is equivalent to $p_{(1)}$ (i.e., rank=1).



(a) Identification (b) Verification

Figure 7. Fusion results of rank and score based summation rules



(a) Identification (b) Verification

Figure 8. Fusion results using the product, sum, max and min combination rules

Based on the combination rules described in Section 4, we examine the results after fusing both static features and dynamic features. Figure 7 (a) and (b) show the results of identification and verification using rank-summation-based and score-summation-based combination rules respectively, and Figure 8 (a) and (b) give the results of identification and verification using the product, sum, max and min combination rules respectively. For comparison, we also plot the results using a single modality in Figure 7 and Figure 8.

5.2. Analysis of results

From Figure 6, we can see that there is indeed identity information in both static and dynamic features derived from walking video that can be explored for the recognition task. The results using dynamic information are somewhat better than those using static information. This is likely due to the fact that the dynamics reflect more essential information of gait motion.

Figure 7 and Figure 8 demonstrate the improved performance of both identification and verification for the integration step than that using any single modality. A summary of CCRs and EERs (Equal Error Rate) is given in Table 1 for clarity. Another interesting observation from

the comparative results is that the score-summation-based rule outperforms other combinations schemes as a whole. Of the last 4 statistical combination rules, the sum rule is the best for identification, which has also been shown in [15] using the sensitivity analysis to demonstrate that the sum rule is the most resilient to estimation errors. However, the product rule is best for verification. The main reason for the poor performance of the min rule is probably because that it suffers more from the noise in score assignment than the relatively robust mean and product rules. Also, it is believed that there will be better results if there are sufficient data to precisely model the probability distributions of scores for the two pattern classifiers. In all, these studies highlight the importance of a careful choice of the whole combination strategy.

Table 1. Summary of CCRs and EERs

	CCR (rank=1)	CCR (rank=3)	EER
Static features	83.75%	92.50%	10.0%
Dynamic features	87.50%	97.50%	8.42%
Rank-summation	87.50%	100%	3.75%
Score-summation	97.50%	100%	3.75%
Product	92.50%	97.50%	3.54%
Sum	96.25%	100%	5.00%
Max	95.00%	100%	4.70%
Min	91.25%	97.50%	5.00%

Although as a whole the results are very encouraging, more experiments on a larger and more realistic database still need to be investigated in future work in order to be more conclusive.

6. Conclusions

This paper has proposed a method based on fusion of static and dynamic body biometrics for gait recognition. A statistical approach based on Procrustes shape analysis is used to obtain a compact representation of the appearance of body shape from spatiotemporal pattern of walking. A model-based approach is employed to track the walker and to recover joint-angle trajectories of lower limbs that reflect the dynamics of gait. Both static and dynamic cues of body biometrics may be independently used for recognition. Also, they have been combined on decision level for improving the performance. Experimental results have demonstrated the feasibility of the proposed method.

Acknowledgements

This work is supported by NSFC (Grant 69825105 and 60105002), Natural Science Foundation of Beijing (Grant 4031004), the National 863 High-Tech R&D Program of China (Grant 2002AA117010), and Institute of Automation, the Chinese Academy of Sciences.

References

[1] J. Boyd, "Video Phase-locked Loops in Gait Recognition", *Proc.* of *Intl. Conf. on Computer Vision*, I: 696-703, 2001.

[2] R. Tanawongsuwan and A. Bobick, "Gait Recognition from Time-normalized Joint-angle Trajectories in the Walking Plane", *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition,* 2001.

[3] G. Shakhnarovich, L. Lee, and T. Darrell, "On Probabilistic Combination of Face and Gait Cues for Identification", *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition,* pp. 176-181, 2002.

[4] A. Bobick and A. Johnson, "Gait Recognition Using Static, Activity-specific Parameters", *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition,* 2001.

[5] S. Wachter and H. Nagel, "Tracking Persons in Monocular Image Sequences", *CVIU*, 74(3): 174-192, 1999.

[6] M. Isard and A. Blake, "Condensation – Conditional Density Propagation for Visual Tracking", *IJCV*, 29(1): 5-28, 1998.

[7] C. BenAbdelkader, R. Culter, H. Nanda, and L. Davis, "EigenGait: Motion-based Recognition of People Using Image Self-similarity", *Proc. of Intl. Conf. on Audio- and Video-based Person Authentication,* pp. 284-294, 2001.

[8] R. Collins, R. Gross, and J. Shi, "Silhouette-based Human Identification from Body Shape and Gait", *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 366-371, 2002.

[9] J. Phillips, H. Moon, S. Rizvi, and P. Rause, "The FERET Evaluation Methodology for Face Recognition Algorithms", *PAMI*, 22(10): 1090-1104, 2000.

[10] C. Y. Yam, M. S. Nixon, and J. N. Carter, "On the Relationship of Human Walking and Running: Automatic Person Identification by Gait", *Proc. of Intl. Conf. on Pattern Recognition*, 2002.

[11] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm", *Proc. of Intl. Conf. on Pattern Recognition*, 2002.

[12] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification", *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 155-162, 2002.

[13] R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues", *PAMI*, 17(10): 955–966, 1995.

[14] L. Hong and A. Jain, "Integrating Faces and Fingerprints for Personal Identification", *PAMI*, 20(12): 1295–1307, 1998.

[15] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On Combining Classifiers", *PAMI*, 20(3): 226–239, 1998.

[16] B. Achermann and H. Bunke, "Combination of Classifiers on the Decision Level for Face Recognition", *Technical Report IAM-96-002*, University Bern, 1996.

[17] M. Isard and A. Blake, "A Mixed-state Condensation Tracker with Automatic Modal Switching", *Proc.* of *Intl. Conf. on Computer Vision*, pp. 107-112, 1998.

[18] H. Ning, L. Wang, W. Hu, and T. Tan, "Articulated Model Based People Tracking Using Motion Models", *Proc. of Intl. Conf. on Multi-modal Interface*, pp. 383-388, 2002.

[19] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic Gait Recognition Based on Statistical Shape Analysis", *IEEE Trans. Image Processing*, August 2003 (to appear).

[20] Available: http://www.darpa.mil/iao/HID.htm.