Efficient Model Training for HMM-based Person Identification by Gait

Muhammad Rasyid Aqmar, Koichi Shinoda and Sadaoki Furui
Department of Computer Science Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku Tokyo 152-8552 Japan
E-mail: rasyid@ks.cs.titech.ac.jp, {shinoda,furui}@cs.titech.ac.jp

***Abstract*—In gait-based person identification, statistical methods such as hidden Markov models (HMMs) have been proved to be effective. Their performance often degrades, however, when the amount of training data for each walker is insufficient. In this paper, we propose walker adaptation and walker adaptive training, where the data from the other walkers are effectively utilized in the model training. In walker adaptation, maximum likelihood linear regression (MLLR) is used to transform the parameters of the walker-independent model to those of the target walker model. In walker adaptive training, we effectively exclude the *inter-walker* variability from the walker-independent model. In our evaluation, our methods improved the identification performance even when the amount of data was extremely small.**

I. Introduction

Human gait refers to the motion of an individual characterized by his/her spatio-temporal movement while walking. Automatic person identification using human gait has been extensively studied. For example, binary silhouette [1], gait energy image (GEI) [2], higher-order shape configuration [3], and higher-order local autocorrelation [5] have been used as gait features. Similarity matching (SM) [1], *k*-nearest neighbors (*k*-NN) [5], dynamic time warping (DTW) [6], and hidden Markov models (HMMs) [4], [7], [8], [9] have been often used as gait classifiers. Statistical methods such as HMMs often achieved higher performance than the others since they are more robust against the variety in gait than the others.

In our previous work [9], we proposed an HMM-based method robust against speed difference, and confirmed its effectiveness. We used cubic higher-order local autocorrelation coefficients mapped to Fisher discriminant space (CHLAC+FDA) [5] and gait-silhouette principal component (GSP) as features. These features discriminate walkers and distinguish gait phases well. We used the Gaussian mixture distribution as an output probability in HMMs in order to achieve robustness against variations in observed features. Our method was superior than the other methods (e.g. [1], [4], [5]) especially for dealing with speed variations across and within a sequence.

Usually, maximum likelihood (ML) estimation is used for estimating HMM parameters, where a large amount of training data is needed. In gait-based person identification, however, it is often difficult to obtain many training samples from one walker. One effective solution to this problem is to utilize *model adaptation* techniques [10] which were first proposed in speech processing. Recently, Xu *et al.* [11] introduced it to gait-based person identification, where they employ Gaussian mixture models as walker models. Model adaptation is a process to robustly estimate the model parameters with small amount of data using a prior knowledge about the target condition. It effectively utilizes the limited amount of available data for estimating target model parameters more precisely.

In this paper, we propose *walker* adaptation and *walker-adaptive training* for gait-based person identification based on HMMs. In walker adaptation, using a small amount of data, a transfer (mapping) function from the parameters of an initial model to those of the target walker is robustly estimated. Walker-adaptive training provides a good initial model for walker adaptation, which represents intra-walker variety well.

This paper is organized as follows. Section 2 reviews our gait recognition framework. Section 3 and 4 explains our proposing methods, walker adaptation and walker adaptive training, respectively. Section 5 describes the experimental setup and reports the results obtained from our experiments. Section 6 concludes the paper.

II. CHLAC+FDA+GSP HMM

In our previous gait-recognition framework, we use CHLAC+ FDA and principal component analysis (PCA)-based gait silhouette (GSP) concatenation as the observation vectors and use an HMM as the classifier. CHLAC+FDA discriminates accurately between classes, GSP distinguishes gait phases precisely, while HMMs robustly classify gait sequences with different speeds. Their combination was proven to be robust against speed variations, even when the speed varied within a gait sequence [9].

CHLAC features are shape and motion features extracted from local autocorrelation [5]. Let $f(x, y, t)$ be pixel intensity, where x and y are pixel coordinates in one frame image, and t is the time index. Let $\boldsymbol{a}_i$ $(i = 1, \ldots, N)$ be a displacement vector from the reference point, $\boldsymbol{r} = (x, y, t)$. A set of independent local mask pattern, $(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{a}_1, \ldots, \boldsymbol{r} + \boldsymbol{a}_N)$, for capturing shape and motion characteristics of a walker's gait, is generated based on the N-th orders of correlation. We use CHLAC features with the orders 0 to 2. The 0^{th} order corresponds to the correlation within a single frame (1 mask pattern). The 1^{st} order corresponds to the correlation within a

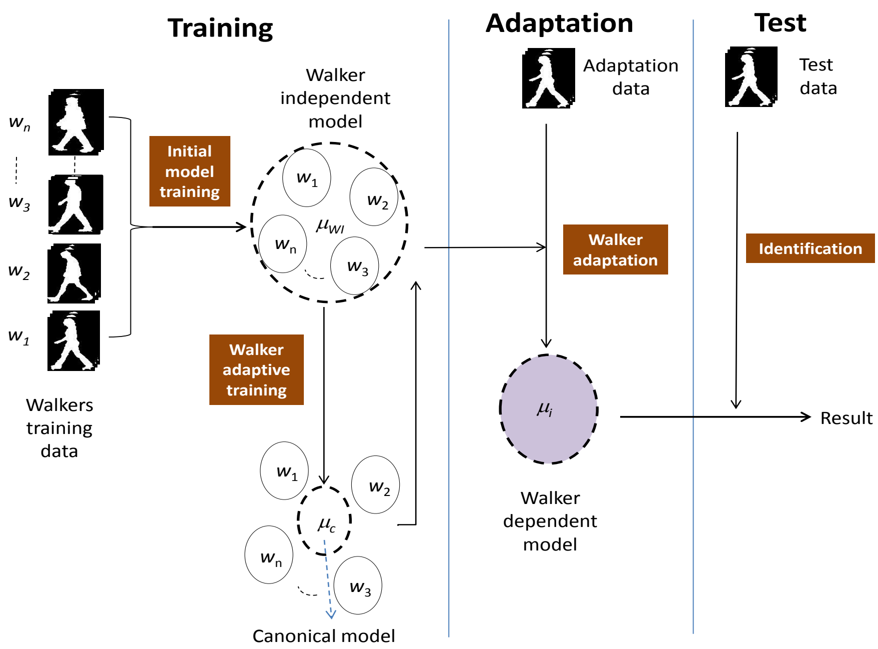

Fig. 1. The implementation of walker adaptation and walker adaptive training in gait-based person identification.

single frame and between two neighbouring frames (13 mask patterns). The 2^{nd} order corresponds to the correlation within a single frame, between two, and up to three neighbouring frames (237 mask patterns). Thus, the total dimensions of a CHLAC feature vector is 251, which corresponds to the 251 independent local mask patterns for calculating the correlation. The correlation coefficient for each pixel are then summed up. CHLAC is governed by three parameters: spatial displacement Δr, frame interval Δt, and window width T. We set $\Delta r = 4$, $\Delta t = 2$, and $T = 5$ frames. Then, the CHLAC features are mapped to the $(c-1)$-dimensional (c is the total number of classes) feature vector using Fisher discriminant analysis (FDA). We call the resulting features CHLAC+FDA.

In addition to CHLAC+FDA features, we introduce another features which have more explicit phase information, silhouette features by principal component analysis (PCA). We call the resulting features Gait Silhouette PCA (GSP). The number of dimension is $\frac{1}{2}(c-1)$. We concatenate CHLAC+FDA features and GSP features to make an input feature vector for an classifier.

We use an HMM as the classifier and employ a *continuous* gait-recognition framework. We prepare one HMM with 8 states for a half-gait cycle assuming there is symmetry between the first and the second half of the cycle in the sagittal plane view. Its topology is left-to-right without any skips. We used a mixture of Gaussian distributions with 16 components as an output probability for each state.

The more detailed explanation of this framework can be found in [9].

III. Walker Adaptation

In our previous framework described in Section I, we directly train *walker-dependent* (WD) model from the scratch using each walker's training samples. In this paper, we introduce walker adaptation (WA) for WD model estimation. By using the parameters of the walker independent (WI) model as the prior knowledge, the effective number of free parameters to be estimated is largely reduced. Accordingly, WA can estimate the model parameters more precisely than the conventional training methods when the amount of data available is small.

We use Maximum Likelihood Linear Regression (MLLR) technique [12] as an adaptation method. In MLLR adaptation, we estimate a transfer (mapping) function from walker-independent (WI) model, which is trained using training samples from all other walkers, to that of the target walker (Fig. 1). The mean vector μ_{I} of each mixture component of the WI HMM is transformed to the walker-dependent (WD) parameter μ_{D} as follows:

$$\mu_{\mathrm{D}} = A\mu_{\mathrm{I}} + b, \tag{1}$$

where A is a $d \times d$ transformation matrix, b is a d-dimension bias vector. d is the dimension of the feature vector. A and b are shared among all the mixture components of the HMM, and estimated by maximizing the likelihood of the target walker's data used for adaptation (adaptation data) using expectation-maximization (EM) algorithm.

IV. Walker Adaptive Training

There exists large variety in shape and motion characteristics of human gait. Since the WI model is built by using many walker's gait data, it represents not only *intra-walker* variety, but also such *inter-walker* variety. On the other hand, ideally, the initial model for WA should have only intra-walker variety and should not have inter-walker variety. We use walker adaptive training (WAT) to obtain a canonical walker model, which has less inter-walker variety than the WI model, and use it as the initial model for WA. Figure 1 illustrates the scheme using WAT. A similar framework has been implemented in speech recognition, in particular speaker adaptive training [13]. Now we implement the framework for the first time to gait-based person identification.

In WAT, a WI model is first trained using data from all walkers in the training set. Next, the model is adapted to each training walker using the walker adaptation in the previous section. Then, the inverse of the transform matrix, A^{-1}, for each walker is used to transform each walker's feature vector o at each time t to walker-independent feature $\hat{o}$.

$$\hat{o} = A^{-1}o - A^{-1}b, \qquad (2)$$

The transformed features $\hat{o}$ from all the time frames of all walkers are used to train the canonical walker model. By transforming the feature vectors of each walker to those of a *canonical* walker, who is the average of all the training walkers, we can effectively reduce the inter-walker variety in features caused by the difference between walkers. This procedure is carried out iteratively to obtain a lower inter-walker variability.

We use the resulting canonical walker model as the initial model for walker adaptation.

V. Experiment

A. Experimental Conditions

For the evaluation, we used USF-NIST (122 walkers) May and November 2001 database [1] for Probe A, B, C (walk on grass), CMU-MoBo database (25 walkers, walk on treadmill) [14], and TokyoTech database A [8] (30 walkers, walk on a treadmill). TokyoTech database A included walkers walking at various speeds. Its detailed description of can be found in [9].

For each image, we first subtracted a background image which was obtained in the preprocessing stage. After a certain threshold for the intensity of all pixels was set, the foreground pixels were then extracted as a binary silhouette image. Next, the bounding box around the silhouette was resized into 128×88 pixels. Silhouette images were kept in the center region.

The dimension of CHLAC+FDA feature was 121 for USF-NIST, 24 for CMU-MoBo, and 29 for Tokyo Tech A. These dimensions were automatically determined by the number of walkers to be classified. We set the dimension of GSP features to be $\frac{1}{2}(c-1)$, where c is the number of classes/walkers for each database. We used 60 for USF-NIST database, 12 for CMU MoBo, and 15 for the TokyoTech database A.

B. Results

Figure 2 shows the average results of the experiment using USF-NIST Probe A, B, and C May and November 2001 database. We compared our method with pHMM [4], GEI [2], Gabor-PDF-NN [11], Gabor-PDF-SR [11], and Gabor-PDF-LGSR [11]. pHMM [4] uses an HMM trained from many walkers for normalizing a sequence before the classification

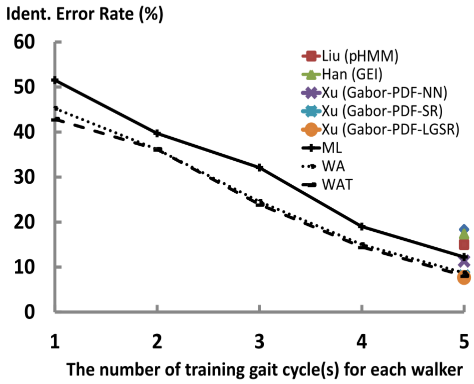

Fig. 2. Average identification error rate (%) for USF-NIST Probe A, B, and C for all walkers.

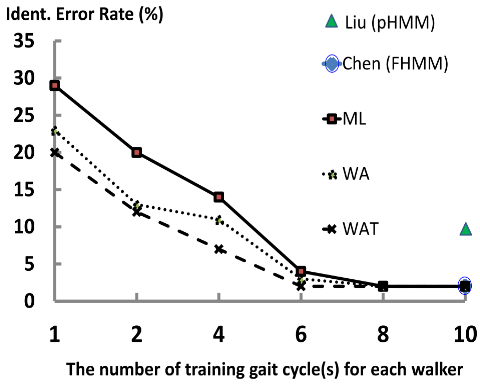

Fig. 3. Average identification error rate (%) across all speeds for CMU-MoBo for all walkers.

stage. A gait energy image (GEI) [2] is obtained by averaging the silhouette images. Gabor-PDF-NN [11] uses GMM-based Gabor features, Gabor-PDF, obtained from GEI, and the conventional nearest neighbour (NN) classifier. Gabor-PDF-SR [11] uses a sparse representation (SR) of Gabor-PDF as features. Gabor-PDF-LGSR [11] uses the local and group information of a gait sequence to obtain SR.

We varied the number of gait cycles used for model estimation to examine the robustness against data insufficiency. The errors for maximum likelihood estimation (ML), WA, and WAT are 51.5%, 45.2%, and 42.8%, respectively, by using 1 gait cycle, and 12.2%, 8.7%, and 8.0%, respectively, by using 5 gait cycles.

The error reduction rates (ERRs) of the proposed methods, WA and WAT, from ML was 12.2% and 16.8%, respectively, by using only 1 gait cycle. Our proposed methods performed better than pHMM [4], GEI [2], Gabor-PDF-NN [11], and performed almost equally as Gabor-PDF-SR [11] and Gabor-PDF-LGSR [11].

Figure 3 shows the average results of the experiment using CMU MoBo. We compared our method with pHMM [4], FHMM [7]. FHMM was employed by combining Frieze and Wavelet features using two layers of HMM. The errors for ML, WA, and WAT are 29.0%, 23.0%, and 20.0%, respectively, by using 1 gait cycle, and 2.0%, 2.0%, and 2.0%, respectively, by using 10 gait cycles. The ERRs of WA and WAT from ML,

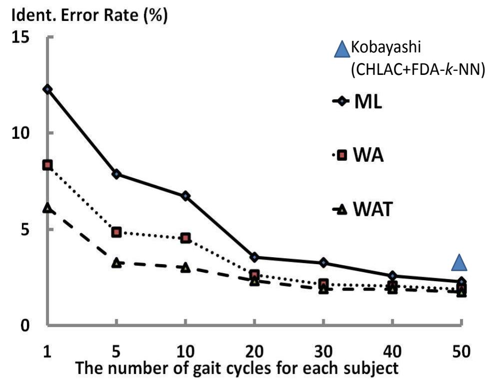

Fig. 4. Average identification error rate (%) across all speeds for TokyoTech A for all walkers.

was 20.7% and 31.0%, respectively, by using only 1 gait cycle. Our proposed methods performed significantly better than pHMM [4], and equivalently to Factorial-HMM (FHMM) [7].

Figure 4 shows the results using TokyoTech A. We compared our method with CHLAC+FDA-k-NN [5]. CHLAC+FDA-*k*-NN used CHLAC+FDA as features and *k*-NN as a classifier. For TokyoTech A (Fig. 4), the errors for ML, WA, and WAT are 12.3%, 8.3%, and 6.1%, respectively, by using 1 gait cycle, and 2.3%, 1.9%, and 1.6%, respectively, by using 50 gait cycles. The ERRs of WA and WAT from ML, was 32.5% and 50.4%, respectively, by using 1 gait cycle. It was 17.4% and 30.4%, respectively, by using 50 gait cycles. Our proposed methods performed better than CHLAC+FDA-*k*-NN [5].

These results show that WA effectively utilizes a small amount of data for walker model parameter estimation, and WAT further improves the identification performance.

In order to confirm the effectiveness of WAT, we compared the variances in the WI model and those in the canonical model. Using USF-NIST database, we calculated a value averaged over all the Gaussian distributions of all the HMM states for each model. The ratio of the averaged variance of the canonical model to that of the WI model was 0.52. This result shows that WAT effectively reduced inter-walker variability.

VI. Conclusions

We proposed walker adaptation (WA) and walker adaptive training (WAT) for gait-based person identification. These methods estimate the walker model parameters more precisely than directly estimating them from the scratch, especially when the amount of data obtained from the target walker is small. By our evaluation using three databases, USF-NIST Probe A, B, and C, CMU MoBo, and TokyoTech A, we confirmed that the proposed methods performed significantly better than the conventional methods. In particular, WAT decreased the identification error rate more than 16% when the amount of data is only 1 gait cycle. Our methods are very useful in real applications; we can decrease the load for each walker to register gait data drastically. While the MLLR transformation was proved to be effective in WAT, it is clear that such simple transformation may not be sufficient to represent speaker differences.

In future work, we plan to apply the proposed methods for reducing the influence of shape variations in gait (e.g. clothing differences). We also plan to combine our MLLR adaptation method with the other adaptation techniques, such as maximum a posteriori (MAP) adaptation, to further improve the performance of gait-based person identification.

References

[1] 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 Transactions on Pattern Analysis and Machine Intelligence, vol. 27(2):162–177, 2005.

[2] Ju Han and Bir Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach.Intell., vol. 28, no. 2, pp. 316–322, 2006.

[3] Worapan Kusakunniran, Qiang Wu, Jian Zhang, and Hongdong Li, Speed-invariant gait recognition based on procrustes shape analysis using higher-order shape conguration, pp. 545–548, 2011.

[4] Z. Y. Liu and S. Sarkar, Improved gait recognition by gait dynamics normalization, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 6, pp. 863–876, 2006.

[5] T. Kobayashi and N. Otsu, Three-way auto-correlation approach to motion recognition, Pattern Recognition Letters, vol. 30, pp. 212–221, 2009.

[6] A. Veeraraghavan, R. Chellapa, and A.K. Roy Chowdury. Rate-invariant recognition of humans and their activities. IEEE Trans. on Image Processing, vol.18, pages 13261339, 2009.

[7] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, Jie Tian, and Jie Tian, Factorial hmm and parallel hmm for gait recognition, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 39, no. 1, pp. 114–23, 2009.

[8] Ibrahim Venkat and Philippe De Wilde, Robust gait recognition by learning and exploiting sub-gait characteristics, International Journal of Computer Vision, vol. 91, no. 1, pp. 7–23, 2011.

[9] Muhammad Rasyid Aqmar, Koichi Shinoda, and Sadaoki Furui, Robust gait-based person identication against walking speed variations, IEICE Transactions, vol. 95-D, no. 2, pp. 668–676, 2012.

[10] K. Shinoda, Acoustic model adaptation for speech recognition, IEICE Trans. Inf. and System, vol. E93-D, 2010.

[11] Dong Xu, Yi Huang, Zinan Zeng, and Xinxing Xu, Human gait recognition using patch distribution feature and locality-constrained group sparse representation, IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 316–326, 2012.

[12] C. J. Leggetter and P. C. Woodland, Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models, Computer Speech and Language, vol. Vol 9, No 2, pp. 171–185, 1995.

[13] T. Anastasakos, J. McDonough, R. Schwartz, and J. Makhoul, A compact model for speaker-adaptive training, Proc. ICLSP96 vol. 2, pp. 1137–1140, 1996.

[14] R. Gross and J. Shi, The CMU Motion of Body (mobo) database, Technical Report CMU-RI-TR-01-18, Robotics Inst., Carnegie Mellon Univ., 2001.