ROBUST GAIT RECOGNITION FROM EXTREMELY LOW FRAME-RATE VIDEOS

Yu Guan[1], Chang-Tsun Li[1], and Sruti Das Choudhury[2]

Department of Computer Science[1], School of Engineering[2]
University of Warwick
Coventry, CV4 7AL, UK
g.yu, c-t.li, s.das-choudhury@warwick.ac.uk

ABSTRACT

In this paper, we propose a gait recognition method for extremely low frame-rate videos. Different from the popular temporal reconstruction-based methods, the proposed method uses the average gait over the whole sequence as input feature template. Assuming the effect caused by extremely low frame-rate or large gait fluctuations are intra-class variations that the gallery data fails to capture, we build a general model based on random subspace method. More specifically, a number of weak classifiers are combined to reduce the generalization errors. We evaluate our method on the OU-ISIR-D dataset with large/small gait fluctuations, and very competitive results are achieved when both the probe and gallery are extremely low frame-rate gait sequences (e.g., 1 fps).

Index Terms— Gait recognition, random subspace method, extremely low frame-rate, biometrics, forensics

1. INTRODUCTION

Human gait is a behavior biometric trait, which can be used for human identification. Compared with other biometric traits like fingerprint or iris, it can be used at a distance without the cooperation from subjects. Due to these reasons, human gait has great potentials in forensic applications [1], and it has gained considerable attentions recently. However, for vision-based gait recognition systems, covariate factors may affect the performance, e.g., low video frame-rate, carrying condition, clothing, walking surface, elapsed time, etc. Most of the existing methods were proposed to ameliorate the problems caused by the covariate factors related to 1) walking environments (e.g., different walking surface [2]) or 2) subjects (different clothes types, shoes, etc. [3, 4]). Only a few works attempted to solve the low frame-rate video problems. In [5], phase synchronization was adopted and low frame-rate probe videos were matched with high frame-rate gallery gait sequences. But it fails for the dual low frame-rate cases, i.e., both probe and gallery are low frame-rate videos. Several temporal reconstruction-based methods were proposed to deal with such dual low frame-rate problems. In [6], Al-Huseiny et al. proposed a level-set morphing approach for temporal interpolation. In [7], temporal Super Resolution (SR) method was employed to construct a period of high frame-rate gait sequence based on multiple periods of low frame-rate gait sequences. Based on [7], Akae et al. [8] applied an exemplar of high frame-rate image sequences to improve the temporal SR quality. A unified framework of example-based and reconstruction-based periodic temporal SR was proposed in [9], and it works well for gait recognition from (dual) extremely low frame-rate videos (with small gait fluctuations). Most of these works attempt to recover the high frame-rate gait sequences in the first place, yet in the applications of human identification, they either face high levels of reconstruction artifacts [6, 8] for extremely low frame-rate input videos (e.g., 1fps), or assume the same motion among the gait periods, which is not applicable when there are large gait fluctuations [9].

In this paper we propose a classifier ensemble approach for gait recognition tasks from the (dual) extremely low frame-rate videos, without using temporal reconstruction. Each gait sequence is averaged to produce a single grayscale image used as input feature template. Principle Component Analysis (PCA) is employed to extract the full feature set. Each weak classifier is constructed based on a small set of random features, which is generalized to the unselected features [10]. The final classification decision is made by majority voting among these weak classifiers. The rationale behind is that we can deem the errors caused by the extremely low frame-rate or gait fluctuations are intra-class variations, which can be solved by a general model [3, 4]. The rest of this paper is organized as follows: the proposed method is described in section 2, and section 3 evaluates the robustness of our method on the OU-ISIR-D dataset. Section 4 concludes this paper.

2. PROPOSED METHOD

Random Subspace Method (RSM) is an effective framework to enhance the generalization accuracy [10]. Each base classifier is generated by randomly sampling on the original feature set, before the final classifier combination, e.g., through

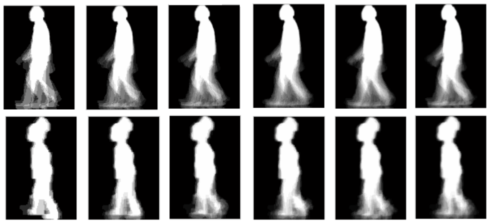

Fig. 1. The Average Gait (AG) samples from different frame-rates videos (1fps-6fps from left to right) in the OU-ISIR-D dataset. AG samples in the top/bottom row denote gait sequences with small/large gait fluctuations.

majority voting. Recently, this concept was used for face recognition [11, 12] and gait recognition [3, 4] for query images under the influence of large intra-class variations. Compared with conventional gait recognition algorithms, RSM has demonstrated its effectiveness against a large number of covariate factors such as clothing [4], shoe, carrying condition, and (small changes in) camera viewpoint [3].

For the input feature template, instead of the popular period-based Gait Energy Image (GEI) [2], in this paper we use the Average Gait (AG) of the whole gait sequence. Given a gait sequence $F_t(t = 1, \ldots, N)$, AG is defined as:

$$AG(x, y) = \frac{1}{N}\sum_{t=1}^{N} F_t(x, y). \tag{1}$$

In the low frame-rate environments, the benefits of using AG as the input feature template are two-fold: 1) gait period is not required, which is hard or impossible to estimate in certain conditions (e.g., for the 1fps gait videos); 2) the noises (due to low frame-rate or large gait fluctuations) can be smoothed by the average operation. Fig. 1 illustrates some AG samples from different frame-rates videos (1fps-6fps), and we can observe that the dynamic features (e.g., legs) are relatively dependent on video frame-rate changes while static features (e.g., head and upper body) are relatively independent. Based on this observation, we deem the unstable dynamic features caused by the low frame-rate or large gait fluctuations are intra-class variations the gallery data fails to capture and RSM concept is borrowed to solve this problem.

Given c classes/gait sequences in the gallery, there are c AGs. After concatenating each two-dimensional AG to the corresponding vector $A_i(i = 1, \ldots, c)$, the scatter matrix S can be estimated:

$$S = \frac{1}{c}\sum_{i=1}^{c}(A_i - \mu)(A_i - \mu)^T, \tag{2}$$

where $\mu = \frac{1}{c}\sum_{i=1}^{c} A_i$. The eigenvectors of S can be computed and the leading d eigenvectors are retained as candidates to construct the random subspaces.

K random subspaces can be generated and each projection matrix is formed by randomly selecting s eigenvectors (from the d candidates). Let the projection matrix be $R^k, k \in [1, K]$, and each $A_i(i = 1, \ldots, c)$ in the gallery can be projected as K sets of coefficients for the new gait representation

$$m_i^k = R^k A_i, \quad i \in [1, c], \quad k = 1, \ldots, K. \tag{3}$$

In the classification phase, R^k is used as a feature extractor (for the k^{th} subspace). Euclidean distance is then adopted to measure the dissimilarity. Given a query gait A_p, the distance between A_p and a certain class m_i^k (in the gallery) in the k^{th} subspace is:

$$D(A_p, m_i^k) = \|R^k A_p - m_i^k\|, \quad i \in [1, c]. \tag{4}$$

Nearest Neighbor (NN) classifier is used for the label assigning in the k^{th} subspace. Considering there are K subspaces/classifiers, majority voting rule is employed for the final classification decision.

3. EXPERIMENTS

3.1. Dataset and Configuration

The proposed method is evaluated on the OU-ISIR-D dataset [13], which consists of 400 gait sequences from 185 subjects. This dataset is divided into 2 parts, i.e., DB-high dataset (i.e., with small gait fluctuations) and DB-low dataset (i.e., with large gait fluctuations). For DB-high/low, there are 100 sequences in the gallery set and 100 sequences in the probe set. The original spatial resolution, frame-rate, and recording time for each sequence are 128×88 pixels, 60 fps, and 6 seconds respectively, while in this work, we use a much lower spatial resolution and frame-rate. Similar to [5], the low resolution sequences are constructed by scaling down the original silhouettes, and the low frame-rate sequences are constructed by selecting the gait silhouettes at a specified interval. Based on both DB-high and DB-low in a spatial resolution of 32×22 pixels, three experiments are conducted, i.e.,

1. To check the performance sensitivity to the feature number s used for each subspace.
2. To evaluate the robustness when both the gallery and probe are from extremely low frame-rate videos (i.e., 1fps).
3. To check the effect when the gallery has relatively higher frame-rates (e.g., 2fps-6fps) than the probe (1fps).

To evaluate the algorithm performance, in this work we use the rank-1 Correct Classification Rate (CCR), which is the percentage of the correct individual ranked as the top 1 candidate.

3.2. The Effect of Feature Number

For each random subspace, the corresponding base classifier has some generalization power for the unselected features,

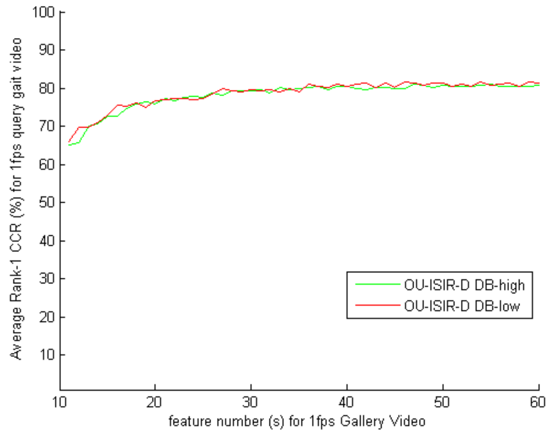

Fig. 2. The rank-1 CCR distribution (%) with respect to different feature numbers. (probe/gallery frame-rate: 1fps)

but it may face the underfitting problem if the feature number is too small. Following [4], we set the classifier number $K = 500$. Then we check the sensitivity of the feature number s within the range $[11, 60]$ and the rank-1 CCR distribution with respect to the feature number s is illustrated in Fig. 2. For both DB-high and DB-low, the performance is relatively stable when $s \in [11, 60]$, and in this work we set $s = 40$. It also should be pointed out that the proposed method is less generalized if larger number of features are employed [10], and with an increasing number of features (out of the total $d = 200$ features), the CCR of our method converges to conventional PCA performance.

3.3. Gait Recognition in the Extremely Low Frame-Rate Videos

Based on 1fps gait videos, we compare our method with several temporal reconstruction-based, i.e., morphing-based reconstruction (Morph) [6], Periodic Temporal SR (PTSR) [8], and Example-based and Reconstruction-based Temporal SR (ERTSR) [9]. Due to the random nature of our method, we repeat each experiment 10 times and report the mean rank-1 CCRs (with the standard deviation) in Table 1. Some observations can be made:

1. The AG-based methods deliver similar results on DB-high and DB-low datasets, which suggests that the feature template AG is less sensitive to large gait fluctuations (in the extremely low frame-rate conditions).
2. Due to the extremely low frame-rate, the temporal reconstruction-based methods may bring high levels of artifacts that hamper the performance. For example, in this experiment, Morph and PTSR even have worse performance than the baseline method (i.e., AG+NN).
3. Although ERTSR outperforms our method on DB-high dataset, it should be pointed out that ERTSR assumes the same motion among the gait periods, which is not applicable when there are large gait fluctuations (e.g., DB-low dataset) [9].
4. Although the proposed method is derived from PCA, it significantly outperforms AG+PCA+NN, which uses a single feature extractor consisting of all the d leading eigenvectors.

Table 1. Algorithms comparison in terms of rank-1 CCR (%) on the OU-ISIR-D dataset. (probe/gallery frame-rate: 1fps)

-	DB-high	DB-low
Morph [6]	52	N/A
PTSR [8]	44	N/A
ERTSR [9]	87	N/A
AG+PCA+NN	69	67
AG+NN	68	68
AG+RSM (our method)	80.40±1.35	80.60±1.26

For real-world scenarios, due to low frame-rate and large gait fluctuations, it is a challenging task for the temporal reconstruction-based methods. Our method suggests a simple yet effective way in such conditions. The feature template AG takes advantage of the low frame-rate videos when there are large gait fluctuations and the reasons are two-fold: 1) the noisy frames may be skipped to some extent due to the low frame-rate nature; 2) the average operation gets the static part reenforced and the unstable dynamic part smoothed. During the classification phase, classifier ensemble based on RSM further boosts the performance, in contrast to single classifier AG+PCA+NN.

3.4. The Effect of Higher Frame-Rate Gallery Videos

Based on the proposed method, we also investigate the effect when the gallery videos have higher frame-rates than the given 1fps query gait videos. Fig. 3 illustrates the rank-1 CCR distribution with respect to the gallery frame-rates from 1fps to 6fps. Generally, for the given 1fps probe gait videos (both DB-high and DB-low), it tends to exhibit better performance with higher frame-rates gallery videos. However, the CCR curve of DB-low grows relatively more slowly than the one of DB-high over the 2fps-6fps gallery videos. We can explain this phenomenon by taking the enrolled data noise into consideration. For DB-low with larger gait fluctuations, although the sensor noise can be reduced with respect to the increasing number of video frame-rates, it instead suffers from the noise caused by large gait fluctuations, since it is more unlikely (for the AG template) to skip the noisy frames in the higher frame-rate environments. For higher frame-rates gallery videos (e.g., 3fps-6fps), by comparing with the results of DB-high in Fig. 3, we can see the effect of such trade-off between sensor noise and large gait fluctuations noise. Nevertheless, given the low frame-rates probe/gallery videos with

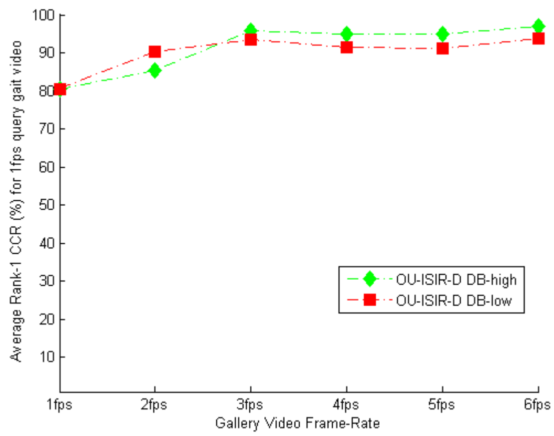

Fig. 3. The rank-1 CCR distribution (%) with respect to different gallery video frame-rates.

large/small gait fluctuations, the general satisfactory performance indicates the effectiveness of our proposed method.

4. CONCLUSIONS

This paper proposes a simple yet effective gait recognition method to address the challenges caused by extremely low frame-rate video conditions or large gait fluctuations. Our method has 3 desirable properties: 1) the performance is less sensitive to the feature number used in the base classifiers; 2) it yields satisfactory results for extremely low frame-rates gait sequences under the influences of large/small gait fluctuations; 3) the performance can be further enhanced if the gallery videos have higher frame-rates. Compared with the temporal reconstruction-based methods [6, 8, 9], our method delivers significant improvements in terms of performance or generalization.

There are several potential applications of our proposed method. For example, since our method is less sensitive to the frame number of a gait sequence, for gait recognition tasks in the noisy environments, we can simply filter out (a large number of) low-quality frames by using some quality metrics (e.g., [14]) to boost the performance. Besides, since each base classifier in the RSM framework is generated in an independent way, parallel implementation can also be adopted for time-critical systems.

5. REFERENCES

[1] E.B. Simonsen P.K. Larsen and N. Lynnerup, "Gait analysis in forensic medicine," *Journal of Forensic Sciences*, vol. 53, pp. 1149–1153, 2008.

[2] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2006.

[3] Y. Guan, C.-T. Li, and Y. Hu, "Random subspace method for gait recognition," in *Proc. Int'l Conf. Multimedia and Expo Workshops*, pp. 284–289, 2012.

[4] Y. Guan, C.-T. Li, and Y. Hu, "Robust clothing-invariant gait recognition," in *Proc. Int'l Conf. Intelligent Information Hiding and Multimedia Signal Processing*, pp. 321–324, 2012.

[5] A. Mori, Y. Makihara, and Y. Yagi, "Gait recognition using period-based phase synchronization for low framerate videos," in *Proc. Int'l Conf. Pattern Recognition*, pp. 2194–2197, 2010.

[6] M.S. Al-Huseiny, S. Mahmoodi, and M.S. Nixon, "Gait learning-based regenerative model: A level set approach," in *Proc. Int'l Conf. Pattern Recognition*, pp. 2644–2647, 2010.

[7] Y. Makihara, A. Mori, and Y. Yagi, "Temporal super resolution from a single quasi-periodic image sequence based on phase registration," in *Proc. Asian Conf. Computer Vision*, pp. 107–120, 2011.

[8] N. Akae, Y. Makihara, and Y. Yagi, "Gait recognition using periodic temporal super resolution for low framerate videos," in *Proc. Int'l Joint Conf. Biometrics*, pp. 1–7, 2011.

[9] N. Akae, A. Mansur, Y. Makihara, and Y. Yagi, "Video from nearly still: An application to low frame-rate gait recognition," in *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 1537–1543, 2012.

[10] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[11] X. Wang and X. Tang, "Random sampling lda for face recognition," in *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 259–265, 2004.

[12] N.V. Chawla and K.W. Bowyer, "Random subspaces and subsampling for 2-d face recognition," in *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 582–589, 2005.

[13] A. Tsuji M. A. Hossain K. Sugiura A. Mori Y. Makihara, H. Mannami and Y. Yagi, "The ou-isir gait database comprising the treadmill dataset," *IPSJ Trans. on Computer Vision and Applications*, vol. 4, pp. 53–62, 2012.

[14] D. Matovski, M. S. Nixon, S. Mahmoodi, and T. Mansfield, "On including quality in applied automatic gait recognition," in *Proc. Int'l Conf. Pattern Recognition*, 2012.(In Press).