Uncooperative Gait Recognition Using Joint Bayesian

Chao Li[1(✉)], Kan Qiao[2], Xin Min[1], Xiaoyan Pang[2], and Shouqian Sun[1]

[1] College of Computer Science and Technology, Zhejiang University,
Hangzhou 310027, China
{superli,minx,ssq}@zju.edu.cn
[2] Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou 310027, China
{qiaok,pangxiaoy}@srrsh.com

Abstract. Human gait, as a soft biometric, helps to recognize people by walking without subject cooperation. In this paper, we propose a more challenging uncooperative setting under which views of the gallery and probe are both unknown and mixed up (uncooperative setting). Joint Bayesian is adopted to model the view variance. We conduct experiments to evaluate the effectiveness of Joint Bayesian under the proposed uncooperative setting on OU-ISIR Large Population Dataset (OULP) and CASIA-B Dataset (CASIA-B). As a result, we confirm that Joint Bayesian significantly outperform the state-of-the-art methods for both identification and verification tasks even when the training subjects are different from the test subjects. For further comparison, the uncooperative protocol, experimental results, learning models, and test codes are available.

Keywords: Gait recognition · Joint Bayesian
Uncooperative gait recognition

1 Introduction

Biometrics refers to the use of intrinsic physical or behavioral traits in order to identify humans. Besides regular features (face, fingerprint, iris, DNA and retina), human gait, which can be obtained from people at larger distances and at low resolution without subjects' cooperation has recently attracted much attention. It also has a vast application prospect in crime investigation and wide-area surveillance. For example, criminals usually wear gloves, dark sun-glasses, and face masks to invalidate finger print, eyes, and face recognition. In such scenarios, gait recognition is the only useful and effective identification method. Previous research [1,2] has shown that human gait, specifically the walking pattern, is difficult to disguise and unique to each people.

In general, video sensor-based gait recognition methods are divided into two families: appearance-based [3–5] and model-based [6–8]. In the appearance-based methods, it focus on the motion of human body and usually operate

on silhouettes of gait. They extract gait descriptors from the silhouettes. The general framework of appearance-based methods usually consists of silhouette extraction, period detection, representation generation, and recognition. A typical example is gait energy image (GEI) [3] which is proposed as a mixture of dynamic and static features. In model-based gait recognition, it focus more on the extraction of the stride parameters of subject that describe the gait by using the human body structure. The model-based methods usually require high resolution images as well as are computationally expensive while gait recognition needs to be real-time and effective at low resolution. However, the performance of gait recognition is often influenced by several variations such as clothing, walking speed, observation views, and carrying bags. For appearance-based methods, view changes are the most problematic variations.

A. Joint Bayesian for modeling view variance

When dealing with view change problems, many appearance-based methods are proposed: (1) view transformation model (VTM) based generative approaches [9,10]. (2) view-invariant feature-based approaches [11]. (3) multi-view gallery-based approaches [12]. (4) subspace-based view-invariant discriminative approaches [13–15]. However, VTM-based approaches(e.g. TCM+ [10], wQVTM [9]) as well as some discriminative approaches (GMMFA [13]) often require view information for a matching pair, while the information usually can't be obtained easily without subject's cooperate.

So, we introduce Joint Bayesian to model the view variance which differs from the above approaches, and the commonly used GEI is adopted as the input gait representation. After the training process, the proposed method can be easily used without any view information in advance.

B. Uncooperative gait recognition and transform learning

Most exist cross-view gait recognition methods [9,10,13] are based on the assumption that gallery and probe views are known as a priori or fixed (cooperative setting) while this assumption is often not valid in practice.

Usually, the gallery and probe view are often unknown and mixed up (uncooperative setting). However, to the best of our knowledge, only a few studies [11,14,15] focus on uncooperative gait recognition. In [14,15], the uncooperative setting just consider two different views every time while our proposed setting consider four different views and is more complex. Our proposed uncooperative setting is same with [11], but they just conduct experiments on OU-ISIR Large Population Dataset (OULP) [16] while we also use the famous CASIA-B Dataset (CASIA-B) [17] as a benchmark. Additionally, the training subjects are often different from the test subjects, so that transfer learning is performed [14,15].

2   Gait Recognition

Usually, gait recognition can be divided into two major tasks: gait verification and gait identification as in face recognition [18–20]. Gait verification is used for

verifying whether two input gait sequences (Gallery, Probe) belong to the same subject. In this paper, we calculated the similar score ($SimScore$) using Joint Bayesian to evaluate the similarity of two given sequences. Euclidean distance was also adopted as a baseline method for comparison. In gait identification, a set of subjects are gathered (The gallery), and it aims to decide which of the gallery identities are similar to the probe at test time. Under the closed set identification condition, a probe sequence is compared with all the gallery identities, then identity which has the largest $SimScore$ is the final result.

2.1   Gait Verification Using Joint Bayesian

In this paper, we modeled gait representations by summing two independent Gaussian variables as:

$$x = \mu + \varepsilon \tag{1}$$

where x represents a mean-subtracted representation vector. For a better performance, L_2 - normalization was applied for gait representations. μ is gait identity following a Gaussian distribution $N(0, S_\mu)$. ε stands for different gait variations (e.g., view, clothing and carrying bags etc.) following a Gaussian distribution $N(0, S_\varepsilon)$. Joint Bayesian models the joint probability of two gait representations using the intra-class variation (I) or inter-class variance (E) hypothesis, $P(x_1, x_2|H_I)$ and $P(x_1, x_2|H_E)$. Given the above prior from Eq. 1 and the independent assumption between μ and ε, the covariance matrix of $P(x_1, x_2|H_I)$ and $P(x_1, x_2|H_E)$ can be derived separately as:

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{bmatrix} \tag{2}$$

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\varepsilon & 0 \\ 0 & S_\mu + S_\varepsilon \end{bmatrix} \tag{3}$$

S_μ and S_ε are two unknown covariance matrices which can be learned from the training set using EM algorithm. During the testing phase, the likelihood ratio ($r(x1, x2)$) is regarded as the similar score ($SimScore$):

$$SimScore(x_1, x_2) = r(x_1, x_2) = log\frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)} \tag{4}$$

$r(x_1, x_2)$ is efficient to obtained with the following closed-form process:

$$r(x_1, x_2) = x_1^T A x_1 + x_2^T A x_2 - 2x_1^T G x_2 \tag{5}$$

where A and G are two final result model, which can be obtained by using simple algebra operations between S_μ and S_ε. Please refer to [21] for more details. We also public our trained model (A and G) and testing codes in https://pan.baidu.com/s/1qYk9HoC for further comparison.

2.2   Gait Identification Using Joint Bayesian

For gait identification, the probe sample x_p is classified as class i, if the final $SimScore$ with all the gallery (x_i) is the maximum as shown in Eq. 6.

$$i = arg \max_{i \in [0, N_{gallery}-1]} SimScore(x_i, x_p) \tag{6}$$

where $N_{gallery}$ is the number of training subjects. In the experiments, we just used the first period of the gait sequence.

3   Experiments

To evaluate the performance of Joint Bayesian under uncooperative setting [11,14,15], extensive experiments have been carried out on the two largest public gait dataset: OU-ISIR Large Population Dataset (OULP) [16] and CASIA-B Dataset (CASIA-B) [17]. For comparison, we just considered four different views on OULP (55°, 65°, 75°, 85°) and CASIA-B (36°, 54°, 72°, 90°), respectively.
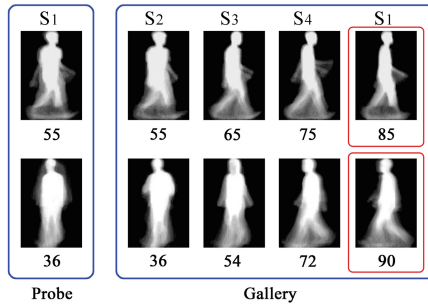
Fig. 1. Examples of GEIs from different people in OULP (top) and CASIA-B (bottom) under four different view conditions. The first S1 appears to be the best match to S2, because they are under the same view, which can easily lead to a wrong match.

Table 1. Comparison of rank-1 (%) and EERs (%) with other existent methods on OULP in uncooperative setting.

	Joint Bayesian	GEINet	LDA	1NN
Rank-1 [%]	96.81 ± 0.60	89.70	91.09 ± 0.89	21.62 ± 1.44
EERs [%]	1.58 ± 0.14	1.60	4.11 ± 0.24	31.53 ± 0.95

Table 2. Comparison of rank-1 (%) and EERs (%) with other existent methods on CASIA-B in uncooperative setting.

|	Joint Bayesian	RankSVM	LDA	1NN
Rank-1 [%]	86.4 ± 2.78	72.2 ± 3.80	76.0 ± 3.56	12.9 ± 3.8
EERs [%]	8.3 ± 1.58	15.5 ± 1.72	12.6 ± 1.34	42.5 ± 3.79

3.1 Experiments Settings

Gait Features. We first computed gait periods in each gait sequence and then extracted the most commonly used gait energy image (GEI) [3] which is proposed as a mixture of dynamic and static features. GEI is calculated by averaging gait silhouettes over a gait cycle. If the gait sequence has more than one cycle, we just chose the first one. For preprocessing, gait silhouette images were scaled to 64×44 pixel-sized images and PCA was adopted to preserve 95% of the variance before Joint Bayesian was applied. GEIs under four view conditions are shown in Fig. 1.

Uncooperative Setting. All experiments are carried out following the uncooperative protocol as follows unless otherwise specified. First of all, the whole set of gait sequences is equally and randomly divided into two groups of the same number of subjects, one for training and the other for testing, i.e. the subjects in the two groups are different and transfer learning is performed. Secondly, the test data is further split into a gallery set and a probe set as the following steps: (1) A gallery view of each subject is drawn randomly from four different views; (2) A probe view of corresponding subject is randomly chosen from the other three views. We have made public details of our division for all the experiments in https://pan.baidu.com/s/1qYk9HoC.

Benchmarks. On the two gait datasets, two commonly used methods are adopted as baseline methods. They are : (1) 1 Nearest Neighbor classifier (1NN). The original gait representation (GEI) are used in this method, and it has a relatively high dimensionality ($64 \times 44 = 2816$); (2) Linear Discriminant Analysis (LDA): Firstly, PCA is adopted along with LDA to achieve the best performance as in [14,15].

Additionally, on CASIA-B, RankSVM [14,15] achieves the best performance under uncooperative setting while they just consider two different views every time. RanSVM are so computationally expensive when the training subjects increase that it is not suitable for OULP which has a large training population (956 subjects). On OULP, GEINet [11] is the state-of-the-art method which uses the deep learning method and the performance is dependent on the number of training subjects, so that it is not suitable for CASIA-B. RankSVM and GEINet are also adopted as the comparison methods separately on the two datasets. The results of GEINet are provided by the authors while RankSVM are implemented by ourselves.

Evaluation Criteria. The recognition performance is evaluated using four metrics: (1) cumulative match characteristics (CMC) curve, (2) rank-1 identification rates, (3) the receiver operating characteristic (ROC) curve of false acceptance rates (FAR) and false rejection rates (FRR), and (4) equal error rates (EERs). CMC curve, and rank-1 identification rates are used for identification task while ROC curve and EERs are used for verification task.

| (a) OULP | (b) CASIA-B |
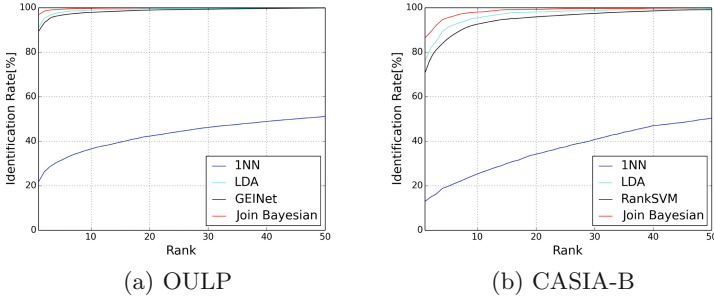
Fig. 2. CMC curves of two different datasets in uncooperative setting.

| (a) OULP | (b) CASIA-B |
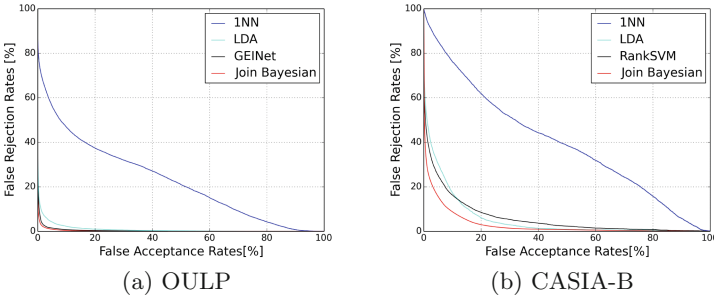
Fig. 3. ROC curves of two different datasets in uncooperative setting.

3.2 Experimental Results on OULP

The OULP has nearly 4000 subjects, and because of the largest population, experimental results can be calculated in a statistically reliable way. Each subject has two video sequences (Gallery, Probe) and is at four view angles (55°, 65°, 75°, 85°). GEIs at different views with four sample subjects are shown in Fig. 1.

We used a subset (1912 subjects) of OULP following the uncooperative protocol of [11] and the subset was further divided into two groups of the same number of subjects, one for training while the other one for testing. To reduce any effect

of random grouping, five 2-fold cross validations were performed. During each train phase, $956 * (956 - 1) = 912980$ intra-class samples and $956 * 1 = 956$ inter-class samples were used for training Joint Bayesian. For preprocessing, gait silhouette images were scaled to 64×44 pixel-sized images and PCA was adopted to preserve 95% of the variance.

We summarize rank-1 identification rate and EERs in Table 1. Furthermore, Figs. 2a, and 3a show more details of CMC and ROC curves. We find that our proposed method significantly outperforms the benchmarks with respect to rank-1 and EERs under uncooperative setting. More specifically, compared with the state-of-the-art method (GEINet), rank-1 identification rates of Joint Bayesian improves from 89.70% to 96.81%; for verification task, our proposed method also achieve a competitive result with respect to EERs. We can also find that learning based method (1NN PCA+LDA) significantly outperforms the existing template matching methods (1NN).

3.3   Experimental Results on CASIA-B

In CASIA-B, totally 124 subjects gait data are captured from 11 views (0°–180°). Between two nearest view directions, the interval angle is 18°. Three covariate condition changes, i.e., clothing, carrying, and view angle are all considered. In this dataset, each subject has 10 gait sequences: 6 normal walking sequences (nm), 2 carrying-bag sequences (bg), and 2 wearing-coat sequences (cl).

For consistency with the result of OULP, we just considered the view covariate condition change (6 nm sequences), and four similar views (36°, 54°, 72°, 90°) were selected. Following the uncooperative protocol, the dataset was also divided into two groups of the same number of subjects, on for training and one for testing. As in Sect. 3.2, five repeated experiments performed. As a result 5 2-fold cross validations were performed, and the same processing are adopted as in 3.2.

We summarize the rank-1 identification rate and EERs in Table 2. CMC and ROC curves are also shown in Figs. 2b and 3b. They show similar trends as those in the OULP experiments that Joint Bayesian achieves the best results for both the identification and verification tasks. We can also find that: (1) the corresponding methods perform better on OULP than on CASIA-B due to OULP's cleaner silhouettes and larger training subjects; (2) RankSVM loses it's stronger power than in [14,15] because our proposed uncooperative setting is more challenging; (3) the insufficient dataset leads to volatile results for all the methods.

4   Conclusion

In this paper, Joint Bayesian is used for model the view variance for uncooperative gait recognition. Extensive experiments have been conducted to validate the effectiveness of our method particularly under our proposed more challenging

uncooperative setting. Our proposed method which learns transferable information independent of the identity of people achieved state-of-the-art results for both the identification and verification tasks through experiments on OULP and CASIA-B datasets. What's more important, Joint Bayesian can be trained from different subjects and performs better which makes it more generally applicable.

In our future works, we will evaluate our proposed method with a wider view variation, or other variations (e.g. clothing, carrying bags). Additionally, cross-dataset gait recognition will be evaluated and the novel deep convolutional features will also be considered.

References

1. Murray, M.P., Drought, A.B., Kory, R.C.: Walking patterns of normal men. J. Bone Joint Surg. Am. **46**(2), 335–360 (1964)
2. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**(5), 353–356 (1977)
3. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette. In: Proceedings of 2004 International Conference on Pattern Recognition, pp. 211–214. IEEE Press, New York (2004)
4. Lam, T.H., Cheung, K.H., Liu, J.N.: Gait flow image: a silhouette-based gait representation for human identification. Pattern Recogn. **44**(4), 973–987 (2011)
5. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: Proceedings of 3rd International Conference on Crime Detection and Prevention, pp. 1–6. IEEE Press, New York (2009)
6. Luo, J., Tang, J., Tjahjadi, T., Xiao, X.: Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. Pattern Recogn. **60**, 361–377 (2016)
7. Bhanu, B., Han, J.: Model-based human recognition 2D and 3D gait. In: Bhanu, B., Han, J. (eds.) Human Recognition at a Distance in Video. Advances in Pattern Recognition, pp. 65–94. Springer, Heidelberg (2010). https://doi.org/10.1007/978-0-85729-124-0_5
8. Nixon, M.S., Carter, J.N., Cunado, D., Huang, P.S., Stevenage, S.: Automatic gait recognition. In: Jain, A.K., Bolle, R., Pankanti, S. (eds.) Biometrics, pp. 231–249. Springer, Boston (1996). https://doi.org/10.1007/0-306-47044-6_11
9. Muramatsu, D., Makihara, Y., Yagi, Y.: View transformation model incorporating quality measures for cross-view gait recognition. IEEE Trans. Cybern. **46**(7), 1602–1615 (2016)
10. Muramatsu, D., Makihara, Y., Yagi, Y.: Cross-view gait recognition by fusion of multiple transformation consistency measures. IET Biometrics **4**(2), 62–73 (2015)
11. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y.: GEINet: view-invariant gait recognition using a convolutional neural network. In: 2016 International Conference on Biometrics, pp. 1–8. IEEE Press, New York (2016)

12. Iwashita, Y., Baba, R., Ogawara, K., Kurazume, R.: Person identification from spatio-temporal 3D gait. In: 2010 International Conference on Emerging Security Technologies, pp. 30–35. IEEE Press, New York (2010)
13. Sharma, A., Kumar, A., Daume, H., Jacobs, D.W.: Generalized multiview analysis: a discriminative latent space. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2160–2167. IEEE Press, New York (2012)
14. Martín-Félez, R., Xiang, T.: Gait recognition by ranking. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 328–341. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33718-5_24
15. Martín-Félez, R., Xiang, T.: Uncooperative gait recognition by learning to rank. Pattern Recogn. **47**(12), 3793–3806 (2014)
16. Iwama, H., Okumura, M., Makihara, Y., Yagi, Y.: The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans. Inf. Forensics Secur. **7**(5), 1511–1521 (2012)
17. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 2006 International Conference on Pattern Recognition, pp. 441–444. IEEE Press, New York (2006)
18. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognitionm, pp. 1891–1898. IEEE Press, New York (2014)
19. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems, pp. 1988–1996. IEEE Press, New York (2014)
20. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900. IEEE Press, New York (2015)
21. Chen, D., Cao, X., Wang, L., Wen, F., Sun, J.: Bayesian face revisited: a joint formulation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 566–579. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33712-3_41