PERSON IDENTIFICATION USING SPATIOTEMPORAL MOTION CHARACTERISTICS

Muhammad Hassan Khan[1,2], *Muhammad Shahid Farid*[2], *Marcin Grzegorzek*[1,3]

[1]Research Group of Pattern Recognition, University of Siegen, Siegen, Germany
[2]College of Information Technology, University of the Punjab, Pakistan
[3]University of Economics in Katowice, Katowice, Poland

ABSTRACT

Biometric gait recognition has received substantial attention of researchers in the recent years due to its applications in numerous fields of computer vision, particularly in visual surveillance and monitoring systems. Most existing gait recognition algorithms solve the problem of person identification either by constructing a human body model based on various skeletal data characteristics such as joints positioning and their orientation, or use gait features, e.g., stride length, gait patterns and other shape templates. Such approaches require the extraction of the human-body's silhouette, contour, or skeleton from the images, and therefore their performance highly depends on the silhouette segmentation accuracy. In this paper, we propose a novel gait recognition algorithm which exploits spatiotemporal motion characteristics of a person, which does not need silhouette or skeleton extraction at all. The proposed algorithm computes a set of spatiotemporal features from the video sequences and uses them to generate a codebook. Fisher vector is used to encode the motion descriptors which are classified using linear Support Vector Machine (SVM). The proposed algorithm is evaluated on three benchmark gait datasets: TUM GAID, CASIA-B, and CASIA-C. It achieved excellent results on all datasets which demonstrate the effectiveness of the proposed algorithm.

Index Terms— Gait recognition, Spatiotemporal features, Fisher vector encoding, Visual surveillance

1. INTRODUCTION

Biometrics has received significant research efforts in the recent years due to its growing applications in authentication, access control and surveillance. Studies [1–3] have shown that individuals can be identified by using different distinguishing biological traits. Biometrics refers to the physiological or behavioral characteristic of the human, e.g., fingerprints, facial features, iris, DNA, voice, and gait, which have proven to be unique for each individual. Gait refers to the walking style of a person and is considered an important cue for person identification. Unlike other biometrics, gait does not require human interaction with the system which makes it the most suitable for surveillance systems. Moreover, gait biometrics can be used at low resolution in a non-invasive and hidden manner. Gait recognition, however, is challenging as many factors may affect it such as clothing, shoes, walking surface and injuries. Gait may not be as powerful as other biometric modalities such as fingerprints to identify the individuals, however its characteristic to recognize human from distance and without any interaction makes it irreplaceable in many applications such as visual surveillance.

The gait recognition approaches in literature can be divided into two broad categories: model-based and model-free approaches. The model-based techniques build the human body structure and motion models by tracking the different body parts and joint position over time using the underlying mathematical structure [4], and use them to recognize the people. These models such as [5–7] may include stick figure, interlinked pendulum and ellipse are generally constructed based on the prior knowledge of the human body shape. Recent studies have demonstrated that such models are capable to deal with the occlusion and rotation problems. However, they are computationally inefficient and sensitive to the quality of video data, and therefore they are not considered suitable for real-world and real-time applications [8].

The model-free gait recognition approaches do not use a structural model of human motion, instead they usually operate on the sequence of extracted human silhouettes. In particular, such algorithms either use the temporal information of human motion [9–11] or construct a template from silhouettes images [12–15], and use them to recognize the individual's gait. The gait recognition methods [16, 17] extract the human silhouette from the background using the depth and skeleton information from Microsoft Kinect, and compute various features for gait recognition. However, the biggest restriction is the field-of-view, which is very limited ($1 - 4$ meters) [18]. In contrast to model-based gait recognition approaches, the model-free techniques have shown more promising recognition results on various gait datasets. Moreover they are computationally efficient too.

In this paper, we present a novel spatiotemporal gait repre-

This research was partially supported by the University of the Punjab and German Federal Ministry of Education and Research within the project "SenseVojta: Sensor-based Diagnosis, Therapy and Aftercare According to the Vojta Principle" (Grant Number: 13GW0166E).

sentation using dense trajectories to characterize the distinctive motion traits of human gait. Unlike most existing gait recognition algorithms that require the extraction of the human body silhouette or other skeletal information, the proposed approach is model-free. It neither involves any kind of human body segmentation nor requires gait cycle estimation. Experiments worked out on three well-known gait databases confirm the effectiveness of the proposed algorithm.

2. PROPOSED METHOD

The proposed gait recognition algorithm works in three steps. First, dense trajectories are generated based on optical flow field and their motion information is encoded using local descriptors. Second, a codebook based on Gaussian Mixture Model (GMM) is built and the local descriptors are encoded using Fisher vector (FV). Finally, the computed features are classified using linear Support Vector Machine (SVM) to recognize the individuals.

2.1. Motion descriptor estimation

Recently, dense trajectories have demonstrated excellent results in action recognition [19, 20]. Our motivation to use dense trajectories is that they encode the local motion patterns of gait and can be easily computed from video sequences. To extract dense trajectories, a set of dense points is selected from each frame and tracked in successive frames using displacement information from a dense optical flow field. Given a trajectory of length L, a sequence S of displacement vector ΔP_t is computed as given below [19]:

$$S = (\Delta P_t, \cdots, \Delta P_{t+L-1}), \tag{1}$$

where $\Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$. P_t and and P_{t+1} represent a point in frame t and $t+1$ respectively. The sequence vector S is then normalized by the sum of the magnitudes of the displacement vector. That is,

$$S' = \frac{(\Delta P_t, \cdots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|} \tag{2}$$

The descriptor S' encodes the shape of the trajectory. Wang et al. [19] proposed the Histogram of Oriented Gradient (HOG) and Histogram of Optical Flow (HOF) features along the dense trajectories. In addition, to encode the relative motion information between pixels, the derivatives along the horizontal and the vertical components of the optical flow are also computed, known as Motion Boundary Histograms and are represented as MBH$_x$ and MBH$_y$ respectively. We evaluated various combinations of these descriptors on TUM GAID database [17], and the results are shown in Fig. 1. These results reveal that HOG in combination with MBH outperform the rest and achieves up to 100% recognition accuracy. HOG captures the characteristics of a person's static appearance and MBH highlights the information about the changes in optical flow field. Therefore, combining the person's appearance information with local motion characteristics has great impact on the identification of a person.

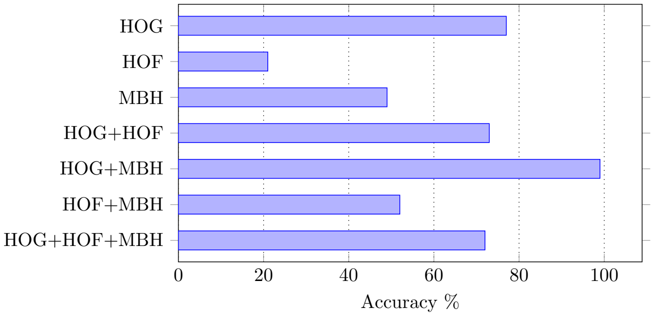

Fig. 1. Performance of various motion descriptors for gait recognition on TUM GAID gait database.

2.2. Feature encoding

Inspired by the recent popularity of FV encoding in image classification, object detection and action recognition [20,21], we encode our local descriptors using FV and a codebook based on GMM. FV is derived from the Fisher kernel [21] that combines the characteristics of both discriminative and generic approaches. The basic idea is to model a feature set by gradient of its log-likelihood function with respect to model parameters. To build a codebook, we used GMM from one million randomly selected features of each descriptor. GMM is a generative model that defines the distribution over feature space and can be described as follows:

$$p(X \mid \theta) = \textstyle\sum_{i=1}^{K} w_i \mathcal{N}(x \mid \mu_i, \sum_i) \tag{3}$$

where $i = 1, 2, ..., K$ is the mixture (i.e., cluster) number, w_i, μ_i and $\sum_i$ are the weight, mean vector and covariance matrix of the ith cluster, respectively. Furthermore, $\theta = \{w_i, \mu_i, \sum_i\}$ is the set of model parameters, and $\mathcal{N}(X \mid \mu_i, \sum_i)$ represents the D-dimensional Gaussian distribution. For a given feature set $X = \{x_t, t = 1, ..., T\}$, the optimal parameters of GMM are learned using maximum likelihood estimation. The soft assignment of data x_t to cluster i can be defined as,

$$q_t(i) = \frac{w_i \mathcal{N}(x_t \mid \mu_i, \sum_i)}{\sum_{j=1}^{K} w_j \mathcal{N}(x_t \mid \mu_j, \sum_j)} \tag{4}$$

We assume that each model represents a specific motion pattern shared by the local descriptors in the codebook. The Expectation Maximization (EM) algorithm of GMM applies soft assignments of the feature descriptor to each mixture component. Therefore, the local descriptors will be assigned to multiple clusters in a weighted manner using the posterior component probability given by the descriptor. The feature set X can be modeled into a vector by computing the gradient vector of its log-likelihood function at the current θ,

$$F_X = \frac{1}{T}\nabla_\theta logp(X|\theta), \qquad (5)$$

where F_X represents the FV and ∇_θ is the gradient of the log-likelihood function, which describes the contribution of parameters in the generation process. Let x_t be the local descriptor, $q_t(i)$ be the soft assignment of x_t to cluster i, σ_i is the diagonal element of $\sum_i$; u_i and v_i are the gradient vector with respect to μ_i and σ_i, respectively [22]:

$$u_i = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^{T} q_t(i) \frac{x_t - \mu_i}{\sigma_i} \qquad (6)$$

$$v_i = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^{T} q_t(i) \left[\frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right], \qquad (7)$$

Equation (6) and (7) are known as the first and the second order differences of descriptor points to cluster centers, respectively. The final gradient vector (i.e., FV encoding for the set of local descriptors X) is computed by concatenating the all u and v for all K clusters. That is,

$$f = [u_1^\top, v_1^\top, u_2^\top, v_2^\top, ..., u_K^\top, v_K^\top]^\top \qquad (8)$$

The total size of encoded vector is $2KD$, where K is the total number of clusters and D is the dimension of the descriptor. We encode our HOG, MBH$_x$ and MBH$_y$ descriptors using the above described method and fuse them using representation level fusion [20].

2.3. Gait classification

The encoded vectors are classified using Linear Support Vector Machine (SVM). SVM is considered a powerful tool for solving classification problems in many applications [23,24]. Due to the high dimensionality of our features, we decided to use SVM as a classifier. In contrast to SVM, the other similarity based classifiers like K-Nearest Neighbor and probability based classifiers such as Naive Bayes do not perform well on high dimensional features [23]. SVM first maps the training samples in high dimensional space and then extracts a hyperplane between the different classes of objects using the principle of maximizing the margin. Because of this principle, the generalization error of SVM is theoretically independent from the number of feature dimensions. We used LIBLINEAR SVM library [25] for classification.

3. EXPERIMENTS AND RESULTS

The performance of the proposed method is evaluated on three popular benchmark gait recognition databases: TUM GAID database [17], CASIA B database [26] and CASIA C dataset [27]. In all experiments, the local descriptors on each video sequence are computed using dense trajectories. The codebook size K is empirically computed and set to 32.

Table 1. Performance evaluation on TUM GAID database. Avg. is the weighted average score of each method. The best results are in bold font.

Method	N	B	S	TN	TB	TS	Avg.
GEI [17]	99.4	27.1	56.2	44.0	6.0	9.0	56.0
GEV [17]	94.2	13.9	87.7	41.0	0.0	31.0	61.4
SVIM [30]	98.4	64.2	91.6	65.6	31.3	50.0	81.4
CNN-SVM [29]	99.7	97.1	97.1	59.4	50.0	**62.5**	94.2
CNN-NN128 [29]	99.7	98.1	95.8	62.5	56.3	59.4	94.2
H2M [31]	99.4	100.0	98.1	71.9	63.4	43.8	95.5
DCS [31]	99.7	99.0	99.0	78.1	62.0	54.9	96.0
PFM [28]	99.7	99.0	99.0	**78.1**	62.0	54.9	96.0
Proposed	**99.7**	**100**	**99.7**	68.8	**71.9**	53.1	**96.5**

3.1. Results on TUM GAID database

TUM GAID is one of the largest gait databases comprising $3,370$ gait sequences of 305 subjects. It was recorded in two seasons, winter and summer, using Microsoft Kinect at 30 frames-per-second (f/s). A subset of 32 subjects participated in both seasons. Therefore, there is a substantial variation in the clothing of the participants which makes it a challenging gait database. Ten walk sequences were captured for each subject, namely normal walk (N), walk with backpack (B) and walk with coating shoes (S). Each subject in the common subset of 32 people has 10 more sequences referred to as normal walk after time (TN), walk with backpack after time (TB), and walk with coating shoes after time (TS).

The gallery and probe set division are done similarly to [17]. The first four recordings of N (i.e., $N_1 - N_4$) for each person are used as gallery set, and the sequences $N_5 - N_6, B_1 - B_2$ and $S_1 - S_2$ are used in probe set, giving three experiments namely N, B and S. In the next set of experiments labeled as TN, TB and TS, the sequences $N_7 - N_8, B_3 - B_4$ and $S_3 - S_4$ are used in probe set, while the gallery set is the same. The recognition results achieved by the proposed algorithm and other gait recognition methods on TUM GAID database are presented in Tab. 1. The proposed algorithm achieves the best results on N, B, S, and TB experiments. In TN and TS experiments PFM [28] and CNN-SVM [29] performs better than our method respectively. On average the proposed algorithm achieved the best recognition rate 96.5%.

3.2. Results on CASIA-B database

The CASIA-B gait database contains the walk sequences of 124 subjects, recorded from 11 different viewing angles in a well controlled laboratorical environment at 25 f/s. Three different variations in walking style namely normal walk (nm), walk with bag (bg) and walk with coat (cl) are recorded for each person. There are 10 walking sequences for each subject: 6 of normal walk, 2 of walk with carrying bag and 2 of walk with wearing-coat. In experiments, the first 4 out of 6

Table 2. Performance evaluation on CASIA-B database. The best results are marked in bold.

Method	nm	bg	cl	Avg.
TM [33]	97.6	52.0	32.7	60.8
Shiqi [26]	97.6	52.0	32.2	60.8
HSD [34]	94.5	62.9	58.1	71.8
M_j+ACDA [33]	100.0	91.0	80.0	90.3
DCS+H2M [31]	100.0	99.2	72.6	90.6
PFM [28]	100.0	100.0	85.5	95.2
SDL [32]	98.4	93.5	**90.3**	94.1
Proposed	**100.0**	**100.0**	86.7	**95.6**

nm sequences of each subject are used in gallery set. Three different experiments are conducted using the remaining two sequences of nm, bg and cl in probe set separately. Performance comparison of the proposed method with the state-of-the-art methods on CASIA-B database is outlined in Tab. 2. The results show that on experiment cl, SDL [32] performs better than our algorithm, while on experiments nm and bg our method achieves the best results. Overall, the proposed method achieved highest average recognition rate 95.6%.

3.3. Results on CASIA-C database

The CASIA-C database contains the gait sequence of 153 subjects with four variations: normal walk (fn), slow walk (fs), fast walk (fq), and walk with a backpack (fb). The videos were captured at night using a low resolution thermal camera at 25 f/s. Each subject has 4 sequences of fn and 2 sequences of each fs, fq and fb. A total of four experiments are conducted. In the first experiment, 3 sequences of fn are used as gallery set and the fourth fn sequence is placed in probe set. In the next three experiments, fs, fq, and fb forms the probe set, while the gallery set is same. The results achieved by the proposed algorithm and the state-of-the-art methods are presented in Tab. 3. In experiments fn, fq and fb our method achieves the best results, whereas in fs, RSM [35] performs marginally better than our algorithm. Our method achieves the best average recognition rate 99.8%.

The results presented in Tables 1-3 confirm the effectiveness of the proposed gait recognition algorithm. On all three databases, the proposed algorithms has shown very convincing results outperforming the state-of-the-art in most experiments. In particular, the average recognition performance of the proposed algorithm is the highest on all three gait databases.

Table 3. Performance evaluation on CASIA-C database. The best results are bolded.

Methods	fn	fs	fq	fb	Avg.
NDDP [36]	97.0	83.0	83.0	17.0	70.0
HSD [34]	97.0	86.0	89.0	65.0	84.2
Dadshahi et al. [37]	93.0	83.0	85.0	21.0	70.5
Tan et al. [38]	98.4	91.3	93.7	24.7	77.0
RSM [35]	100.0	**99.7**	99.6	96.2	98.9
SDL [32]	95.4	91.2	92.5	81.7	90.2
PFM [28]	100.0	98.7	100.0	99.3	99.5
Proposed	**100.0**	99.4	**100**	**99.7**	**99.8**

4. CONCLUSION

In this paper, we presented a novel model-free gait recognition algorithm which exploits the spatiotemporal characteristics of a human motion. In contrast to most existing gait recognition methods, the proposed solution does not involve any human body segmentation. The proposed method extracts dense trajectories by tracking a set of points in the successive frames of the walk sequence. Local descriptors based on MBH and HOG features are computed and encoded using Fisher vector encoding. The classification is performed using linear support vector machine. The experimental results on three popular gait benchmark databases reveal that the proposed algorithm is highly accurate.

5. REFERENCES

[1] F. Loula, S. Prasad, K. Harber, and M. Shiffrar, "Recognizing people from their movement.," *J. Exp. Psychol.-Hum. Percept.*, vol. 31, no. 1, pp. 210, 2005.

[2] A.K. Jain et al., "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004.

[3] Sarah V. Stevenage, Mark S. Nixon, and Kate Vince, "Visual analysis of gait as a cue to identity," *Applied Cognitive Psychology*, vol. 13, no. 6, pp. 513–526, 1999.

[4] M. Nixon et al., "Model-based gait recognition," in *Encyclopedia of Biometrics*. 2009, pp. 633–639, Springer.

[5] I. Bouchrika and M.S. Nixon, "Model-based feature extraction for gait analysis and recognition," in *ICCV*. Springer, 2007, pp. 150–160.

[6] Y. Chai et al., "A novel human gait recognition method by segmenting and extracting the region variance feature," in *IEEE ICPR*, 2006, vol. 4, pp. 425–428.

[7] L. Wang et al., "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, 2004.

[8] Y. Yang, D. Tu, and G. Li, "Gait recognition using flow histogram energy image," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, 2014, pp. 444–449.

[9] K. Bashir et al., “Gait representation using flow fields.,” in *BMVC*, 2009, pp. 1–11.

[10] S. Sivapalan et al., “Histogram of weighted local directions for gait recognition,” in *IEEE CVPR*, 2013, pp. 125–130.

[11] J. Little and J. Boyd, “Recognizing people by their gait: the shape of motion,” *Videre: Journal of Computer Vision Research*, vol. 1, no. 2, pp. 1–32, 1998.

[12] J. Man and B. Bhanu, “Individual recognition using gait energy image,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, 2006.

[13] K. Bashir, T. Xiang, and S. Gong, “Gait recognition without subject cooperation,” *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 2052–2060, 2010.

[14] C. Chen et al., “Frame difference energy image for gait recognition with incomplete silhouettes,” *Pattern Recognit. Lett.*, vol. 30, no. 11, pp. 977–984, 2009.

[15] C. Wang et al., “Human identification using temporal information preserving gait template,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 2164–2176, 2012.

[16] S. Sivapalan et al., “Gait energy volumes and frontal gait recognition using depth images,” in *Int. Joint Conf. Biometrics (IJCB),*, 2011, pp. 1–6.

[17] M. Hofmann, S. Bachmann, and G. Rigoll, “2.5D gait biometrics using the depth gradient histogram energy image,” in *IEEE BATS Conf.*, 2012, pp. 399–403.

[18] M. H. Khan et al., “Multiple human detection in depth images,” in *Proc. Int. Workshop Multimed. Signal Process. (MMSP)*. IEEE, 2016, pp. 1–6.

[19] H. Wang and C. Schmid, “Action recognition with improved trajectories,” in *IEEE ICCV*, 2013, pp. 3551–3558.

[20] X. Peng, L. Wang, X. Wang, and Y. Qiao, “Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice,” *Comput. Vis. Image Underst.*, vol. 150, pp. 109 – 125, 2016.

[21] J. Sánchez et al., “Image classification with the fisher vector: Theory and practice,” *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013.

[22] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in *ECCV*. Springer, 2010, pp. 143–156.

[23] M.H. Khan et al., “Automatic recognition of movement patterns in the vojta-therapy using RGB-D data,” in *Proc. Int. Conf. Image Process. (ICIP)*, 2016, pp. 1235–1239.

[24] M. H. Khan et al., “An automatic vision-based monitoring system for accurate vojta-therapy,” in *Int. Conf. Comput. Inf. Sci. (ICIS)*. IEEE, 2016, pp. 1–6.

[25] R-E. Fan et al., “Liblinear: A library for large linear classification,” *J. Mach. Learn. Res*, vol. 9, no. Aug, pp. 1871–1874, 2008.

[26] S. Yu et al., “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in *IEEE ICPR*, 2006, vol. 4, pp. 441–444.

[27] D. Tan et al., “Efficient night gait recognition based on template matching,” in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, 2006, vol. 3, pp. 1000–1003.

[28] F.M. Castro et al., “Fisher motion descriptor for multiview gait recognition,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 01, pp. 1756002, 2017.

[29] F.M. Castro et al., “Automatic learning of gait signatures for people identification,” *arXiv preprint arXiv:1603.01006*, 2016.

[30] T. Whytock, A. Belyaev, and N.M. Robertson, “Dynamic distance-based shape features for gait recognition,” *J. Math. Imaging Vis.*, vol. 50, pp. 314–326, 2014.

[31] F.M. Castro, M.J. Marín-Jiménez, and N. Guil, “Multimodal features fusion for gait, gender and shoes recognition,” *Mach. Vis. Appl.*, pp. 1–16, 2016.

[32] W. Zeng, C. Wang, and F. Yang, “Silhouette-based gait recognition via deterministic learning,” *Pattern Recognit.*, vol. 47, no. 11, pp. 3568–3584, 2014.

[33] K. Bashir, T. Xiang, and S. Gong, “Feature selection for gait recognition without subject cooperation.,” in *BMVC*, 2008, pp. 1–10.

[34] W. Kusakunniran, “Attribute-based learning for gait recognition using spatio-temporal interest points,” *Image Vis. Comput.*, vol. 32, no. 12, pp. 1117–1126, 2014.

[35] Y. Guan and C-T. Li, “A robust speed-invariant gait recognition system for walker and runner identification,” in *IEEE Int. Conf. on Biometrics (ICB)*, 2013, pp. 1–8.

[36] D. Tan, S. Yu, K. Huang, and T. Tan, “Walker recognition without gait cycle estimation,” in *Int. Conf. on Biometrics*, 2007, pp. 222–231.

[37] F. Dadashi et al., “Gait recognition using wavelet packet silhouette representation and transductive support vector machines,” in *IEEE CISP*, 2009, pp. 1–5.

[38] D. Tan, K. Huang, S. Yu, and T. Tan, “Recognizing night walkers based on one pseudoshape representation of gait,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2007, pp. 1–8.