Multiple HOG Templates for Gait Recognition

Yushu Liu[1], Junping Zhang[1], Chen Wang[1], Liang Wang[2]
[1] *School of Computer Science, Fudan University*
[2] *National Key of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences*
liuyushu1@126.com, jpzhang@fudan.edu.cn
chenwang0512@gmail.com, wangliangnlpr@gmail.com

Abstract

In gait recognition domain, template-based approaches such as Gait Energy Image (GEI) and Chrono-Gait Image (CGI) can achieve good recognition performance with low computational cost. Meanwhile, CGI can preserve temporal information better than GEI. However, they pay less attention to the extraction of local shape features. To preserve temporal information and generate more abundant local shape features, we generate multiple HOG templates by extracting Histogram of Oriented Gradients (HOG) of GEI and CGI templates. Experiments show that compared with several published approaches, our proposed multiple HOG templates can achieve better performance for gait recognition.

1 Introduction

Human identification by gaits is a promising biometric authentication technique since it is non-invasive and can be recognized at a distance. However, its performance suffers from many exterior factors such as footwear, terrain, fatigue [7].

To address these issues, model-based approaches aim to recover the underlying behavior of gait with a structure/motion model [14] [4]. However, it is not easy to quantify models for discrimination. Model-free approaches recognize human based on either gait sequences or gait templates. For example, Hidden Markov models [8] or Dynamic Time Warping [9] use gait sequence for gait recognition directly without breaking the internal temporal relationship between gait frames. However, they are computational-costly. By averaging a gait sequence into a template, Gait Energy Image (GEI) [3] achieves real-time recognition with relatively low accuracy. The recent proposed Chrono-Gait Image (CGI) [11] figured out a new template by mapping temporal information into color space of a single gait sequence which achieves higher accuracy with the same computational cost as GEI. However, both GEI and CGI pay less attention to extract the dense and local shape features from the templates which may be crucial to the improvement of gait-based identification.

In this paper, we propose to extract the dense and local shape features from GEI and CGI templates separately by utilizing Histogram of Oriented Gradient proposed by [6]. Then we project these features into two low-dimensional subspaces by using principal component analysis and linear discriminant analysis (PCA+LDA). Finally, we classify each probe gait image according to nearest neighbor rules.

With this way, we can better preserve temporal information and extract more abundant features than other template-based methods, and still achieve lower computational cost than sequence-based approaches. Experiments in a benchmark gait database show that our proposed multiple HOG templates can achieve better performance compared with other published algorithms.

2 Multiple HOG Templates

In this section, we will present the proposed Multiple HOG templates. For better understanding, we show its flow chart in Fig. 1. And the following introduction are in line with the flow chart. From the figure we can see that to obtain a template, it is necessary to subtract background from input video followed by detecting gait period of each sequences. Since the two steps have well-studied in literature and less related to our refinements, we will omit their introductions and assume that the images have been well processed in this paper.

Currently, there are two typical approaches to construct a gait template from a period of gait sequences: GEI and CGI. Given a binary gait silhouette image B_t at the t-th frame of a sequence, the gray-level GEI is $G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y)$. Here N is the number

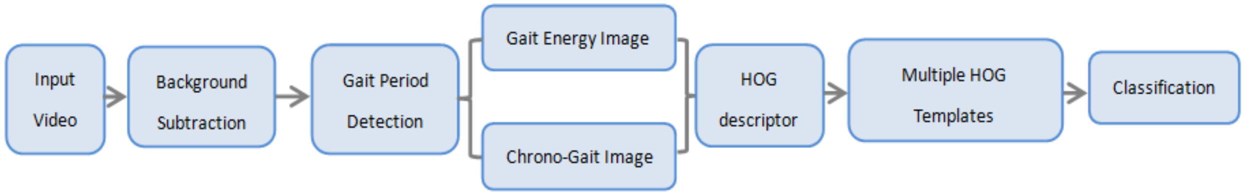

Figure 1. The Flow Charts of the Combined HOG features for Gait Recognition

of frames in a complete cycle of a silhouette sequence, and x, y are the coordinate values of the frame. Different from GEI, CGI encodes temporal information into a template via three-channel color mapping [12]. Furthermore, it selects the outer contour of each silhouette image rather than the silhouette itself to overcome the overlapping issue of gait silhouettes that will degenerate the performance of color encoding. The differences between GEI and CGI are illustrated in Fig. 2.

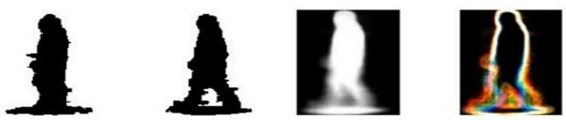

Figure 2. The left two images are key silhouettes in one walking circle. The third is GEI and the fourth is CGI.

Note that both GEI and CGI pay less attention to the extraction of local shape features, we utilize HOG operator to better characterize the local object appearances and shapes based on the distribution of local intensity gradients with oriented directions [6]. Specifically, we extract the gradient of each pixel (x, y) with $g(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2}$ according to 1-order gradient operator $[-1, 0, 1]$ and its transpose, and compute the corresponding direction with $O(x, y) = g_x(x, y)/g_y(x, y)$. After dividing the orientation into 9 bins, we generate a set of 3-D histogram features by using weighted vote based on the relationship of the orientation and two spatial directions of each pixel and its neighboring pixels. More details can be found in [6]. Note that for grayscale GEI the computation of HOG is very convenient while for CGI we get its HOG features by averaging three components in RGB space.

As illustrated in Fig. 3, the histograms of HOG on GEI and CGI are very similar in general rise and fall but different in extent. It indicates not only the consistency of HOG in preserving shape information but also the difference between GEI and CGI. As we know, CGI excels GEI in preserving the temporal information and this difference are clearly reflected from the histograms in Fig. 2. To include the spatial information and temporal information, the simplest is to take advantage of these two templates, and as illustrated before

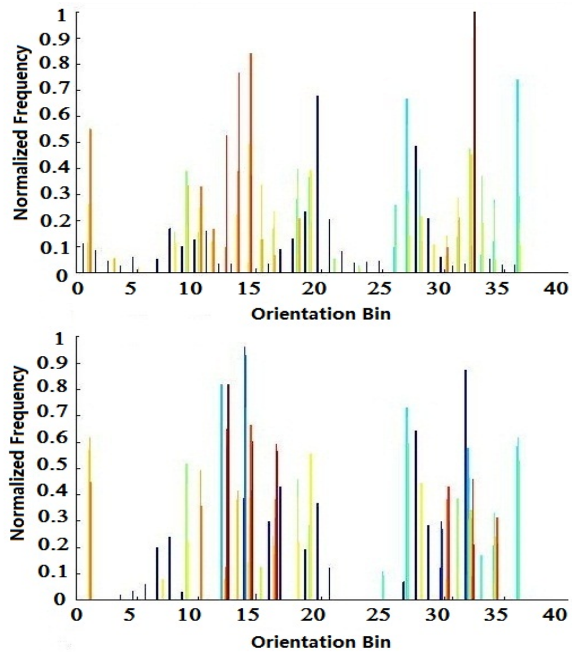

Figure 3. The Histograms of Orientation Gradients for GEI (Top) and CGI (Bottom)

HOG play an active role in keeping the local appearance and frequency information, We propose our mixed features by exploring all the advantage of the templates and descriptors as Multiple HOG Templates. Firstly we extract HOG features from GEI and CGI template. once we get these features, we perform principal component analysis following by linear discriminant analysis (PCA+LDA) to obtain a discriminant subspace. Given a set of gallery set $\mathbf{G}$, we classify the probe set $\mathbf{P}$ based on nearest neighbor rule as follows:

$$C(i) = \arg\min_j \quad \|(P^{HC}(i) - G^{HC}(j)) + (P^{HG}(i) - G^{HG}(j))\|^2$$

where $G^{HC}(j)$ and $G^{HG}(j)$ denote the low-dimensional HOG+CGI/HOG+GEI features of the j-th gallery sample, respectively. $P^{HC}(i)$ and $P^{HG}(i)$ are defined similarly for the i-th probe sample.

The advantages of our proposed algorithms are: 1) we can characterize the local object and shape appearances without losing temporal information; 2) the HOG descriptor is simple and easy to implement.

3 Experiments

We evaluate our algorithm on the USF HumanID Gait Database (silhouette version 2.1). The database consists of 122 individuals' walking data in elliptical paths on concrete and grass surface, with/without a briefcase, wearing different shoes, and sampling in elapsed time. Sarkar et al. [7] selected 122 individuals' sequences with "Grass, Shoe Type, Right Camera, No briefcase, and Time t1" for the gallery set, and developed 12 probe sets (A to K as in Tab. 3), each of them reflects specific conditions. We report Rank1 and Rank5 recognition performances in this paper. To avoid the influence from tuning the parameters in PCA and LDA, we empirically choose the same contribution ratio 0.995 for PCA and factor 1e8 for LDA for all the experiments.

We evaluate the "Rank1" and "Rank5" performances of several recent approaches including baseline algorithm (based on silhouette shape matching) [7], HMM [5], IMED+LDA [10], 2DLDA [10], MTP [1] and Tensor Locality Preserving Projections (TLPP) [2]. The Rank1 performance denotes the percentage of the correct subjects ranked first while the Rank5 performance denotes the percentage of the correct subjects appeared in any of the first five places in the rank list. We also report the average performance by computing the ratio of correctly recognized subjects to the total number of subjects [12].

From the Tab. 2 we can see that the proposed Multiple HOG templates obtain the best performance in averaging performance. A possible reason is that CGI and GEI play complement roles and HOG further improves the accuracy by extracting their local shape features.

We also investigate the influence of real, synthetic and fusion templates, which are used to enhance the robustness of gait recognition in different environment [7]. From Tab. 3 it can be seen that Multiple HOG templates achieve further improvements. It means that the proposed new templates can be more robust for different environments.

It is worth mentioning that the multiple HOG templates have outstanding performance under parameter "V" - View (0^o, 45^o, 90^o and 135^o), and perform not so impressive in probe set D, E, F sharing a common environmental parameter "S" - surface (see Tab. 3). It indicates that our method can deal with different views well and has slightly shortage to deal with surface environment.

4 Conclusion

In this paper, we propose multiple HOG features based on CGI and GEI templates to characterize the local shape and temporal information of gait sequence. Experiments on a benchmark database show that the proposed templates can attain better performance than other published algorithms. In the future, we will study how to produce more real and synthetic HOG templates for further improvement. Furthermore, how to better preserve the temporal information of CGI in our templates deserves studying.

References

[1] C. Chen, J. Zhang, and R. Fleischer. Multilinear tensor-based non-parametric dimension reduction for gait recognition. *ICB*, pages 1030–1039, 2009.

[2] C. Chen, J. Zhang, and R. Fleischer. Distance approximating dimension reduction of riemannian manifolds. *IEEE TSMCB*, 40(1):208–217, 2010.

[3] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE TPAMI*, 28(2):316–322, 2006.

[4] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A.K.Roy-Chowdhury, V.Kruger, and R. Chellappa. Indentification of humans using gait. *ICCV*, pages 1163 – 1173, 2004.

[5] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE TIP*, 13:1163–1173, 2004.

[6] D. N, I. Rhone-Alps, and B. Montbonnot Triggs. Histograms of oriented gradients for human detection. *CVPR*, pages 886 – 893, 2005.

[7] S. Sarkar, P. K. Phillip, Z. Liu, I. Vega, P. Grother, and K. V.Browyer. The humanid gait challenge problem: datasets, performance and analysis. *IEEEE TPAMI*, 27(2):162–177, 2005.

[8] A. Sundaresan, A. RoyChowdhury, and R. Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. *ICIP*, 2003.

[9] G. T and W. R. Differential radon transform for gait recognition. *ICASSP*, pages 834–837, 2010.

[10] D. Tao, X. Li, X. Wu, and S. J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE TPAMI*, 29(10):1700–1715, 2007.

[11] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang. Chrono-gait image: A novel temporal template for gait recognition. *ECCV*, pages 257–270, 2010.

[12] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE TPAMI*, in press, 2012.

[13] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. jiang Zhang. Human gait recognition with matrix representation. *IEEE TCSVT*, 16(7):896–903, 2006.

[14] C. Yam and M. S. Nixon. Model-based gait recognition. *Encyclopedia of Biometrics*, pages 633–639, 2009.

Experiment Label	A	B	C	D	E	F	G	H	I	J	K	L
Size of the Probe Set	122	54	54	121	60	121	60	120	60	120	33	33
Gallery/Probe Difference	V	H	VH	S	SH	SV	SHV	B	BS	BV	THC	TS

V-View, H-Shoe, S-Surface, B-Briefcase, T-Time, and C-Clothing

Table 1. Twelve experiments designed for individual recognition in USF HumanID Database.

RANK1	A	B	C	D	E	F	G	H	I	J	K	L	Avg
Baseline [3]	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM [7]	89	88	68	35	28	15	21	85	80	58	17	15	53.54
IMED+LDA [13]	88	86	72	29	33	23	32	54	62	52	8	13	48.63
2DLDA [10]	89	93	80	28	33	17	19	74	71	49	16	16	50.98
TLPP [5]	87	93	72	25	35	17	18	62	62	43	12	15	46.95
MTP [2]	90	91	83	37	43	23	25	56	59	59	9	6	51.57
GEI	84	87	69	19	18	10	13	54	55	40	9	3	39.01
HOG_GEI	92	89	82	31	32	23	23	92	82	68	12	6	52.63
CGI	85	87	78	38	35	23	18	93	80	60	9	9	51.27
HOG_CGI	92	89	83	34	38	22	25	88	82	69	3	3	57.31
Multi-HOG Templates	96	91	83	33	33	18	25	91	82	82	9	6	**59.39**
RANK5	A	B	C	D	E	F	G	H	I	J	K	L	Avg
Baseline [3]	88	93	78	66	55	42	38	85	78	62	12	5	64.54
HMM [7]	-	-	-	-	-	-	-	-	-	-	-	-	-
IMED+LDA [13]	95	95	90	52	63	42	38	85	78	62	21	19	68.60
2DLDA [10]	97	93	93	57	59	39	47	91	94	75	37	34	70.95
TLPP [5]	94	94	87	52	55	35	42	85	78	68	24	33	65.18
MTP [2]	94	93	91	64	68	51	52	88	83	82	18	15	71.38
GEI	92	94	93	45	53	29	37	77	77	69	15	15	58.00
HOG_GEI	98	94	91	60	48	43	45	96	93	87	24	24	67.11
CGI	94	94	87	64	52	41	45	96	92	87	18	18	65.66
HOG_CGI	98	94	91	65	55	42	47	97	93	93	30	18	74.11
Multi-HOG Templates	98	94	93	66	52	44	47	96	93	93	30	21	**74.32**

Table 2. The Rank1 and Rank5 performances of different features on the USF Gait Dataset

Rank1	A	B	C	D	E	F	G	H	I	J	k	L	Avg
GEIreal	88	87	76	28	27	17	15	58	57	44	9	6	45.51
GEIsyn	83	91	70	19	20	10	15	49	45	31	9	6	38.83
GEIfusion	89	94	76	41	42	23	23	60	63	60	9	6	48.90
HOGGEIreal	94	93	83	33	30	16	16	85	80	68	15	3	55.74
HogGEIsyn	81	91	61	40	33	27	20	92	83	58	15	6	55.32
HogGEIfusion	93	93	87	36	33	20	25	88	80	68	12	6	57.83
CGIreal	90	89	82	28	30	15	13	83	75	60	3	3	51.98
CGIsyn	84	87	67	27	28	18	15	63	55	41	12	6	44.89
CGIfusion	84	87	80	41	38	29	35	78	67	57	6	12	55.11
HogCGIreal	93	94	87	28	35	11	17	87	82	74	12	9	56.26
HogCGIsyn	85	87	70	40	42	22	25	84	83	63	3	6	55.64
HogCGIfusion	92	94	83	33	33	17	18	88	87	64	12	6	56.47
Multi-HOG Templates real	94	93	85	35	40	18	27	86	73	68	9	6	57.31
Multi-HOG Templates syn	82	91	63	37	38	25	25	85	78	53	3	9	53.44
Multi-HOG Templates fusion	90	94	76	37	40	28	32	87	77	70	9	6	**58.77**
Rank5	A	B	C	D	E	F	G	H	I	J	k	L	Avg
GEIreal	94	94	89	57	58	36	42	83	82	73	12	12	65.64
GEIsyn	93	94	89	40	40	30	33	69	70	63	24	21	58.04
GEIfusion	96	94	96	69	68	53	53	84	80	78	21	30	68.72
HOGGEI	98	94	93	57	57	43	47	96	95	88	30	21	72.96
HogGEIsyn	92	93	87	66	60	49	48	96	93	82	27	27	73.07
HogGEIfusion	99	94	94	63	55	44	45	96	95	88	27	27	73.80
CGIreal	96	94	93	64	62	46	50	94	93	85	27	33	73.90
CGIsyn	89	89	82	53	52	37	48	80	87	72	30	30	65.41
CGIfusion	90	93	91	65	63	48	52	89	90	82	27	27	72.13
HogCGIreal	99	96	96	61	60	44	43	96	97	88	27	24	74.11
HogCGIsyn	90	91	85	61	57	48	52	93	88	80	24	24	70.88
HogCGIfusion	99	96	96	58	57	37	45	96	95	86	27	15	72.03
Multi-HOG Templates real	98	96	96	65	58	48	43	92	93	89	30	24	74.32
Multi-HOG Templates syn	94	94	83	65	55	46	52	93	90	80	18	30	71.71
Multi-HOG Templates fusion	98	96	94	66	55	51	48	93	92	87	27	21	**74.53**

Table 3. The Rank1 and Rank5 performances of Gait Recognition on USF real templates