IMPROVED GAIT RECOGNITION USING GRADIENT HISTOGRAM ENERGY IMAGE

Martin Hofmann, Gerhard Rigoll

Institute for Human-Machine Communication, Technische Universität München, Germany
{martin.hofmann, rigoll}@tum.de

ABSTRACT

We present a new spatio-temporal representation for Gait Recognition, which we call Gradient Histogram Energy Image (GHEI). Similar to the successful Gait Energy Image (GEI), information is averaged over full gait cycles to reduce noise. Contrary to GEI, where silhouettes are averaged and thus only edge information at the boundary is used, our GHEI computes gradient histograms at all locations of the original image. Thus, also edge information inside the person silhouette is captured. In addition, we show that GHEI can be greatly improved using precise segmentation techniques (we use α-matte segmentation). We demonstrate great effectiveness of GHEI and its variants in our experiments on the large and widely used HumanID Gait Challenge dataset. On this dataset we reach a significant performance gain over the current state of the art.

Index Terms— Biometrics, Gait Recognition, Histogram of Oriented Gradients, Gradient Histogram Energy Image

1. INTRODUCTION

Many current person identification systems rely on close proximity of the subject to a sensing device and often also require cooperation of the person to be identified. At a distance, however, many typical *physiologic* features, such as fingerprint, DNA, hand, ear, retina and face, are obscured or cannot be obtained at all. By contrast, *behavior based* features such as gait features can be extracted from walking people at a distance and without the person's cooperation.

A multitude of gait recognition algorithms (see Table 1) have so far been proposed. The most successful of currently used methods are based on the Gait Energy Image (GEI) [1]. In GEI, binarized silhouettes are extracted and averaged over full gait cycles. While this binarization and averaging reduces noise, it also destroys a lot of the available information. We therefore present the Gradient Histogram Energy Image (GHEI) which not only captures edges at the silhouette boundaries, but also within the person's silhouette. We observe that good foreground segmentation has a significant impact on GHEI performance. By precise segmentation of the person from the background, recognition rates can be further improved. Experiments are run on the widely used HumanID Gait database [2].

2. RELATED WORK

There are two kinds of gait recognition methods: Model-based methods model-free methods. Model-based methods are typically very demanding and good results are hard to achieve. Model-free methods [1][2][3][4][5][6][7][8] on the other hand have shown more success in the recent past. Here, the person identity is directly inferred from the features without an intermediate person model. Most methods build on a silhouette extraction for each frame in a gait cycle. Silhouettes are either averaged (as in Gait Energy Image) [1][5][7], or all silhouettes are used simultaneously [4][2]. Different classifiers ranging from nearest neighbor [1] to SVM and HMM [4] have been applied with similarly good results.

3. THE GRADIENT HISTOGRAM ENERGY IMAGE

Human walking is generally regarded as a periodic motion. Therefore, it can be assumed that all gait information is captured within a full gait cycle. A prominent exponent of this approach is the Gait Energy Image [1], where all information withing a gait cycle is averaged. This kind of gait signature representation has proven to be very robust and efficient. First, using background modeling, silhouettes are extracted at each frame and are tracked over time. Using a preprocessing step, the resulting bounding boxes are then normalized to have the same size. Also, persons are horizontally aligned to be in the center of the the bounding box. Many gait recognition methods use all silhouettes within a gait cycle (e.g. by analyzing their mutual correlation). By contrast, in Gait Energy Image, the aligned silhouettes S_t within a gait cycle are averaged and are thus compressed to a single image: $G(x,y) = \frac{1}{T}\sum_{t=1}^{T} S_t(x,y)$. While in this averaging step, information is seemingly lost, it has also proven to greatly reduce noise from imprecise foreground segmentation.

3.1. GHEI Representation

Our GHEI representation therefore builds on the averaging concept used in GEI. However, we note that in standard GEI, binary silhouettes are averaged. Since these silhouettes can only capture edge information at the boundary, a lot of information is being discarded. Therefore, with GHEI, we present

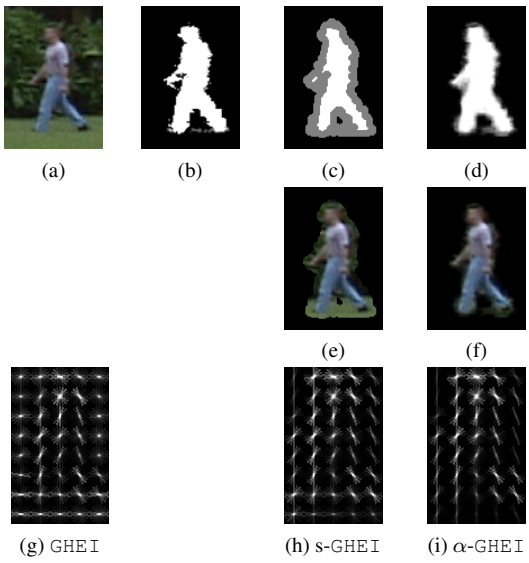

Fig. 1: (a) input image; (b) foreground segmentation using Gaussian mixture models; (c) tri-state labeling using morphologic operations; (d) alpha matte; (e) coarse segmentation based on tri-state labeling; (f) precise segmentation based on alpha matte; (g) GHEI calculated directly from input images; (h) s-GHEI calculated on coarse segmentation; (i) α-GHEI calculated on precise α-matte segmentation

a method that also captures edges inside the person's silhouette. This is done by extracting gradients at all locations and aggregating this information into orientation histograms. The process of GHEI extraction is detailed as follows:

First, at each pixel of the tracked bounding box I, magnitude r and orientation θ of intensity gradients are computed:

$$r(x,y) = \sqrt{u(x,y)^2 + v(x,y)^2} \quad (1)$$

$$\theta(x,y) = \text{atan2}(u(x,y), v(x,y)) + \pi \quad (2)$$

with $u(x,y) = I(x-1,y) - I(x+1,y)$ and $v(x,y) = I(x,y-1) - I(x,y+1)$. Gradient orientations at each pixel are discretized into 9 orientations:

$$\hat{\theta}(x,y) = \left\lfloor \frac{9 \cdot \theta(x,y)}{2\pi} \right\rfloor \quad (3)$$

These discretized gradient orientations are then aggregated into a dense grid of non-overlapping square image regions, the so called "cells" (each containing typically 8×8 pixels). Each of these cells is thus represented by a 9-bin histogram of oriented gradients. Finally, each cell is normalized four times (by blocks of four surrounding cells each) leading to $9 \cdot 4 = 36$ values for each cell. (Details to be found in [9]).

At each frame t in a gait cycle, a gradient histogram descriptor $h_t(i,j,f)$ is computed on the size and position normalized RGB-image inside the bounding box. Here, i and j are pointing to the histogram cell at position (i,j) and $f = \{1 \ldots 36\}$ is the index to the histogram bin. The GHEI is then obtained by averaging the gradient histogram representations over a full gait cycle consisting of T frames:

$$H(i,j,f) = \frac{1}{T} \sum_{t=1}^{T} h_t(i,j,f) \quad (4)$$

Each gait cycle is finally represented by a multidimensional feature vector $H(i,j,f)$.

In summary, the Gradient Histogram Energy Image representation can therefore be seen as a combination of Histograms of Oriented Gradients (HOG) and Gait Energy Image (GEI). The GHEI leverages the averaging and noise reducing step of GEI with the advantages of gradient histograms.

3.2. Improvement using Segmentation

First experiments with the basic GHEI showed good results. However, we notice that background information is not fully averaged out over one gait cycle and degrades recognition performance. Recognition can therefore be improved using good foreground segmentation.

For precise segmentation we use automated α-matte segmentation which was previously applied to GEI [10]: Typical foreground segmentation methods (we use Gaussian mixture models [11]) all lead to a noisy, binary segmentation as depicted in Figure 1b. However, due to the nature of the image capturing, there is a band on the silhouette which belongs partially to foreground and partially to background. Thus, at each pixel (x,y), the image I can be modeled as a linear composition of the foreground F and the background B:

$$I(x,y) = \alpha(x,y)F(x,y) + (1 - \alpha(x,y))B(x,y) \quad (5)$$

Here, $\alpha(x,y)$ is the opacity of the pixel at (x,y). $F(x,y)$, $B(x,y)$ and $\alpha(x,y)$ are unknown. For a typical color image with three color channels we thus have $3+3+1 = 7$ unknowns to solve for at each pixel. This is the typical matting problem. To leverage the high number of unknowns, proximity and smoothness assumptions are made. In addition, the typical matting application has a human in the loop who has to provide some scribbles for foreground and background, leading to the so called *trimap*. This map contains regions which are definitely foreground ($\alpha(x,y) = 1$), some which are definitely background ($\alpha(x,y) = 0$) and some unknown regions for which the matting method determines the $\alpha(x,y)$.

For automated gait recognition it is infeasible to have a human in the loop. We therefore automatically generate the trimap from the noisy foreground segmentation. We get the definite-foreground regions ($\alpha(x,y) = 1$) by eroding the foreground segmentation with a circular structure element with radius $r = 4$. The definite-background regions are obtained by eroding the background region with the same circular structure element. The resulting trimap is shown in Figure 1c.

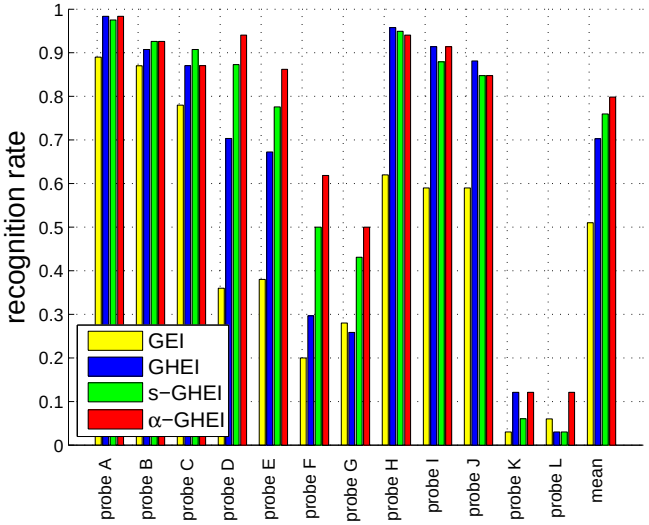

Fig. 2: Quantitative results of GHEI on the HumanID Gait database [2] compared to standard GEI. (1) GEI (for reference); (2) basic GHEI, (3) s-GHEI with coarse segmentation, (4) α-GHEI with precise α-matte segmentation

To solve the alpha matting problem, we use closed form matting [12]. The resulting foreground segmentation – the alpha-matte – is depicted in Figure 1d. It can be seen that this segmentation is superior to the initial background segmentation. Holes are closed, erroneous pixels are removed and most of all, the smooth transition from the foreground to the background is captured. Furthermore, by $F(x, y) = I(x, y) \cdot \alpha(x, y)$ we can approximate a precise color segmentation of the foreground object (see Figure 1f).

3.3. Variants

We define three variants of GHEI. In basic GHEI (Figure 1g), the HOG descriptors are calculated directly on the sequence of color input images (Figure 1a). The segmented GHEI (s-GHEI) (Figure 1h) is calculated on a coarse foreground segmentation (Figure 1e) which is generated using morphologic dilate. Here, most of the background information is removed, but some is still present. Finally in α-GHEI (Figure 1i), the HOG features are calculated on precise α-matte segmentation (Figure 1f).

4. PERSON IDENTIFICATION USING GHEI

For the pattern recognition and classification part of the gait recognition system we use the same strategy used in the GEI system [1]. Dimension reduction is done by Principal Component Analysis followed by Linear Discriminant Analysis (PCA+LDA). Classification is done using nearest-neighbor.

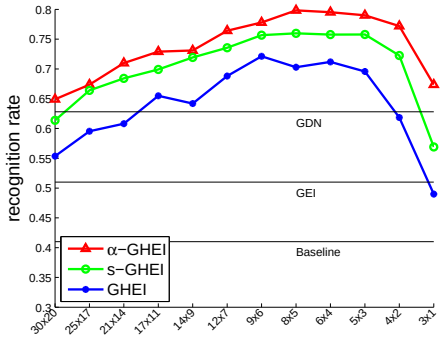

Fig. 3: Influence of the number of HOG cells on the average recognition rate for α-GHEI, s-GHEI and GHEI. Baseline, Gait Energy Image (GEI) as well as Gait Dynamics Normalization (GDN) (representing the current state of the art) are plotted for reference

5. RESULTS AND COMPARISON

For performance evaluation, many databases have been recorded. One of the most widely used databases is the HumanID Gait database [2]. This database features video sequences of a total of 122 subjects, which walk perpendicular to the camera at a distance. Twelve experiments (A to L) have been defined to evaluate the influence of view, shoes, surface, briefcase and clothing. Many current gait recognition methods have been applied to the HumanID Gait database which makes it an ideal benchmark for competitive comparison.

5.1. Results of GHEI

Figure 2 shows the quantitative results of the three proposed GHEI variants compared to GEI. The GEI system is similar, since the same classifier is used. Synthetic templates as in [1] are not used. It can be seen that all variants of GHEI greatly outperform the standard GEI. In fact, the basic GHEI (with no segmentation) outperforms GEI by 38% (relative) on average recognition rate. Using segmentation, the α-GHEI even beats standard GEI by 56%. With precise segmentation, α-GHEI is 14% better than standard GHEI. Thus, using good segmentation is quite important for good recognition rates.

5.2. Influence of the number of HOG cells

Figure 3 shows the influence of the number of HOG cells on the recognition rate. For a large range, all three GHEI variants outperform the state of the art. In the range from $[9 \times 6]$ to $[5 \times 3]$ recognition rates are approximately constant, which shows that GHEI is not very sensitive to the number of cells. Interestingly, for person detection, similar sizes are used [9]. Thus, both detection and identification can be carried out on the same HOG data. For our comparative evaluation we use patches of $[8 \times 5]$.

5.3. Comparison to other methods

For performance evaluation, we compare GHEI to several state-of-the-art results. Summarizing results are shown in Table 1. Here, recognition rates for all twelve experiments (A to L), as well as the weighed recognition averages are shown.

It can be seen that, in almost all of the twelve experiments, all three variants of GHEI perform comparably to other methods or even outperform them. For simple experiments (A to C), GHEI slightly outperforms most other methods. Major performance increases can be seen for the experiments with surface change (exp. D to G). Also for experiments with briefcase (H to J), GHEI greatly outperforms most other methods. This shows that GHEI is very capable of handling surface and briefcase changes. On the most difficult experiments (K and L, with time and cloth variations), GHEI can achieve an average recognition rate compared to other methods.

Variations	view, shoe			surface, view, shoe				briefcase, view, shoe			time, clothing		
Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	avg.
Probe Size	122	54	54	121	60	121	60	120	60	120	33	33	-
Baseline [2]	73	78	48	32	22	17	17	61	57	36	3	3	41.0
IMDE [7]	75	83	65	25	28	19	16	58	60	42	2	9	42.9
IMDE+LDA [7]	88	86	72	29	33	23	32	54	62	52	8	13	48.6
GEI [1]	89	87	78	36	38	20	28	62	59	59	3	6	51.0
HMM [4]	89	88	68	35	28	15	21	85	80	58	17	15	53.5
α-GEI [10]	89	87	79	30	36	21	19	83	69	63	6	6	53.6
GEI+Synth [1]	90	91	81	56	64	25	36	64	60	60	6	15	57.7
DATER [13]	89	93	80	44	45	25	33	80	79	60	18	21	58.5
MMFA [8]	89	**94**	80	44	47	25	33	85	83	60	**27**	21	59.9
GTDA [6]	91	93	86	32	47	21	32	95	90	68	16	19	60.6
I-to-C [3]	93	89	81	54	52	32	34	81	78	62	12	9	61.2
GDN [5]	85	89	72	57	66	46	41	83	79	52	15	**24**	62.8
α-GEI+face [10]	93	85	81	56	45	38	31	89	90	82	3	6	65.2
GHEI	**98**	91	87	70	67	30	26	**96**	**91**	**88**	12	3	70.2
s-GHEI	**98**	93	**91**	87	78	50	43	95	88	85	6	3	76.1
α-GHEI	**98**	93	87	**94**	**86**	**62**	**50**	94	**91**	85	12	12	**79.8**

Table 1: Comparison of GHEI to other methods (all rank 1): Baseline [2]; IMage Euclidean Distance (IMED) [7]; Gait Energy Image (GEI) [1]; Hidden Markov Models (HMM) [4]; Alpha Gait Energy Image (α-GEI) [10]; Discriminant Analysis with Tensor Representation (DATER) [13]; Marginal Fisher Analysis (MMFA) [8]; General Tensor Discriminant Analysis (GTDA) [6]; Image-to-Class Distance (I-to-C) [3]; Gait Dynamics Normalization (GDN) [5]

6. CONCLUSION AND OUTLOOK

In this work we presented a new and highly efficient feature extraction method for person identification. By taking HOG features instead of silhouettes in the Gait Energy Image representation, a basic pattern recognition framework easily outperforms the current state of the art. It can be foreseen that when combined with more sophisticated pattern recognition techniques (which have already been tested with standard GEI), recognition rates can be improved even further. Based on our results, we encourage researchers to replace the GEI representation in their systems by GHEI representation. In many cases this might lead to new and improved gait recognition performance.

The limits of GHEI feature extraction have thus by far not yet been reached. It will be interesting to see how GHEI representation will stand the test on other databases. For example in multi-view gait recognition, in clothing invariant recognition, or in speed invariant recognition.

7. REFERENCES

[1] Ju Han and Bir Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 316–322, 2006.

[2] Sudeep Sarkar, P. Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 162–177, 2005.

[3] Yi Huang, Dong Xu, and Tat-Jen Cham, "Face and human gait recognition using image-to-class distance," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 3, pp. 431 –438, march 2010.

[4] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. RoyChowdhury, and V. Krueger, "Identification of humans using gait," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163–1173, 2004.

[5] Zongyi Liu and Sudeep Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 863–876, 2006.

[6] Dacheng Tao, Xuelong Li, Xindong Wu, and S.J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 10, pp. 1700 –1715, oct. 2007.

[7] Liwei Wang, Yan Zhang, and Jufu Feng, "On the euclidean distance of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1334 –1339, aug. 2005.

[8] Dong Xu, Shuicheng Yan, Dacheng Tao, S. Lin, and Hong-Jiang Zhang, "Marginal fisher analysis and its variants for human gait recognition and content- based image retrieval," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2811 –2821, nov. 2007.

[9] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, june 2005, vol. 2, pp. 886–893.

[10] Martin Hofmann, Stephan M. Schmidt, AN. Rajagopalan, and Gerhard Rigoll, "Combined face and gait recognition using alpha matte preprocessing," *IAPR/IEEE International Conference on Biometrics*, march 2012.

[11] Chris Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 2246–2252, 1999.

[12] Anat Levin, Dani Lischinski, and Yair Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 228–242, 2008.

[13] Dong Xu, Shuicheng Yan, Dacheng Tao, Lei Zhang, Xuelong Li, and Hong-Jiang Zhang, "Human gait recognition with matrix representation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 7, pp. 896 –903, july 2006.