EXPLOITING GRADIENT HISTOGRAMS FOR GAIT-BASED PERSON IDENTIFICATION

Martin Hofmann, Gerhard Rigoll

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

ABSTRACT

In this paper, we exploit gradient histograms for person identification based on gait. A traditional and successful method for gait recognition is the Gait Energy Image (GEI). Here, person silhouettes are averaged over full gait cycles, which leads to a robust and efficient representation. However, binarized silhouettes only capture edge information at the boundary of the person. By contrast, the Gradient Histogram Energy Image (GHEI) also captures edges within the silhouette by means of gradient histograms. Combined with precise α-matte preprocessing and with a new part-based extension, recognition performance can be further improved. In addition, we show, that GEI can even be outperformed by directly applying gradient histogram extraction on the already binarized silhouettes. We run all experiments on the widely used HumanID gait database and show significant performance improvements over the current state of the art.

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

1. INTRODUCTION

Gait is an important biometric modality of recognizing humans. Especially at large distances, where other modalities such as face and fingerprint cannot be extracted, gait signatures are still available and can be obtained from low resolution video streams and without the person's cooperation. Thus, there is growing interest in gait recognition approaches and a multitude of recent methods and approaches have shown the capability of gait as a biometric.

Many of the most successful gait recognition approaches are probably those that are based on the Gait Energy Image (GEI) [1]. In GEI, binarized silhouettes are extracted at each frame and are averaged over full gait cycles (see Fig. 1a). While this binarization and averaging reduces noise, it also destroys a lot of the available information. Due to the extraction of binarized silhouettes, in essence, only edge information at the boundary is captured. Any edges or gradients inside the silhouette are completely discarded. For example, the arm swinging in front of the person is largely lost due to the binarized silhouette. An efficient way to overcome this limitation of GEI is the Gradient Histogram Energy Image (GHEI) [2]. Here, gradient histograms are extracted at all locations without any early binarization, then the histograms are averaged. This method captures gradients also within the person's silhouette.

In this paper we present improvements by using the robust and efficient gradient histogram representation, which significantly outperform standard GEI. Additional improvements can be reported when using precise α-matte segmentation, as well as precise part localization for body parts like head, torso and legs. The resulting α-part-based Gradient Histogram Energy Image (α-pb-GHEI) significantly outperforms the full-body approach. Furthermore, to directly assess the power of gradient histograms, we perform an additional investigation by applying GHEI directly on the binarized silhouettes and show that even in this case GEI is outperformed (even though not as much as using α-pb-GHEI). We test our algorithms on the popular and widely used HumanID Gait database [3].

2. RELATED WORK

Generally speaking, gait recognition methods can be categorized into model-based and model-free approaches. In model-based methods [4][5], in a first step, a human body model is fitted to the input data. Recognition is then performed based on the model parameters or the change of these model parameters. However, in practice, fitting a body model has turned out to be extremely difficult and fitting results are too noisy to be used for individual identification. In contrast, model-free methods [1][2][3][6][7][8][9][10][11][12][13][14][15][16] directly extract features from the input data without an intermediate human body model and thus a robust statistical person model can be built. Due to its robustness and efficiency, most current methods, including ours, are model-free.

It is interesting to note that almost all of the successful gait recognition methods rely more or less on the silhouette averaging idea of the Gait Energy Image (GEI) [1]. For a robust recognition system, apparently reducing noise by taking the mean of gait cycles outweighs the loss of data which comes with averaging. Histograms are a very good way to handle noise. In [17], similar to our approach, the authors use Histograms of Oriented Gradient (HOG) to model the input data. However, in their approach, they run HOG only once on the Gait Energy Image. This is contrary to our approach, since we calculate HOG on each frame separately and average the resulting gradient histograms.

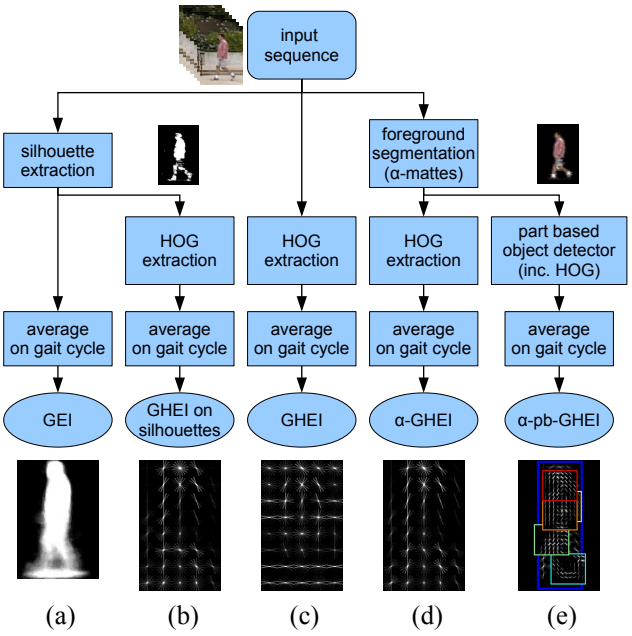

Fig. 1: Extraction of the different variations of Gradient Histogram Energy Images (GHEI) compared to the Gait Energy Image (GEI).

3. THE GRADIENT HISTOGRAM ENERGY IMAGE

The Gradient Histogram Energy Image (GHEI) is in essence a combination of Gait Energy Image (GEI) and Histograms of Oriented Gradients (HOG).

The idea is driven by the observation that in GEI, edge information inside the body silhouette is lost due to the binarization in the foreground segmentation step. Histograms of Oriented Gradients have proven to be a highly efficient way of capturing edges. Therefore, combining the basic idea of GEI with the edge capturing capabilities of HOG leads to robust and efficient feature extraction.

The process of gradient histogram calculation can be 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 [18]). Thus, the final HOG description of a patch containing $m \times n$ cells has dimension of $m \times n \times 36$.

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 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 $H(i,j,f)$.

3.1. Precise Foreground Segmentation

Averaging over gradient histograms also means that gradients in the background are averaged. It is assumed, since the background is changing within the tracked bounding box that background gradients would average out over a full gait cycle. However, in practice this is only partly the case. There are still distinct and high-magnitude edges in the background, which degrade the GHEI and lead to worse results. Figure 1 (on top) shows a typical input sequence with disturbing background. The resulting GHEI (Figure 1c) captures the major gradients, however, also background edges have a high impact on the gradient histograms.

An appropriate way to circumvent this issue is by setting all background pixels to zero before calculating the gradient histograms. As proposed in [19], this is done by precise background modeling using α-matte segmentation. The α-matte segmentation (we use closed form matting [20]) is an efficient and precise way to segment the foreground from the background. This way, all the background pixels are discarded, while all edges in the foreground can be captured in the gradient histograms. Figure 1 (top right) shows the result after α-matte segmentation. The final α-GHEI descriptor (Figure 1d) is much cleaner and shows more relevant gradients.

3.2. Part-Based Gradient Histogram Energy Image

For typical gait recognition methods, first the person has to be detected and tracked. However, in most current gait recognition approaches, detection and tracking were never the focus. Typical databases usually show only one person at a time and the background is relatively controlled. So standard background modeling followed by a simple blob tracker is sufficient for person localization. For GHEI extraction, as described above, the same localization method was used.

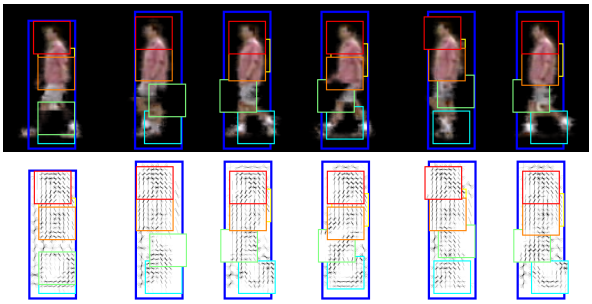

Fig. 2: Top row: Results of a part-based person detector on a sequence of frames. Here, foreground is already segmented using α-matte segmentation. Bottom row: Corresponding HOG features for all parts.

Since the HOG features are typically used in object detection, it is possible to simultaneously use the HOG features for both detection and identification using GHEI. Thus, instead of blob-tracking, we employ a tracking-by-detection approach using a HOG-based person detector.

This has the major advantage that at each frame, the detector precisely locates the position and the scale of the person. This precise detection is often not guaranteed in the commonly used blob tracker, where erroneous background segmentation can lead to bad position and scale detection.

Recently, part-based object detectors have received remarkable attention [21][22]. Here not only one sliding window (root) is used for object search, but in addition, multiple smaller object windows (parts) are used. The joint appearance and relative location of the parts with respect to the root part are used for object detection. In case of a person model, this means that body parts like head, torso, hips and feet are separately detected and located within the image.

We leverage this separate detection and localization of body parts for the purpose of gait feature extraction. First, HOG features h_t^k are generated for all body parts, as well as for the root part. Gradient Histogram Energy Images H^k are then calculated for the root window r and all l part windows $p_1 \dots p_l$ by averaging over a full gait cycle:

$$H^k(i,j,f) = \frac{1}{T}\sum_{t=1}^{T} h_t^k(i,j,f) \qquad k \in \{r, p_1, \dots, p_l\} \quad (5)$$

In addition, the final feature vector also contains the mean of the relative positioning x_k of each part with respect to the root part: $X^k = \frac{1}{T}\sum_{t=1}^{T} x_k$. This way, a rough posture of the person can be captured. Finally, all data is concatenated to form the final α-pb-GHEI feature vector:

$$H_{\alpha\text{-pb-GHEI}} = \left[H^{root}, H^{p_1}, \dots, H^{p_l}, X^{p_1}, \dots, X^{p_l}\right] \quad (6)$$

The part-based object detection and the corresponding part-based HOG features are depicted in Figure 2. After averaging over a full gait cycle, a α-pb-GHEI such as shown in Figure 1e) is obtained.

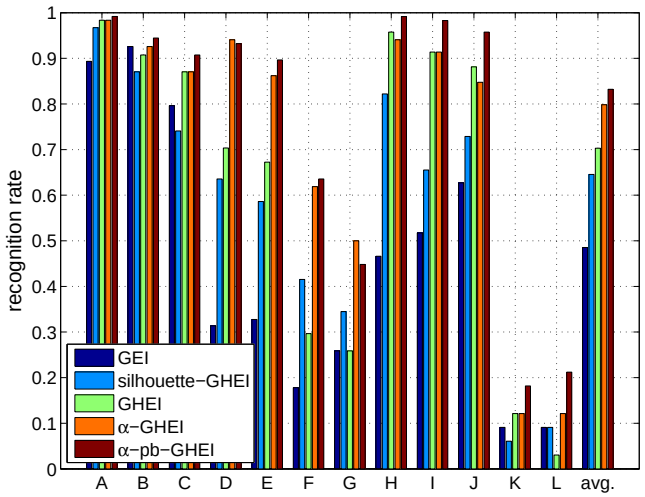

Fig. 3: Results of the GHEI and the three variants α-GHEI, silhouette-GHEI, α-pb-GHEI compared to baseline Gait Energy Image (GEI).

4. EXPERIMENTS

To test the GHEI features, we use an identification system similar to the one presented in [1]. Here, dimension reduction is done by Principal Component Analysis followed by Linear Discriminant Analysis (PCA+MDA). Classification is done using nearest neighbor. This combination of dimension reduction and classifier has proven highly effective for problems with a small amount of training data, which is typical for the gait recognition problem.

To evaluate the effectiveness of the proposed Gradient Histogram Energy Image and its variants, we run experiments on the widely used USF HumanID gait database [3]. In this database, a total of 122 subjects walk in elliptical paths with multiple variations. These variations include surface, briefcase, shoes and sampling time. In [3], 12 experiments (A to L) using different combinations of variations have been defined. Experiment A-C are the easiest with only view and shoe type changing. Experiments D-G in addition have surface variation. In experiments H-J, in addition to view and shoe type, subjects carry a briefcase. For the most challenging setups (experiment K-L), probe samples were recorded significantly after the gallery samples and subjects were wearing different clothes.

4.1. Performance of Part-Based Gradient Histogram Energy Image

Figure 3 shows the performance of α-pb-GHEI with respect to the GEI baseline and the other GHEI variants on all twelve experiments of the HumanID gait database. It can be seen, that α-pb-GHEI greatly outperforms all these methods in almost all experimental settings. The α-pb-GHEI uses the same

α-matte preprocessing as used in α-GHEI, but represents an extension using precise localization of the body and its parts. It can be seen that, compared to GHEI, α-pb-GHEI is better capable of capturing the person's identity even in cases of severe changes in visual appearances. This can especially be observed in cases with the briefcase (experiments H-J), where a high relative performance gain can be reported. The part and root localization is thus capable of reliably capturing relevant parts and does not get diverted by local but strong degradations like the briefcase. Thus, no explicit removal of carried objects (such as done in [23]) has to be used.

4.2. Gradient Histogram Energy Image on Binary Silhouettes

The idea of Gradient Histogram Energy Image is to capture edges not only at the silhouette boundary, but also inside the silhouette. In general, gradient extraction is therefore performed on the RGB sequence. As a reference experiment, we applied GHEI extraction directly on binarized silhouettes which were obtained using a simple (and noisy) Gaussian Mixture Model extraction. As shown in Figure 3, it can be observed that the resulting silhouette-GHEI still greatly outperforms standard GEI, especially in scenarios with change in surface and in briefcase condition. This substantiates the conclusion that the gradient histograms are indeed a much better and more robust way to capture gradients than the pure averaging of silhouettes in GEI.

4.3. Comparison to Other Methods

We compare all our approaches to several current state-of-the-art results reported in the literature. Rank 1 and Rank 5 performance on all 12 experiments, as well as a weighed average rate are summarized in Table 1.

It can be seen that GHEI, α-GHEI and α-pb-GHEI outperform the other methods in almost all experiments.

With an average recognition rate of $83.2\,\%$ the α-pb-GHEI surpasses the standard GEI by $63\,\%$. Even when run on limiting silhouette data, silhouette-GHEI still outperforms GEI by $27\,\%$, which shows the superiority of gradient histogram representation for person identification.

5. CONCLUSION AND OUTLOOK

In this work we made several improvements to the Gradient Histogram Energy Image. We presented a part-based extension with precise detection of body parts, which is beneficial, especially for cases with degrading visual conditions such as briefcase and cloth changes. In addition, we applied GHEI directly on binary silhouettes and were able to show that the gradient histogram representation outperforms the classical Gait Energy Image, even when using the same binary silhouette data.

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	-
Variations	view, shoe			surface, view, shoe				briefcase, view, shoe			time, clothing		
Rank 1 Performance													
Baseline [3]	73	78	48	32	22	17	17	61	57	36	3	3	41.0
GEI [1]	89	87	78	36	38	20	28	62	59	59	3	6	51.0
α-GEI [19]	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
HMM [7]	89	88	68	35	28	15	21	85	80	58	17	15	53.5
DATER [24]	89	93	80	44	45	25	33	80	79	60	18	**21**	58.5
MMFA [10]	89	**94**	80	44	47	25	33	85	83	60	**27**	**21**	59.9
GTDA [9]	91	93	86	32	47	21	32	95	90	68	16	19	60.6
I-to-C [6]	93	89	81	54	52	32	34	81	78	62	12	9	61.2
GDN [8]	85	89	72	57	66	46	41	83	79	52	15	24	62.8
silhouette-GHEI	97	87	74	64	59	42	34	82	66	73	6	9	64.7
GHEI	98	91	87	70	67	30	26	96	91	88	12	3	70.2
α-GHEI	98	93	87	**94**	86	62	**50**	94	91	85	12	12	79.8
α-pb-GHEI	**99**	**94**	**91**	93	**90**	**64**	45	**99**	**98**	**96**	18	**21**	**83.2**
Rank 5 Performance													
Baseline [3]	88	93	78	66	55	42	38	85	78	62	12	15	64.5
GEI [1]	93	93	89	65	60	42	45	88	79	80	6	9	68.7
α-GEI [19]	93	94	91	55	59	39	40	92	88	81	18	15	68.7
GEI+Synth [1]	93	96	93	75	71	54	53	78	82	64	33	42	72.1
HMM [7]	-	-	-	-	-	-	-	-	-	-	-	-	-
DATER [24]	96	96	94	74	79	53	57	93	91	83	40	36	77.9
MMFA [10]	98	98	94	76	76	57	60	95	93	84	**48**	**39**	79.9
GTDA [9]	98	**99**	**95**	58	64	41	52	98	**99**	87	31	37	74.9
I-to-C [6]	97	98	93	81	74	59	55	94	95	83	30	33	79.2
GDN [8]	96	94	89	85	81	68	69	96	95	79	46	39	82.1
silhouette-GHEI	99	94	91	86	76	69	62	97	88	91	12	21	81.5
GHEI	**100**	94	91	87	74	62	60	99	98	96	15	15	82.0
α-GHEI	**100**	93	93	**97**	**93**	81	**78**	99	98	93	21	24	88.1
α-pb-GHEI	**100**	94	93	**97**	**93**	**83**	72	**100**	98	**98**	30	30	**89.4**

Table 1: Comparison of GHEI variants to other methods: Baseline [3]; Gait Energy Image (GEI) [1]; Hidden Markov Models (HMM) [7]; Alpha Gait Energy Image (α-GEI) [19]; Discriminant Analysis with Tensor Representation (DATER) [24]; Marginal Fisher Analysis (MMFA) [10]; General Tensor Discriminant Analysis (GTDA) [9]; Image-to-Class Distance (I-to-C) [6]; Gait Dynamics Normalization (GDN) [8]; original Gradient Histogram Energy Image (GHEI) [2].

It is important to note that for all our experiments we use a relatively simple feature reduction and classification scheme. An obvious extension would thus be to combine better classification methods (tensor discriminant analysis, sparse representation, ...) with GHEI. In addition, the presented part-based α-pb-GHEI can be improved. Currently parts are selected by optimal detection performance and not by optimal identification performance. Furthermore, temporal information could be used for better part alignment.

In summary, GHEI has shown high person identification performance. Also, GHEI nicely integrates to part-based person detectors. Thus, combined detection, tracking and identification systems can emerge, which will be necessary for more realistic gait recognition in actual surveillance videos.

6. 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] Martin Hofmann and Gerhard Rigoll, "Improved gait recognition using gradient histogram energy image," in *IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 1389–1392.

[3] 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.

[4] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proceedings Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2002, pp. 372–377.

[5] Chew Yean Yam, Mark S Nixon, and John N Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, 2004.

[6] 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.

[7] Amit Kale, Aravind Sundaresan, A. N. Rajagopalan, Naresh P. Cuntoor, Amit K. Roy-Chowdhury, Volker Krüger, and Rama Chellappa, "Identification of humans using gait," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163–1173, 2004.

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

[9] 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.

[10] 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.

[11] Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan, and Clinton B. Fookes, "Gait energy volumes and frontal gait recognition using depth images," in *Biometrics (IJCB), 2011 International Joint Conference on*, oct. 2011, pp. 1–6.

[12] Martin Hofmann, Sebastian Bachmann, and Gerhard Rigoll, "2.5d gait biometrics using the depth gradient histogram energy image," in *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012, pp. 399–403.

[13] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll, "The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits," *Journal of Visual Communication and Image Representation*, 2013.

[14] Xiaxi Huang and N. V. Boulgouris, "Gait recognition with shifted energy image and structural feature extraction," vol. 21, no. 4, pp. 2256–2268, 2012.

[15] Naoki Akae, Al Mansur, Yasushi Makihara, and Yasushi Yagi, "Video from nearly still: An application to low frame-rate gait recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1537–1543.

[16] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan, "Human identification using temporal information preserving gait template," vol. 34, no. 11, pp. 2164–2176, 2012.

[17] Bing Sun, Junchi Yan, and Yuncai Liu, "Human gait recognition by integrating motion feature and shape feature," in *Multimedia Technology (ICMT), 2010 International Conference on*, oct. 2010, pp. 1–4.

[18] 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.

[19] 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*, pp. 390–395, march 2012.

[20] 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.

[21] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, sept. 2010.

[22] Martin A. Fischler and Robert A. Elschlager, "The representation and matching of pictorial structures," *Computers, IEEE Transactions on*, vol. C-22, no. 1, pp. 67–92, jan. 1973.

[23] Brian DeCann and Arun Ross, "Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76670Q–76670Q.

[24] 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.