2.5D Gait Biometrics using the Depth Gradient Histogram Energy Image

Martin Hofmann, Sebastian Bachmann, Gerhard Rigoll
Institute for Human-Machine Communication
Technische Universitt Mnchen
martin.hofmann@tum.de, sebastian.bachmann@mytum.de, rigoll@tum.de

Abstract

Using gait recognition methods, people can be identified by the way they walk. The most successful and efficient of these methods are based on the Gait Energy Image (GEI). In this paper, we extend the traditional Gait Energy Image by including depth information. First, GEI is extended by calculating the required silhouettes using depth data. We then formulate a completely new feature, which we call the Depth Gradient Histogram Energy Image (DGHEI). We compare the improved depth-GEI and the new DGHEI to the traditional GEI. We do this using a new gait database which was recorded with the Kinect sensor. On this database we show significant performance gain of DGHEI.

1. Introduction

Identifying people has been studied intensively in recent years. Besides physiologic features, such as fingerprint, iris, retina, DNA and face, also behavior based features such as signature, gait and voice (which strictly speaking is also a physiological trait) have been applied. Gait recognition has interesting applications, because gait features can be obtained from people at larger distances when other features such as face are obscured. In addition, capturing gait features does not require the cooperation of the subject as it is necessary for example for fingerprint recognition. Thus, gait recognition has great potential in access control, law enforcement, video surveillance as well as tracking and monitoring.

A multitude of methods and techniques in feature extraction as well as in classification have been developed. Major approaches include model-based and model-free (appearance based) methods. In model based methods, a human pose model is extracted at each frame and the underlying kinematics are used for individual identification. While this is in some sense "true gait recognition", in practice pose estimation proves highly difficult and results show limited performance. In contrast model-free methods bypass the model fitting and extract a variety of features directly on the input data. This way, a correspondence of the person's appearance to its identity is created. In most experiments, model-free methods greatly outperform model-based approaches.

The most successful representatives of model-free methods are those that are based on the Gait Energy Image (GEI). Here, the concept of averaging features at each frame within a full gait cycle has proven to greatly reduce noise and therefore leads to a robust and efficient identity representation.

In this paper, we extend the GEI by using depth information. To this end, we recorded a new database with the Kinect sensor, which on the one hand captures standard color images and on the other hand features a depth channel which gives the physical distance of the camera to the object at each pixel. We show that simply by using depth information instead of color images for silhouette extraction, the GEI can be improved. This, however, also binarizes the depth data and most of the depth information is discarded. In order to capture this information, we present the Depth Gradient Histogram Energy Image (DGHEI). Here, depth gradients are extracted at each position and are then aggregated into direction histograms. We also compare our results to the Gait Energy Volume (GEV), which was presented in [10]. The proposed DGHEI representation greatly outperforms standard GEI, the depth-GEI, as well as the GEV.

The remainder of this paper is structured as follows. First, we present related work on depth data in Section 2. In Section 3 we present the feature extraction, by first reviewing the standard GEI and then presenting the GEI improvement, as well as the new DGHEI. Section 4 then briefly summarizes the overall gait recognition system. Experiments, including database description are given in Section 5. Finally, we conclude in the last section.

2. Related Work

As outlined in the introduction, there are model-based and model-free gait recognition methods. In model-based

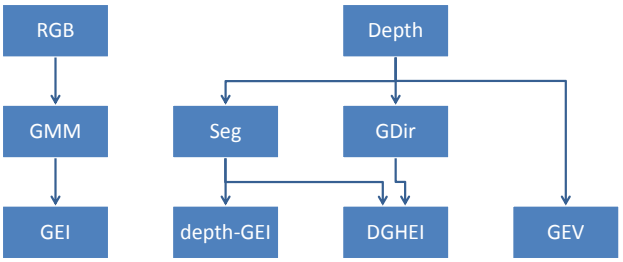


Figure 1: Feature extraction: The GEI is calculated using GMM background modeling on the RGB stream. Depth-GEI, DGHEI and GEV are extracted on the depth data.

methods [1][15], 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. While this is conceptionally solid, 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 to model-based methods, model-free methods [3][5][6][7][8][13][14] 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.

Many model-free methods build on silhouette extraction for each frame in a gait cycle. Silhouettes are either averaged (as in the prominent Gait Energy Image) [3][7], or all silhouettes are used simultaneously [6][8][12]. Different classifiers ranging from nearest neighbor [3], SVM and HMM [6][12] have been applied with similarly good results.

A majority of current model-free methods use only 2D data. Only a few works have so far addressed 3D gait recognition. For example in [9], a multi-camera system together with a structure from motion algorithm is used to build binary 3D voxel representations of the human. The voxel set is then back-projected to the side, front and top view, where 2D gait recognition methods are applied. In [10], the authors use a similar voxel reconstruction and in addition they use the Kinect sensor to obtain depth data. They define the Gait Energy Volume (GEV) as a 3D extension to the Gait Energy Image (GEI).

Our work is similar in the sense that we also do 3D gait recognition. Our work however differs in the way we make use of 3D information.

3. Feature Extraction

In the following, we first recap the traditional Gait Energy Image, which we improve in the subsequent section simply by using depth segmentation. Finally, we present the Depth Gradient Energy Image (DGHEI), which yields significant performance gain. Figure 1 shows an overview of how the features are generated from the color and depth channels.

3.1. Baseline: Gait Energy Image

The Gait Energy Image (GEI) is often used as a feature for gait recognition. The idea is simple, yet has proven highly efficient. Assuming that all gait information is captured in a full gait cycle, the information of each frame within this gait cycle is averaged. This averaging seemingly discards information, however, assuming that the noise in each frame is independent, averaging removes a substantial part of the degradation.

The simplest feature (which is assumed to capture the gait information) is the silhouette. Thus, for the Gait Energy Image, in a first step, binary silhouettes are extracted at each frame, for example using Gaussian mixture models [11]. After some possible processing with morphologic operations, the person is tracked by finding the largest blob. This found blob is extracted from the binary image and is resized, such that all blobs have the same size. In addition, the horizontal position is normalized such that the torso in each frame is roughly at the same location. Finally, the aligned silhouettes are averaged yielding the Gait Energy Image G:

$$G(x,y) = \frac{1}{T}\sum_{t=1}^{T} S_t(x,y). \quad (1)$$

3.2. Gait Energy Image on Depth Data

Extracting the Gait Energy Image as above assumes on the one hand, that the binary silhouettes actually capture the gait information, and second, it assumes that errors in silhouette extraction result from independent noise at each frame. However, in practice the silhouette extraction is performed using background modeling methods which are run on the color images. Due to difficulties in the segmentation process, silhouettes can be of quite bad quality and certainly do not reliably capture the boundary of the subject. In addition, errors often result to local similarities of the person to the background. In these regions, the error is certainly not independent at each frame.

To overcome these limitations, we use depth information instead of color information to obtain binary silhouettes $S_t(x,y)$. Depth information can be used to reliably segment the object from the background. Here, the background model is defined by the depth in the empty scene. Since the distance between object and background are relatively large, a large margin exists and simple thresholding of the depth difference results in good segmentation.

3.3. Depth Gradient Histogram Energy Images

In this section we present the Depth Gradient Histogram Energy Image (DGHEI). As mentioned above, the noise re-

ducing property by averaging feature vectors of each frame within a full gait cycle has prove highly efficient. Thus, for DGHEI, we also make use of this concept. It is interesting to note, that in the standard GEI, all information is reduced to binary silhouettes. With the newly available depth information, the edges and depth gradients within the person's silhouettes can also be used. In order to capture all gradients and edges in a robust and efficient manner, we propose the use of histogram binning. This idea is motivated by the concept of 'histograms of oriented gradients' (HOG) as they are frequently used for object detection [2].

Extraction of DGHEI therefore in a first step consists of calculating histograms of oriented gradients at each frame t:

To this end, magnitude r and orientation θ of the depth data are computed in a first step:

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

$$\theta(x,y) = \text{atan2}(u(x,y), v(x,y)) + \pi \quad (3)$$

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

$$\hat{\theta}(x,y) = \left\lfloor \frac{9 \cdot \theta(x,y)}{2\pi} \right\rfloor \quad (4)$$

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 [2]).

Next, following the averaging concept of GEI, the calculated gradient histograms are finally averaged over a full gait cycle consisting of T frames and result in the DGHEI.

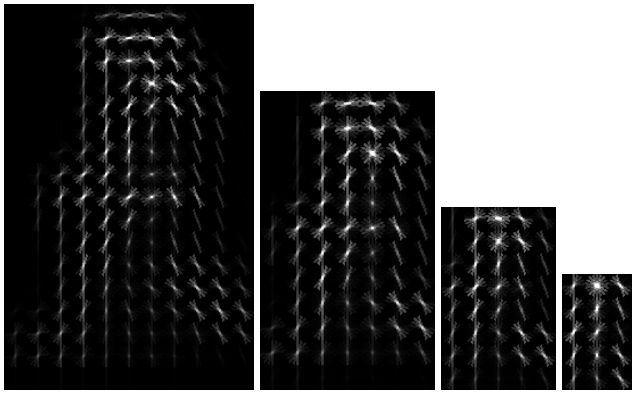
$$H(i,j,f) = \frac{1}{T} \sum_{t=1}^{T} h_t(i,j,f) \quad (5)$$

Here, i and j are pointing to the histogram cell at position (i,j) and $f = \{1 \dots 36\}$ is the index to the histogram bin. Each gait cycle is finally represented by a multidimensional feature vector $H(i,j,f)$.

Therefore, our new representation extends the GEI by using depth information. Instead of simply averaging the depth information (which has resulted in bad recognition rates), depth information is first aggregated in gradient direction histograms of non overlapping regions. The final DGHEI representation is visualized in Figure 2 using different cell quantizations.

4. Gait Recognition System

For person identification, we use a method similar to the one presented in [3]. Dimension reduction is done by



(a) 17×11 (b) 12×7 (c) 8×5 (d) 5×3

Figure 2: Visual representation of DGHEI using different quantizations

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 small amount of training data, such as it is typical for gait recognition.

4.1. Feature Space Reduction

Because of the high dimension of the DGHEI features $H(i,j,f)$, we use unsupervised Principal Component Analysis and supervised Linear Discriminant Analysis (PCA + MDA) for dimension reduction. A combination of PCA and MDA, as proposed in [4], results in the best recognition performance. While PCA seeks a projection that best represents the data, MDA seeks a projection that best separates the data.

Assume that the training set, consisting of N d-dimensional training vectors $\{g_1, g_2, \dots, g_N\}$, is given. Then the projection to the $d' < d$ dimensional PCA space is given by

$$y_k = U_{pca}(g_k - \overline{g}), \qquad k = 1, \dots, N \quad (6)$$

Here U_{pca} is the $d' \times d$ transformation matrix with the first d' orthonormal basis vectors obtained using PCA on the training set $\{g_1, g_2, \dots, g_N\}$ and $\overline{g} = \sum_{k=1}^{N} g_k$ is the mean of the training set. After PCA, MDA is performed. It is assumed that the reduced vectors $\mathcal{Y} = \{y_1, y_2, \dots, y_N\}$ belong to c classes. Thus, the set of reduced training vectors $\mathcal{Y}$ is composed of its c disjunct subsets $\mathcal{Y} = \mathcal{Y}_1 \cap \mathcal{Y}_2 \cap \dots \mathcal{Y}_c$. The MDA projection has by construction $(c-1)$ dimensions. These $(c-1)$ dimensional vectors z_k are obtained as follows

$$z_k = U_{mda} y_k, \qquad k = 1, \dots, N \quad (7)$$

where U_{mda} is the transformation matrix obtained using MDA. This matrix results from optimizing the ratio of the

between-class scatter matrix S_B and the within-class scatter matrix S_W:

$$J(U_{mda}) = \frac{|\widetilde{S}_B|}{|\widetilde{S}_W|} = \frac{|U_{mda}^T S_B U_{mda}|}{|U_{mda}^T S_W U_{mda}|}. \quad (8)$$

Here the within-class scatter matrix S_W is defined as $S_W = \sum_{i=1}^{c} S_i$, with $S_i = \sum_{y \in \mathcal{Y}_i} (y - m_i)(y - m_i)^T$ and $m_i = \frac{1}{N_i} \sum_{y \in \mathcal{Y}_i} y$. Where $N_i = |\mathcal{Y}_i|$ is the number of vectors in $\mathcal{Y}_i$. The between-class scatter S_B is defined as $S_B = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T$, with $m = \frac{1}{N} \sum_{i=1}^{c} N_i m_i$.

Finally, for each Gradient Histogram Energy Image, the corresponding gait feature vector is computed as follows

$$z_k = U_{pca} U_{mda} (g_k - \overline{g}) = T(g_k - \overline{g}), \qquad k = 1, \ldots, N \quad (9)$$

4.2. Classification

To classify samples, nearest-neighbor classification is used. Thus, the class label L_i is assigned to each test sample according to its minimal distance to the sample in the gallery set:

$$L_i = \underset{c}{\operatorname{argmin}}\, D_i(c) \quad (10)$$

For distance measure $D_i(c)$, we use Euclidean distance for all our experiments.

5. Experiments

For the evaluation of the proposed depth based methods, we needed a database with depth information. In the lack of a publicly available dataset, we recorded our own dataset, which is described below. We reimplemented the Gait Energy Image (GEI) [3] as well as the Gait Energy Volume (GEV) [10] and tuned both methods to our new dataset. Thus, we have a good relative performance evaluation of our two new approaches (depth-GEI and DGHEI) to the two established methods.

5.1. The TUM-GAID Database

For our experiments, we use a newly recorded database (the TUM Gait from Audio, Image and Depth (GAID) database). This database was recorded with the Kinect sensor and therefore features both a video stream, a depth stream as well as a four channel audio stream. Even though audio was recorded (in order to potentially allow audio based gait recognition), it is not used in this work. Both video and depth have a resolution of 640×480 at a frame rate of 30 fps. A total of 176 people were recorded, 103 male and 73 female, which makes the database roughly balanced in gender. Each person is captured in 10 sequences. Of these 10 sequences, 6 are recorded in normal walking ("normal"), 2 are carrying a backpack ("backpack") and the remaining 2 are captured with disposable coating shoes ("shoes"). The backpack is meant to degrade the visual appearance, while the coating shoes are meant to degrade the audio stream. In each sequence, people walk perpendicular to the camera at a distance of approximately 4 meters.

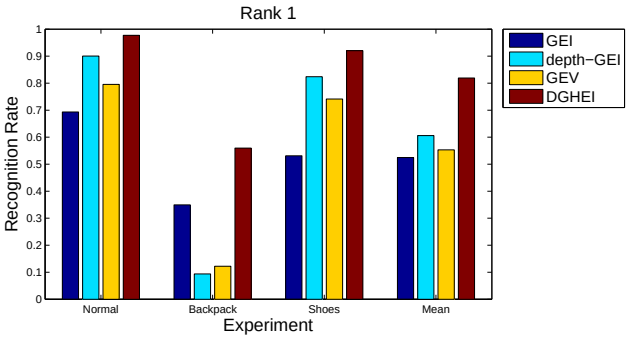


Figure 3: Results

5.2. Experimental Setup

We define three experiments for evaluation. In all three experiments, we use the first two sequences of normal walking as training data. These sequences are at the same time used as the gallery set. The test data is composed of two sequences of each "normal walking", "backpack" as well as "shoes", respectively. Furthermore, two disjoint sequences of normal walking are reserved for the development set which might be used for parameter tuning, however, in this work no parameter tuning was necessary.

5.3. Results

Results of depth-GEI and our DGHEI are shown in Table 1 and 2. In Figure 3 the results of the rank 1 performance are visualized as bar graphs. It can be seen that the depth-GEI outperforms the regular GEI, except in the scenario with the backpack. This leads to the conclusion, that the new depth-GEI representation is very well capable of precisely capturing the silhouette, however in case of large visual degradation (like the backpack), an over-precise modeling is not beneficial. The same goes for the Gait Energy Volume (GEV), which seems to work better than GEI, however fails with large silhouette degradation such as with the backpack.

Using the proposed DGHEI, these limitations can be overcome. Throughout all experiments, DGHEI surpasses the other methods by a significant margin. Using a one-tailed z-test, it can be shown, that the DGHEI significantly (on an $\alpha = 0.001$ level) outperforms all other methods in all configurations. Thus, the use of gradient orientations, as well as the used histogram binning turn out to be a robust and efficient representation, which is especially capable of handling large silhouette degradations.

	Rank 1			
	GEI	depth-GEI	GEV	DGHEI
normal	0.6932	0.9006	0.7955	**0.9773**
backpack	0.3494	0.0938	0.1222	**0.5597**
shoes	0.5313	0.8239	0.7415	**0.9205**
mean	0.5246	0.6061	0.5530	**0.8191**

Table 1: Recognition rate, rank 1

	Rank 5			
	GEI	depth-GEI	GEV	DGHEI
normal	0.8097	0.9545	0.9006	**0.9858**
backpack	0.5455	0.2841	0.3324	**0.8239**
shoes	0.6818	0.9063	0.8523	**0.9688**
mean	0.6790	0.7150	0.6951	**0.9261**

Table 2: Recognition rate, rank 5

6. Outlook and Conclusion

In this paper, we have used depth for gait recognition, which is a relatively new approach for gait recognition, since so far, most approaches focus exclusively on the visual channel. First, we did a straight forward extension of the Gait Energy Image by using depth segmentations instead of foreground/background segmentation. This worked well and showed a performance gain over the standard Gait Energy Image as long as there is no large degradation like a backpack. However, in practical applications, visual degradations such as the backpack play a crucial role. To overcome these limitations, we presented the Depth Gradient Histogram Energy Image (DGHEI), which is a very robust representation of the person's identity and shows good performance in spite of visual degradations. Thus, in a direct comparison, depth data can be beneficial for gait recognition if combined with a robust feature representation such as the proposed gradient histograms.

To continue research in depth based gait recognition, we will extend the database in order to allow for time and cloth variations, which has proven to add additional challenges [3]. Furthermore, gait recognition often (including our work) focuses on closed-set recognition, which is not very applicable in practice. A larger dataset (with more people in the test set than in the training set) will allow to model the unknown person category.

References

[1] C. BenAbdelkader, R. Cutler, and L. 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*, pages 372–377. IEEE, 2002. 2

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, june 2005. 3

[3] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 316–322, 2006. 2, 3, 4, 5

[4] P. Huang, C. Harris, and M. Nixon. Recognising humans by gait via parametric canonical space. *Journal of Artificial Intelligence in Engineering*, 13(4):359–366, November 1999. 3

[5] Y. Huang, D. Xu, and T.-J. Cham. Face and human gait recognition using image-to-class distance. *Circuits and Systems for Video Technology, IEEE Transactions on*, 20(3):431 –438, march 2010. 2

[6] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. RoyChowdhury, and V. Krueger. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004. 2

[7] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 863–876, 2006. 2

[8] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The HumanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 162–177, 2005. 2

[9] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. *Biometrics: Theory, Applications and Systems*, sept. 2008. 2

[10] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. Gait energy volumes and frontal gait recognition using depth images. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1 –6, oct. 2011. 1, 2, 4

[11] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, pages 2246–2252, 1999. 2

[12] A. Sundaresan, A. Chowdhury, and R. Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. *Proceedings IEEE International Conference on Image Processing*, 2:II – 93–6 vol.3, 2003. 2

[13] D. Tao, X. Li, X. Wu, and S. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(10):1700 –1715, oct. 2007. 2

[14] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang. Marginal fisher analysis and its variants for human gait recognition and content- based image retrieval. *Image Processing, IEEE Transactions on*, 16(11):2811 –2821, nov. 2007. 2

[15] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004. 2