Gait Recognition in the Presence of Occlusion: A New Dataset and Baseline Algorithms

Martin Hofmann[1], Shamik Sural[2], Gerhard Rigoll[1]

[1] Institute for Human-Machine Communication, Technische Universität München, Germany
{martin.hofmann,rigoll}@tum.de
[2] Indian Institue of Technology Kharagpur, India
shamik.sural@gmail.com

ABSTRACT

Human gait is an important biometric feature for identification of people. In this paper we present a new dataset for gait recognition. The presented database overcomes a crucial limitation of other state-of-the-art gait recognition databases. More specifically this database addresses the problem of dynamic and static inter object occlusion. Furthermore this dataset offers three new kinds of gait variations, which allow for challenging evaluation of recognition algorithms. In addition to presenting the database we present two baseline algorithms (Color histograms, Gait Energy Image) to perform person identification using gait. These algorithms already show promising results on the presented database.

Keywords: biometrics, gait recognition, database, occlusion, gait energy image.

1 INTRODUCTION

Person identification by biometric features is a well established research area. The main focus has so far been on physiologic features such as face, iris and fingerprint. In addition, behavior based features such as voice, signature and gait can be used for person identification. In this work we contribute to the research of person identification using gait. The main advantage of using these features over other physiologic features is the possibility to identify people from large distances and without the person's direct cooperation. For example, in low resolution images, a person's gait signature can be extracted, while the face is not even visible. Also no direct interaction with a sensing device is necessary, which allows for undisclosed identification. Thus gait recognition has great potential in video surveillance, tracking and monitoring.

Studies suggest [13] that if all gait movements are considered, gait is unique. These findings are the basis of the assumption that recognition using only gait must also be possible for a computer system. Over the last decade the field of recognizing people using gait features has received remarkable attention. A multitude of methods and techniques in feature extraction as well as in classification have been developed. Experiments are promising and encouraging.

While good datasets for training and evaluation are available (See summary in Section 3), we find that all of them ignore to address one important challenge: The challenge of occlusions. Occlusions are annoying but are unfortunately omnipresent in practice. Especially in a real word surveillance scenario, occlusions occur frequently. Typical gait recognition algorithms require a full gait cycle[1] for recognition. In the case of occlusion, however, it becomes a challenging problem to extract a full gait cycle. In heavy occlusion, parts of the gait cycle might be visible, while other parts are obscured by another person walking in front. The challenge then lies in stitching together parts of different gait cycles in order to obtain one complete gait cycle. Alternatively gait recognition algorithms could be developed for which parts of the gait cycle are sufficient. While to date, no algorithm is capable of handling partially observable gait cycles, we here present the TUM-IITKGP gait dataset, which can be used to specifically address occlusions.

To this end the presented database includes recordings with two kinds of occlusions. On the one hand dynamic occlusions by people walking in the line of sight of the camera and on the other hand static occlusions by people who are occluding the person of interest by standing in the scene. In addition to specifically addressing the occlusion challenge, the TUM-IITKGP dataset also features three new configuration variations, which allows to test algorithms for their capability of handling changes in appearance.

We present two baseline algorithms for recognition on this dataset. The first algorithm uses appearance

[1] A full gait cycle is the time interval between successive instances of initial foot-to-floor contact for the same foot

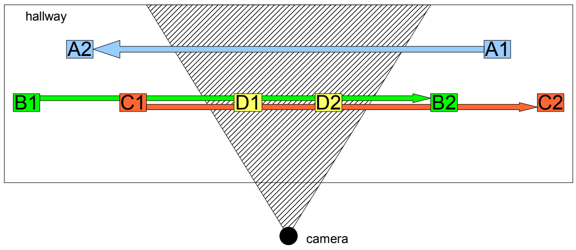


Figure 1: Physical setup of the recording

information based on color histograms. Thus this algorithm is not precisely a gait recognition algorithm, but shows promising results. The second algorithm is an actual gait recognition algorithm based on the well known Gait Energy Image (GEI) features [6][7]. It can be seen that this baseline algorithm which uses only motion information and no color information already shows excellent results.

Section 2 gives a summary of related gait recognition databases. Section 3 presents the new dataset in details. We then present in Section 4.1 a simple baseline recognition system based on color histograms, as well as an actual gait recognition baseline algorithm in Section 4.2. Results are given in Section 6 and we conclude in Section 7.

2 RELATED GAIT DATABASES

Since the field of gait recognition has been in existence for roughly a decade, the research community has long utilized publicly available databases for comparative performance evaluation.

Table 1 summarizes the most prominent gait recognition corpora. This table also shows the important features of the particular databases. The most important features of a database are the number of subjects (which should be high), as well as a good set of person variations. These variations include, but are not limited to, view angle, clothing, shoe types, surface types, carrying condition, illumination, and time.

The first available dataset was the 1998 UCSD Dataset [11], which contains merely 6 subjects. Most of the following early gait recognition databases were published in 2001 from various institutions [2][3][5][9][10][12]. Those datasets feature a medium number (about 25) of subjects. It was then found, that for meaningful evaluation, datasets should contain at least 30 subjects and possibly more.

The most comprehensive database to date, which features a large set of subjects as well as a substantial set of variations is probably the HumanID Gait Challenge [15]. Other databases such as CASIA (Dataset B) [1] also feature high numbers of subjects and a significant number of variations. CASIA additionally features an exhaustive number of views, which allows for precise 3D reconstruction.

3 THE TUM-IITKGP DATABASE

As established in the introduction, the rationale behind recording a new gait recognition dataset is to specifically address the problem of occlusions, which would frequently occur in real world applications. The TUM-IITKGP Database currently consists of 840 sequences from 35 individuals.

The physical setup can be seen in Figure 1. The camera is set up in a rather narrow hallway, reflecting a realistic setup in a potential real world surveillance application. The camera is positioned at a medium height of 1.85 meters and is oriented perpendicular to the hallway direction. Thus people are walking from right to left and from left to right in the image.

Each person is captured in six different configurations. Furthermore, each of the configurations is repeated two times in a right-to-left motion and two times in a left-to-right motion, resulting in a total of 840 sequences. Table 2 and Figure 2 show the six configurations for each person. Each person was primarily recorded in a *regular* walking configuration, followed by three degenerated configurations including hands in *pocket*, *backpack* and *gown*. These configurations can be used to evaluate recognition methods in the presence of different kinds of gait variation.

Furthermore two configurations are specifically designed to evaluate performance in the presence of occlusions. One is with two people walking past (*dynamic occlusion*). The other is with two people just standing in the line of sight (*static occlusion*).

In all of the six recordings, the person of interest (the subject) is starting to walk at point A1 and ending at point A2. In case of *dynamic occlusions* (configuration 5), the two other people are walking from B1/C1 to B2/C2, respectively. For *static occlusions* (configuration 6), the two additional people are standing at D1 and D2, respectively.

Overall, each configuration is repeated 4 times. For the second iteration the walking directions are inversed. Thus the subject is walking back from A2 to A1, and in case of occlusion configuration 5, the occluding people are also walking in the opposite direction. The third iteration is equivalent to the first and the fourth is equivalent to the second.

	Short Name	Description
Conf. 1	regular	Regular walking
Conf. 2	pocket	Walk with hands in pocket
Conf. 3	backpack	Walk with a backpack
Conf. 4	gown	Walk with gown
Conf. 5	dynamic occlusion	Occlusion by two walking people
Conf. 6	static occlusion	Occlusion by two standing people

Table 2: Walking configurations

Database, Ref.	#subjs.	#seqs.	Environment	Time	Variations
UCSD ID [11]	6	42	Outdoor, Wall background	1998	Time (minutes)
CMU Mobo [5]	25	600	Indoor, Tredmill	2001	Viewpoint, Walking speeds, Carrying conditions, Surface incline
Georgia Tech [9]	15	268	Outdoor	2001	Time(6 months), viewpoint
	18	20	Magnetic tracker	2001	Time(6 months)
HID-UMD Dataset 1 [10]	25	100	Outdoor	2001	
HID-UMD Dataset 2 [3]	55	222	Outside, Top mounted	2001	viewpoints (front, side), time
MIT, 2001 [2]	24	194	Indoor	2001	view, time (minutes)
Soton Small Database [12]	12	-	Indoor, green background	2001	carrying condition, clothing, shoe, view
Soton Large Database [12]	115	2128	Indoor, Outdoor, Tredmill	Summer 2001	view
HumanID Gait Challenge [15]	122	1870	Outdoor	May & Nov. 2002	viewpoint, surface, shoe, carrying condition, time (months)
CASIA Database A [1]	20	240	Outdoor	Dec. 2001	3 viewpoints
CASIA Database B [1]	124	13640	Indoor	Jan 2005	11 viewpoints, clothing, carrying condition
CASIA Database C [1]	153	1530	Outdoor, night, thermal camera	2005	speed, carrying condition
TUM-IITKGP	35	840	Indoor, Hallway, Occlusions	Apr. 2010	time (minutes), carrying condition, occlusions

Table 1: Comparison of other gait recognition databases


(a) regular


(b) pocket


(c) backpack


(d) gown


(e) dynamic occlusion

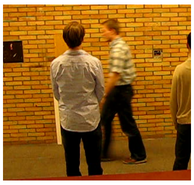

(f) static occlusion

Figure 2: Example images from all six configurations

4 BASELINE ALGORITHMS

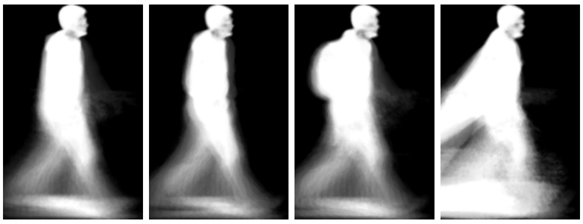
In order to show first recognition results and in order to have a means of comparing other algorithms for future performance evaluation, we applied two baseline algorithms to the database.

Both methods are non-model based. The first method uses color histograms for feature extraction, the second method uses Gait Energy Image (GEI) [7]. Obviously using only color information has a multitude of drawbacks, most importantly the fact that this kind of feature is not invariant to change of clothing.

The second method however is a true gait recognition method, because the Gait Energy Image captures temporal motion over a gait cycle and is independent from any appearance based features such as color.

4.1 Baseline Algorithm using Color Histograms

Using color histograms is a widely used technique for recognition and re-identification of people. This holds especially true for short-time recognition, where people do not change their appearance and clothing. Color histograms are extremely fast and easy to compute. Furthermore no detection of body parts is necessary, because the feature can be extracted globally from the full person. Besides the problem that color features fail in case of change in clothing, another drawback is that they are very sensitive to lighting differences especially when recognition is to be performed between differently calibrated cameras. This however can be handled using adaptive appearance transformations such as the Brightness Transfer Function [14].For this work, we use 4-by-4-by-4 3D color histograms H. Thus each person in the database is represented by a 4096 dimensional sparse feature vector. To extract this feature vector, we first use background modeling based on Gaussian Mixture Models [16] to segment foreground blobs. The color histograms are then computed over all foreground segments on the full sequence.For recognition we use nearest neighbor classification, where H_j is the



(a) regular (b) pocket (c) backpack (d) gown

Figure 3: Gait Energy Images for four configurations

j-th sample from the test set, and $\overline{H}_i$ is the mean of the samples in the i-th class from the training set.

$$L_j = \underset{i}{\operatorname{argmin}} d_X(H_j, \overline{H}_i) \tag{1}$$

Here $d_X = \{d_{euclid}, d_{corr}, d_{bhatt}, d_{chi}\}$ is the distance function of one of four different histogram comparison measures: Euclidean distance, normalized correlation, Bhattacharyya distance and Chi Squared distance with the following respective formulas:

$$d_{euclid} = \sqrt{\sum_{Bins} (H_1 - H_2)^2} \tag{2}$$

$$d_{corr} = 1 - \frac{\sum_{Bins} (H_1 - \bar{H}_1)(H_2 - \bar{H}_2)}{\sum_{Bins} \sqrt{(H_1 - \bar{H}_1)^2 (H_2 - \bar{H}_2)^2}} \tag{3}$$

$$d_{bhatt} = 1 - \sum_{Bins} \sqrt{\frac{H_1}{|H_1|} \frac{H_2}{|H_2|}} \tag{4}$$

$$d_{chi} = \sum_{Bins, H_1 + H_2 \neq 0} \frac{(H_1 - H_2)^2}{H_1 + H_2} \tag{5}$$

In the experiments it turned out that all four of these distance measures performed similarly well with a slight tendency of the Chi Squared distance being the best. See results in Section 6.

4.2 Baseline Algorithm based on Gait Energy Image

In contrast to the color histogram method presented in the previous section, GEI [6] is considered a true gait recognition method, because the used features only make use of silhouette and motion information. Appearance and color information is discarded.

4.3 Feature Extraction using GEI

In essence, the Gait Energy Image is an arithmetic mean of the binarized foreground blobs. Denote B_t the foreground silhouette in frame t. Then, the Gait Energy Image g is formally defined as the silhouette average over one full gait cycle:

$$g(x,y) = \frac{1}{T} \sum_{t=1}^{T} B_t(x,y) \tag{6}$$

Here, T is the number of frames in a full gait cycle. Using this kind of feature greatly reduces the available data, since all the gait information is compressed to only one gray level image. Figure 3 shows Gait Energy Images for the first four configurations. It has been shown that this representation suffices for person identification [7].

4.4 Feature Space Reduction

The gait energy images $g(x,y)$ have a resolution of 130×200 pixels. Thus the feature vector is still large with 26000 dimensions. We apply principal component analysis (PCA) followed by multiple discriminant analysis (MDA) to reduce the size of the feature vector. A combination of PCA and MDA, as proposed in [8], results in the best recognition performance. While PCA seeks a projection that best represents the data [4], MDA seeks a projection that best separates the data [8].

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

$$y_k = U_{pca}(g_k - \overline{g}), \qquad k = 1, \ldots, N \tag{7}$$

Here U_{pca} is the $d' \times d$ transformation matrix with the first d' orthonormal basis vectors obtained using PCA on the training set $\{g_1, g_2, \ldots, g_N\}$ 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 $\mathscr{Y} = \{y_1, y_2, \ldots, y_N\}$ belong to c classes. Thus the set of reduced training vectors $\mathscr{Y}$ is composed of its c disjunct subsets $\mathscr{Y} = \mathscr{Y}_1 \cap \mathscr{Y}_2 \cap \ldots \mathscr{Y}_c$. The MDA projection has by construction $(c-1)$ dimensions. These $(c-1)$ dimensional vectors z_k are obtained as follows

$$z_k = V_{mda} y_k, \qquad k = 1, \ldots, N \tag{8}$$

where V_{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:

$$J(V) = \frac{|\widetilde{S}_B|}{|\widetilde{S}_W|} = \frac{|V^T S_B V|}{|V^T S_W V|}. \tag{9}$$

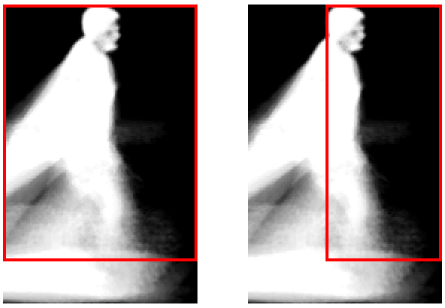
Here the within-class scatter matrix S_W is defined as $S_W = \sum_{i=1}^{c} S_i$, with $S_i = \sum_{y \in \mathscr{Y}_i} (y - m_i)(y - m_i)^T$ and $m_i = \frac{1}{N_i} \sum_{y \in \mathscr{Y}_i} y$. Where $N_i = |\mathscr{Y}_i|$ is the number of vectors in $\mathscr{Y}_i$. The between-class scatter 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 Gait Energy Image, the corresponding gait feature vector is computed as follows

$$z_k = U_{pca} V_{mda}(g_k - \overline{g}) = T(g_k - \overline{g}), \qquad k = 1, \ldots, N \tag{10}$$

4.5 Classification

For further classification, we use nearest neighbor classification on this reduced set of feature vectors. To this



(a) Baseline 2a　　(b) Baseline 2b

Figure 4: Cropped GEI regions used for recognition

end, first the mean feature vector $\bar{z}_i$ is calculated for each class.

$$\bar{z}_i = \frac{1}{|\mathscr{Z}_i|} \sum_{z \in \mathscr{Z}_i} z. \tag{11}$$

For each Gait Energy Image from the test set $\hat{g}_j$, we perform the identical transformation to get the reduced feature vector

$$\hat{z}_j = T(\hat{g}_j - \bar{g}) \tag{12}$$

Person identification then becomes a nearest-neighbor classification. We assign a class label L_j to each test gait image according to

$$L_j = \underset{i}{\operatorname{argmin}} ||\hat{z}_j - \bar{z}_i|| \tag{13}$$

4.6 Implementation details

Besides the principle approach as it was described above, there are several technical details that had to be considered. First, for our experiments, we align the foreground blobs B_t before calculating the GEI. This is done by centering each blob B_t based on the centroid of the top 10% of each blob. This way it is guaranteed that the heads, which are most stable in recognition, are all aligned at the same position.

Second, we found (just like others have [7] [15]), that using the full Gait Energy Images for recognition does not result in the best performance. Especially the lower region of the image is quite troublesome, because of shadows and reflections on the ground, as well as different floor types (as in [15]). Therefore we decided to use only the top 80% of the GEIs. Figure 4 depicts the cropping regions.

In addition we experimented with a second cropped variation of the GEIs. Here we use the top 80% of the image, and only the rightmost 60% of the image. This way, only the frontal part of the persons are included. This is beneficial, because this way the gown and the backpack have a much smaller impact on the Gait Energy Images. In Section 6 we show that this cropping indeed leads to improved recognition rates.

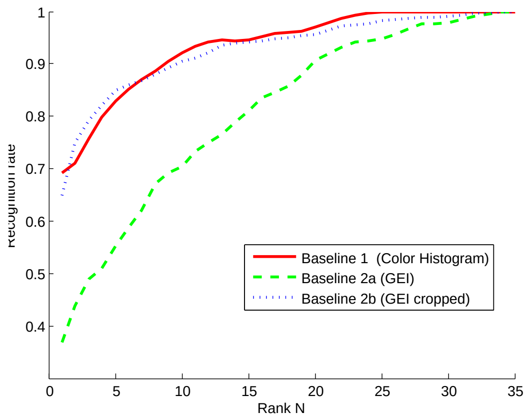


Figure 5: Rank N recognition rates for the three Baseline variants

5 EVALUATION METHOD

The presented gait recognition database is meant to be a basis for performance evaluation of various present and future gait recognition algorithms. Therefore, evaluation should be carried out the same way for all algorithms. We propose the following procedure:

The goal is to recognize a person, which has only been seen once before. Thus the training set consists of merely one single sequence of each person. We define that this sequence is one from the first configuration (regular walking). Consequently the test set consists of 23 sequences for each person (three for *regular* and four each for the other five configurations). Because the database consists of 35 people, the overall test set consists in total of 805 sequences.

Since there are four sequences for the first configuration, a 4-fold cross validation is performed. This means that there are 4 rounds of evaluation, each time with one of the four sequence in regular configuration as the sole training sample, and all the rest as the test. The result is then averaged over the 4 rounds of evaluation.

6 RESULTS

We evaluate on the one hand the color histogram based recognition method (Baseline 1), described in Section 4.1 and on the other hand we evaluate GEI approach (Baseline 2) described in Section 4.2. In the second case we evaluate the recognition rate for the two differently cropped Gait Energy Images. However, at this point, the Gait Energy Image could not be calculated for configuration 5 and 6, thus there are no results in those cases.

Results are shown in Table 3. We evaluate the recognition rate for each of the configurations separately and we report overall recognition rates. For all three variants, we report rank 1, rank 5 and rank 10 recognition rates. Figure 5 additionally shows the cumulative matching characteristic (CMC) for the rank n recognition rates for all three baseline variant.

	Top 1			Top 5			Top 10		
	BL 1	BL 2a	BL 2b	BL 1	BL 2a	BL 2b	BL 1	BL 2a	BL 2b
Conf. 1	97.9%	68.6%	77.1%	100%	76.2%	94.3%	100%	85.7%	97.1%
Conf. 2	93.3%	67.1%	75.7%	93.3%	80.7%	94.3%	100%	90.0%	97.8%
Conf. 3	75.0%	11.4%	77.1%	91.7%	45.7%	90.0%	100%	66.4%	94.3%
Conf. 4	20.0%	8.6%	32.9%	60.0%	23.6%	63.6%	73.3%	43.5%	74.3%
All(1-4)	69.9%	36.9%	64.9%	85.2%	55.2%	84.9%	92.6%	70.5%	90.5%
Conf. 5	43.7%	-	-	60.4%	-	-	77.1%	-	-
Conf. 6	70.0%	-	-	90%	-	-	100%	-	-
All(1-6)	65.8%	-	-	81.9%	-	-	91.1%	-	-

Table 3: Results for Baseline 1 (Color Histogram), Baseline 2a (Gait Energy Image) and Baseline 2b (Cropped Gait Energy Image)

It can be seen that the simple color based recognition method outperforms the GEI approach. However, the GEI approach also shows excellent results, and in case of the cropped GEI, the performance of GEI surpassed the performance of the color histogram method.

7 CONCLUSIONS

In this paper we have presented a new gait recognition database, which is focused on the problem of occlusions. Besides addressing the occlusion problem, the database also addresses three new kinds of variations which have not yet been addressed by other datasets. More specifically these variations include hands in *pocket*, wearing *backpack* and waring a *gown*.

We have presented two baseline algorithms which perform excellent on the given dataset for the case of no occlusion. However, so far neither of these two algorithms specifically addresses the occlusion problem, resulting in low performance for those cases. Thus it remains future work of actually utilizing the databases capabilities to show good performance in spite of occlusions.

8 ACKNOWLEDGMENTS

This work has been partially funded by the European Projects FP-214901 (PROMETHEUS) as well as by Alexander von Humboldt Fellowship for experienced researchers.

REFERENCES

[1] Center for biometrics and security research, CASIA. http://www.cbsr.ia.ac.cn.

[2] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *Proceedings of IEEE Conference on Face and Gesture Recognition*, pages 351–356, 2002.

[3] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *Proc. ICASSP*, pages 6–10, 2003.

[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, New York, 2. edition, 2001.

[5] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical report, 2001.

[6] J. Han and B. Bhanu. Statistical feature fusion for gait-based human recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:842–847, 2004.

[7] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(2):316–322, 2006.

[8] 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.

[9] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 301–311, 2001.

[10] A. Kale, N. Cuntoor, and R. Chellappa. A framework for activity-specific human identification. In *International Conference on Acoustics, Speech and Signal Processing*, 2002.

[11] J. Little and J. E. Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1:1–32, 1996.

[12] J. S. M. Nixon, J. Carter and M. Grant. Experimental plan for automatic gait recognition. Technical Report 2001.

[13] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46-A(2):335–360, 1964.

[14] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *British Machine Vision Conference*, 2008.

[15] 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*, 27:162–177, 2005.

[16] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:2246, 1999.