Gait Recognition Using Motion Trajectory Analysis

Chapter · May 2017

DOI: 10.1007/978-3-319-59162-9_8

4 authors:

Muhammad Hassan Khan
Universität Siegen

Frédéric Li
Universität Siegen

Muhammad Shahid Farid
University of the Punjab

Marcin Grzegorzek
Universität Siegen

Gait Recognition using Motion Trajectory Analysis

Muhammad Hassan Khan[1], Frederic Li[1], Muhammad Shahid Farid[2], and Marcin Grzegorzek[3]

[1] Research Group of Pattern Recognition, University of Siegen, Siegen, Germany
[2] College of Information Technology, University of the Punjab, Lahore, Pakistan
[3] Faculty of Informatics and Communication, University of Economics in Katowice, Katowice, Poland

Abstract. Gait recognition has received significant attention in the recent years due to its applications in numerous fields of computer vision, particularly in automated person identification in visual surveillance and monitoring systems. In this paper, we propose a novel algorithm for gait recognition using spatio-temporal motion characteristics of a person. The proposed algorithm consists of four steps. First, motion features are extracted from video sequence which are used to generate a codebook in the second step. In a third step, the local descriptors are encoded using Fisher vector encoding. Finally, the encoded features are classified using linear Support Vector Machine (SVM). The performance of the proposed algorithm is evaluated and compared with state-of-the-art on two widely used gait databases TUM GAID and CASIA-A. The recognition results demonstrate the effectiveness of the proposed algorithm.

Keywords: Gait recognition, Spatiotemporal model, Fisher vector encoding, Visual surveillance

1 Introduction

Gait is a walking style of a person and can be used to recognize the individuals. Unlike other identification modalities such as fingerprints, faces or iris biometric which require a cooperation and physical contact of human, gait collection do not require any interaction of human with the system and can be performed at distance or at low resolution in a noninvasive and hidden manners, which is unobtrusive. They are extremely useful in many applications such as surveillance system and service robots interacting with human in daily life. Gait recognition is a challenging task, considering it is affected by various factors, such as the type of clothing or shoes, the walking pace, the nature of the floor, injuries or other similar reasons. Although, gait may not be as powerful as fingerprints or other modalities to identify the person but the characteristic to recognize a person from distance makes it irreplaceable in many cases such as visual surveillance [1].

In this paper, we introduce a novel spatio-temporal gait representation to characterize the distinctive motion information of an individual's gait using dense trajectories. It is important to mention that the proposed algorithm does not require silhouette extraction or information related to it like contour and skeleton of a human-body. A sample of dense points is selected in each frame and being tracked in successive frames based on displacement information from a dense optical flow field. Five different local descriptors are computed along dense trajectories. In particular, various combinations of local descriptors are evaluated for gait recognition and the best combination is found. We used Principal Component Analysis (PCA) to project our local descriptors into a low-dimensional subspace to alleviate the curse of dimensionality problem. A codebook approach based on Gaussian Mixture Model (GMM) is exploited to encode the features. We encode our local descriptors using Fisher vectors (FV) and fuse them using representation level fusion [2]. Finally, linear Support Vector Machine (SVM) is employed to classify the encoded features.

2 Related Work

Numerous gait recognition methods have been proposed in literature and they can be categorized into two classes: (1) model-based approaches and (2) model-free approaches. In model-based approaches, the structure of human body and its motion are used as a model by tracking the different body parts and joint position over time [3]. Such models which may include stick figure, interlinked pendulum and ellipse are generally constructed based on the prior knowledge of human body shape. Bouchrika et al. [4] proposed a motion-based model using the elliptic Fourier descriptors to extract features from human joints and incorporated them to establish a model for person identification. In [5], the author split the human's body region into three parts and the variance of these parts are combined and used as gait features. Wang et al. [6] constructed structure-based and motion-based models using a condensation framework to refine the feature extraction for gait recognition. Recent studies [3, 4, 7] have shown that such models can deal, to some extent, with the occlusion and rotation problems, however, the performance of the model-based approaches is highly dependent on the localization of torso, which is not easy to extract. Model-based techniques are computationally expensive and also sensitive to the quality of the video, therefore they are not considered suitable for real-world applications [7].

The model-free approaches operate directly on the sequence of extracted human silhouettes. These techniques either exploit the temporal information of human motion in the recognition process [1, 8–10] or convert the images of a complete gait sequence into a single template [11–13], and used them to recognize the individual's gait. In [1], the human's body silhouettes are extracted using background modeling and averaged over the time in a gait-cycle. This new representation is known as gait energy image (GEI) and then classified using Bayesian classifier. Several improvement in GEI, such as frame difference energy image (FDEI) [9], Gait Entropy Image (GEnI) [8], Chrono-Gait Image (CGI) [10] and Gait Energy Volume (GEV) [14], Depth Gradient Histogram Energy Image (DGHEI) [15] have also been proposed.

In [16], various parameters of a person's silhouette and contour (e.g., height and width ratio, silhouette area, width and centroid of the contour) are extracted and gait is approximated by radial basis function (RBF). Microsoft Kinect is also used for gait recognition due to its built-in features of depth data and skeleton information [14,15,17]for human-body segmentation. However, its biggest restriction is the field-of-view, which is very limited ($1-4$ meters) [18]. In comparison with model-based approaches, the model free techniques have shown more convincing recognition results on various gait datasets and they are computationally efficient as well.

Several algorithms have used the motion information for gait recognition. Little et al. [13] developed the shape of motion using optical flow field and used it to generate a feature vector for gait recognition. The method in [11] captured the motion intensity and its direction information using optical flow field, and formulated in a histogram based gait representation. In [19], authors calculated the optical flow fields on extracted silhouettes and formulated a new representation for gait recognition, known as Gait Flow Image (GFI). Sivapalan et al. [12] proposed the fusion of HOG and Local Directional Patterns (LDP) to identify the gait and claims significant improvement in recognition. The authors in [7] compute the HOF using the silhouette images of a gait sequence and average them on a full gait cycle. This new gait representation is known as Flow Histogram Energy Image (FHEI). The major disadvantage with silhouette based gait recognition techniques is that they require the precise segmentation of silhouette from the background images which could not always be possible and still a challenging problem in the literature [16].

The rest of the paper is organized as follows: Sect. 3 describes the proposed gait recognition algorithm. Experiments and results are reported in Sect. 4 and the conclusions are drawn in Sect. 5.

3 Proposed Method

The proposed gait recognition algorithm works in four steps. In the first step, dense trajectories are generated based on optical flow field and their motion information is encoded using local descriptors. In second step, a codebook is built based on GMM from one million randomly selected motion descriptors. In third step, we encode the local descriptors using Fisher vector encoding and fuse them in a representation level fusion. In the final step, the computed features are classified using linear Support Vector Machine (SVM). Each step of the algorithm is explained in the following sections.

3.1 Motion Descriptor

Numerous features have been proposed in recent years and have been successfully exploited in various computer vision problems; SIFT (Scale Invariant Feature Transform) [20], SURF (Speeded-Up Robust Feature) [21], HOG (Histogram of Oriented Gradient) [22], HOF (Histogram of Optical Flow) [23], MBH (Motion

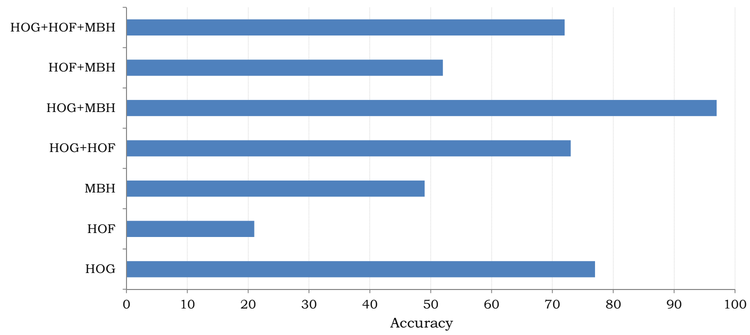

Fig. 1: Performance of various motion descriptors for gait recognition on TUM GAID gait database.

Boundary Histogram) [24], and trajectory [25] are a few to mention. Recently, dense trajectories have shown excellent results in action recognition [2, 25]. Our motivation to use dense trajectories for gait recognition is because they encode the local motion patterns and they can be easily extracted from video sequences. A sample of dense points is selected from each frame and being tracked in successive frames based on displacement information obtained from a dense optical flow field. Specifically, each point $P_t = (x_t, y_t)$ at frame t is tracked in frame $t+1$ by median filtering in a dense optical flow field. Given a trajectory of length L, a sequence of displacement vector S is formed as follow:

$$S = (\Delta P_t, ..., \Delta P_{t+L-1}), \tag{1}$$

where $\Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$. The resulting vector S is then normalized by the sum of the magnitudes of the displacement vector. That is,

$$S' = \frac{(\Delta P_t, ..., \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|} \tag{2}$$

The descriptor S' encodes the shape of trajectory. Wang et al. [25] proposed the HOG and HOF features along the dense trajectories. Moreover, two derivatives along the horizontal and the vertical components of the optical flow are computed, known as MBH_x and MBH_y respectively, to encode the relative motion information between pixels along the respective axis. The orientation information of each local descriptors are quantized into 8-bin histograms, with an additional zero bin for HOF (i.e., in total 9 bins) and normalized with L_2-norm separately. We tested various combinations of all local descriptors on TUM GAID dataset [15] to evaluate their performance for gait recognition. Fig. 1 shows the accuracy achieved by various combinations of these features. From the experimental results, we concluded that HOG in combination with MBH outperforms the others. HOG contains the information of static appearance whereas the MBH highlights the information about the changes in optical flow field (i.e., motion boundaries). Combining information about person appearance and local motion characteristics therefore greatly improves the result of gait recognition.

3.2 Codebook Generation

To build a codebook, we randomly select one million features from each descriptors. Gaussian mixture model (GMM) is a generative model to describe the distribution over feature space [2]:

$$p(X \mid \theta) = \sum_{i=1}^{K} w_i \mathcal{N}(x \mid \mu_i, \textstyle\sum_i) \tag{3}$$

where i is the mixture number (i.e., cluster number), K is total number of clusters, w_i is the weight of ith cluster and μ_i, $\sum_i$ are the mean and covariance matrix of the ith cluster, respectively. Moreover, $\theta = \{w_i, \mu_i, \sum_i, i = 1, 2, ..., K\}$ is the set of model parameters. $\mathcal{N}(x \mid \mu_i, \sum_i)$ is the D-dimensional Gaussian distribution and can be expressed as,

$$\mathcal{N}(x \mid \mu_i, \textstyle\sum_i) = \frac{1}{\sqrt{(2\pi)^D \mid \sum_i \mid}} e^{\frac{-1}{2}(x-\mu_i)^\top \sum_i^{-1}(x-\mu_i)} \tag{4}$$

In a given feature set $X = \{x_1, ..., x_t\}$, the optimal parameters of GMM are learned through maximum likelihood estimation [26]. Then the soft assignment of data x_t to cluster i (also known as posterior probability) is defined as,

$$q_t(i) = \frac{w_i \mathcal{N}(x_t \mid \mu_i, \sum_i)}{\sum_{j=1}^{K} w_j \mathcal{N}(x_t \mid \mu_j, \sum_j)} \tag{5}$$

We consider that each model represents a specific motion pattern shared by the descriptors in the codebook. Unlike the k-means clustering, which performs hard assignment of feature descriptor to codeword, Expectation maximization (EM) algorithm of GMM performs soft assignments of feature descriptor to each mixture component. In this way, the local descriptors will be assigned to multiple clusters in a weighted manner using the posterior component probability given by the descriptor.

3.3 Feature Encoding

Once the local descriptors are extracted, they are used to construct a signature to describe the video. Feature encoding converts the local descriptors into a fixed length vector and this task is usually accomplished by vector quantization of feature vectors and building a histogram of visual words [27]. Inspired by the recent success of Fisher vector (FV) encoding in image classification, object detection and action recognition [28, 29], we choose it to encode our descriptors under GMM. FV is an extension of bag-of-visual words (BoV) and derived from Fisher kernel [28] which combines the characteristics of both discriminative and generic approaches.

A given feature set $X = \{x_t, t = 1, ..., T\}$ of local descriptors can be modeled into a vector by using the probability density function $p(X|\theta)$ (See Eq. 3). The X can be mapped into a vector by computing the gradient vector of its log-likelihood function at the current model parameters θ [27],

$$F_X = \frac{1}{T} \nabla_\theta logp(X|\theta), \tag{6}$$

where F_X representing the FV and its dimensionality depends on the number of parameters in θ. The gradient of the log-likelihood function ∇_θ describes the contribution of parameters in the generation process. Assuming that x_t is D-dimensional local descriptor, its gradient vector with respect to mean μ_i and covariance $\sum_i$ is defined as [30]:

$$u_i = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^{T} q_t(i) \frac{x_t - \mu_i}{\sum_i}, \tag{7}$$

$$v_i = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^{T} q_t(i) \left[\frac{(x_t - \mu_i)^2}{\sum_i^2} - 1 \right], \tag{8}$$

where $q_t(i)$ is the soft assignments of t descriptor to ith Gaussian component (5) and u, v are D-dimensional vectors. Eq. 7 and Eq. 8 are known as the first and second order differences of descriptor points to cluster centers, respectively. The Fisher encoding for the set of local descriptors X is computed by concatenating the all u and v for all K components. That is,

$$f = [u_1^\top, v_1^\top, u_2^\top, v_2^\top, ..., u_K^\top, v_K^\top]^\top \tag{9}$$

The total size of encoded vector is $2KD$ where K is total number of clusters and D is the dimension of descriptor. We encode the HOG, MBH_x and MBH_y descriptors using the above description and fuse them using representation level fusion [2].

3.4 Classification

The encoded vectors are classified using Linear Support Vector Machine (SVM). SVM is considered a powerful tool for solving classification problems in many applications. Due to the high dimensionality of our features, we decided to use SVM as a classifier. In the comparison of SVM, the other similarity based classifiers like K-Nearest Neighbor and probability based classifiers such as Naive Bayes do not perform well on high dimensional features [31,32]. SVM first maps the training samples in high dimensional space and then extracts a hyper-plane between the different classes of objects using the principle of maximizing the margin. Because of this principle, the generalization error of the SVM is theoretically independent from the number of feature dimensions. In this paper, we used LIBLINEAR SVM library [33] for classification.

Table 1: Comparison of recognition performance on TUM GAID gait database. GS is Gallery Set and PS is Probe Set. The best results in each experiment are indicated in bold font.

Experiment	GS	PS	DGHEI [15]	SVIM [35]	GEI [15]	GEV [15]	Proposed
A	N	N	**99.4**	99.0	96.8	94.2	97.74
B	N	B	27.1	18.4	3.9	13.9	**39.35**
C	N	S	56.2	**96.1**	88.7	87.7	85.16
D	N	TN	44	15.6	28	28	**46.88**
E	N	TB	6	3.1	0	0	**18.75**
F	N	TS	9	28.1	22	22	**34.38**

4 Experiments and Results

The performance of the proposed method is evaluated on two widely used gait recognition databases: TUM GAID database [15] and CASIA dataset A [34]. In all experiments, we compute the local descriptors on each video sequence using dense trajectories as described in the previous section. Each of our local descriptor is 96 dimensional long. We used PCA to project them into lower dimension space to alleviate the curse of dimensionality problem, prior to compute codebook and FV. We observed from experiments that the reduced dimension of 48 produces the best results. The number of clusters in GMM modeling is fixed to 256 in all experiments.

4.1 Performance Evaluation on TUM GAID Database

The TUM GAID [15] is one of the biggest gait databases. It consists of 3370 video sequences of 305 subjects captured in outdoor environment. The database was recorded at 30 frames/second at Technical University of Munich, Germany in two different sessions using Microsoft Kinect camera. The first session was recorded in January 2012, which is the winter season in the region and temperature is around $-15°$C. Thus, the subjects were wearing heavy jackets and mostly winter boots. A total of 176 subjects were recorded in this session. The second session was recorded in April 2012 when temperature is around $+15°$C, thus, subjects were wearing significantly different clothes. Therefore, there was a substantial variation in the clothing of participants. A total of 161 subjects were participated in the second session. There was a subset of 32 subjects who participated in both sessions.

Ten walk sequences were captured for each subject namely: normal walk (N), walk with bag-pack (B) and walk with coating shoes (S). Each subject has 6 sequences of N, two sequences of B and two sequences of S, recorded while the subject is moving from left-to-right and right-to-left in the lateral view. Each subject in the subset of 32 peoples (who participated in both sessions) has 10 more walk sequences which are represented as: normal walk after time (TN),

Table 2: Details of experiment on CASIA-A gait database.

Experiment	A
Gallery Set	First 3 sequences of each subject
Probe Set	Forth sequence of each subject
Gallery Size	60
Probe Size	20

Table 3: Comparison of recognition performance on CASIA-A gait database. The best results in each experiment are indicated in bold font.

Experiment	Wang [34]	Ning [36]	Bashir [11]	Kusakunniran [37]	Proposed
A	82.5	88.75	97.5	**100**	**100**

walk with bag-pack after time (TB), and walk with coating shoes after time (TS). This subset allows the research community to analyze the effectiveness of their methods under time and clothing variations. The author in [15] has divided the dataset of 305 subjects in two sets: (1) development set consist of 150 subjects; (2) test set contains the recording of 155 subjects. We used the same division to evaluate the performance of our algorithm.

We compare the performance of the proposed gait recognition algorithm with several competing approaches, including DGHEI [15], SVIM [35], GEI [15], and GEV [15]. The results are reported in Tab. 1, the best results in each experiments are indicated in bold font. The results show that in most experiments the proposed algorithm outperforms the other approaches by significant margins. In experiment C, SVIM [35] shows better performance and in experiment A, DGHEI [15] performs marginally better then ours.

4.2 Performance Evaluation on CASIA-A Database

The CASIA-A dataset contains walking sequences of 20 subjects. Each subject in the dataset has four sequences in three different directions; parallel to the image plane (i.e., lateral view), 45° (i.e., oblique view) and 90° (i.e., frontal view). Each subject walks twice, from left to right and right to left. The proposed method is evaluated on the sequences recorded in lateral view. We used the first 3 out of 4 sequences in gallery set and the fourth one is used in probe set. Further details of the experiment are described in Tab. 2.

We also compared the performance of the proposed algorithm with the state-of-the-art methods: Wang [34], Ning [36], Bashir [11], and Kusakunniran [37]. The results are outlined in Tab. 3, which shows that both proposed algorithm and Kusakunniran [37] approach achieve 100% recognition.

The results of the experiments obtained on these two large gait databases show the effectiveness of the proposed algorithm. On TUM GAID database, the difference of recognition rate of the proposed algorithm and the state-of-the-art is up to 35%. On CASIA-A database, the proposed algorithm achieves 100% recognition.

5 Conclusion

A new gait recognition algorithm based on spatio-temporal characteristics of a human motion is presented in this paper. The proposed method constructs a set of dense trajectories by tracking a set of points in successive frames. Five different local descriptors are computed along the dense trajectories and various combinations of all local descriptors are evaluated to test their performance for gait recognition. The descriptors are encoded using Fisher vector encoding and classified using Linear SVM. The experimental results on two large public datasets reveal that proposed algorithm is highly accurate.

References

1. Man, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2) (2006) 316–322
2. Peng, X., Wang, L., Wang, X., Qiao, Y.: Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. Comput. Vis. Image Underst. **150** (2016) 109 – 125
3. Nixon, M.S., Tan, T., Chellappa, R.: Human identification based on gait. Volume 4. Springer Science & Business Media (2010)
4. Bouchrika, I., Nixon, M.S.: Model-based feature extraction for gait analysis and recognition. In: IEEE ICCV, Springer (2007) 150–160
5. Chai, Y., Wang, Q., Jia, J., Zhao, R.: A novel human gait recognition method by segmenting and extracting the region variance feature. In: IEEE ICPR. Volume 4. (2006) 425–428
6. Wang, L., Ning, H., Tan, T., Hu, W.: Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(2) (2004) 149–158
7. Yang, Y., Tu, D., Li, G.: Gait recognition using flow histogram energy image. In: IEEE ICPR, IEEE (2014) 444–449
8. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. Lett. **31**(13) (2010) 2052–2060
9. Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J.: Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognit. Lett. **30**(11) (2009) 977–984
10. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11) (2012) 2164–2176
11. Bashir, K., Xiang, T., Gong, S., Mary, Q.: Gait representation using flow fields. In: BMVC. (2009) 1–11
12. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Histogram of weighted local directions for gait recognition. In: IEEE CVPR Workshop. (2013) 125–130
13. Little, J., Boyd, J.: Recognizing people by their gait: the shape of motion. Videre: Journal of Computer Vision Research **1**(2) (1998) 1–32

14. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Gait energy volumes and frontal gait recognition using depth images. In: IEEE IJCB. (2011) 1–6
15. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5D gait biometrics using the depth gradient histogram energy image. In: IEEE BTAS. (2012) 399–403
16. Zeng, W., Wang, C., Yang, F.: Silhouette-based gait recognition via deterministic learning. Pattern Recognit. **47**(11) (2014) 3568–3584
17. Gianaria, E., Balossino, N., Grangetto, M., Lucenteforte, M.: Gait characterization using dynamic skeleton acquisition. In: Proc. Int. Workshop Multimed. Signal Process. (MMSP). (Sept 2013) 440–445
18. Khan, M.H., Shirahama, K., Farid, M.S., Grzegorzek, M.: Multiple human detection in depth images. In: IEEE Int. Workshop on MMSP. (2016) 1–6
19. Lam, T.H., Cheung, K.H., Liu, J.N.: Gait flow image: A silhouette-based gait representation for human identification. Pattern Recognit. **44**(4) (2011) 973–987
20. Lowe, D.G.: Object recognition from local scale-invariant features. In: EEE ICCV. Volume 2. (1999) 1150–1157 vol.2
21. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: ECCV 2006. (2006) 404–417
22. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE CVPR. Volume 1. (June 2005) 886–893 vol. 1
23. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: IEEE CVPR. (June 2008) 1–8
24. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: ECCV. (2006) 428–441
25. Wang, H., Schmid, C.: Action recognition with improved trajectories. In: IEEE ICCV. (2013) 3551–3558
26. McLachlan, G., Peel, D.: Finite mixture models. John Wiley & Sons (2004)
27. Sun, C., Nevatia, R.: Large-scale web video event classification by use of fisher vectors. In: IEEE WACV. (2013) 15–22
28. Sanchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the fisher vector: Theory and practice. Int. J. Comput. Vis. **105**(3) (2013) 222–245
29. Peng, X., Zou, C., Qiao, Y., Peng, Q.: Action recognition with stacked fisher vectors. In: ECCV, Springer (2014) 581–595
30. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: ECCV, Springer (2010) 143–156
31. Khan, M.H., Helsper, J., Boukhers, Z., Grzegorzek, M.: Automatic recognition of movement patterns in the vojta-therapy using rgb-d data. In: IEEE ICIP. (2016) 1235–1239
32. Khan, M.H., Helsper, J., Yang, C., Grzegorzek, M.: An automatic vision-based monitoring system for accurate vojta-therapy. In: IEEE/ACIS ICIS. (2016) 1–6
33. Fan, R.E., et al.: Liblinear: A library for large linear classification. J. of machine learning research **9**(Aug) (2008) 1871–1874
34. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12) (2003) 1505–1518
35. Whytock, T., Belyaev, A., Robertson, N.M.: Dynamic distance-based shape features for gait recognition. J. Math. Imaging Vis. **50**(3) (2014) 314–326
36. Wang, L., Tan, T., Hu, W., Ning, H.: Automatic gait recognition based on statistical shape analysis. IEEE Trans. Image Process. **12**(9) (2003) 1120–1131
37. Kusakunniran, W.: Attribute-based learning for gait recognition using spatio-temporal interest points. Image Vis. Comput. **32**(12) (2014) 1117–1126