ISSN 0361-7688, Programming and Computer Software, 2019, Vol. 45, No. 4, pp. 213–220. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Programmirovanie, 2019, Vol. 45, No. 4.

Methods of Gait Recognition in Video

A. Sokolova[a,*] and A. Konushin[a,b,]**

[a] *National Research University Higher School of Economics, Moscow, 101000 Russia*

[b] *Lomonosov Moscow State University, Moscow, 119991 Russia*

**e-mail: ale4kasokolova@gmail.com*

***e-mail: anton.konushin@grapics.cs.msu.ru*

Received February 16, 2019; revised March 23, 2019; accepted March 23, 2019

Abstract—Human gait is an important biometric index that allows to identify a person at a great distance without direct contact. Due to these qualities, which other popular identifiers such as fingerprints or iris do not have, the recognition of a person by the manner of walking has become very common in various areas where video surveillance systems can be used. With the development of computer vision techniques, a variety of approaches for human identification by movements in a video appear. These approaches are based both on natural biometric characteristics (human skeleton, silhouette, and their change during walking) and abstract features trained automatically which do not have physical justification. Modern methods combine classical algorithms of video and image analysis and new approaches that show excellent results in related tasks of computer vision, such as human identification by face and appearance or action and gesture recognition. However, due to the large number of conditions that can affect the walking manner of a person itself and its representation in video, the problem of identifying a person by gait still does not have a sufficiently accurate solution. Many methods are overfitted by the conditions presented in the databases on which they are trained, which limits their applicability in real life. In this paper, we provide a survey of state-of-the-art methods of gait recognition, their analysis and comparison on several popular video collections and for different formulations of the problem of recognition. We additionally reveal the problems that prevent the final solution of gait identification challenge.

DOI: 10.1134/S0361768819040091

1. INTRODUCTION

The problem of people identification in video by the gait is relevant in today's world. According to biometric research the manner of walk is individual for each person and is almost impossible to fake, which makes the gait a unique identifier such as fingerprints or iris. However, unlike these classical features the gait can be observed from afar without any direct contact with a person, therefore, it is the gait that becomes the most applicable index for recognition with the high-quality surveillance system development. The main application field of gait recognition is security area, where it is often necessary to identify a human being captured by the camera, for example, to catch the criminals or control the access to restricted areas. Gait recognition is a very specific problem due to many factors changing the gait visually (presence of the heels or uncomfortable shoes, carried heavy objects, clothing that hides parts of the body) or affecting the internal gait representation in the model (view, lightning, various camera settings). Therefore, despite the success the modern computer vision methods, the problem of identification by gait is not yet solved. This paper provides an overview of methods for recognizing a person by gait in a video and their comparison on popular datasets.

Currently, there are two main approaches to obtaining gait features and their classification: the hand-crafted construction of the descriptors and their training. The first method is more traditional and is usually based on the calculation of various properties of silhouette binary masks or on the study of joints positions, their relative distances and speeds and other kinetic parameters. Feature training is usually made by artificial neural networks that have become very popular in recent years due to outstanding results in many computer vision problems, such as video and image classification, image segmentation, object detection, visual tracking, and others. Features that are trained using neural networks often have a higher level of abstraction which is necessary for high-quality recognition. In addition, high quality of identification is achieved by methods combining the two described approaches. Initially, the basic characteristics of the gait are computed manually, and then they are fed to a neural network to extract more abstract features. Despite the success of deep methods, the best results on some datasets are still achieved by non-deep algo-

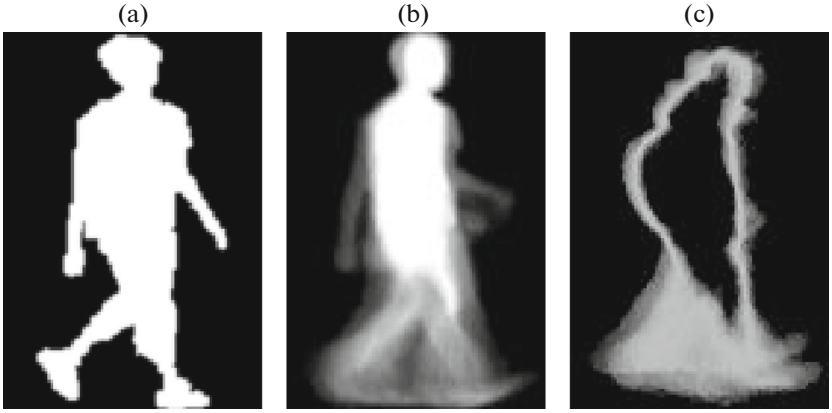

Fig. 1. The examples of basic gait descriptors: a) binary silhouette mask, b) gait energy image, c) gait entropy image.

rithms, so both global approaches are worthy of attention.

2. BASIC GAIT FEATURES

Let us firstly discuss some classical basic approaches in which the gait descriptors are extracted manually for natural reasons.

2.1. Binary Silhouettes

The most common gait characteristic is the gait energy image (GEI) [11], the average over one gait cycle of binary silhouette masks of moving person. Such spatial-temporal description of human gait is computed under the assumption that the movements of the body are periodically repeated during the walking. The resulting images characterize the frequency of a moving person being in a particular position. This approach is widely used and many other gait recognition methods are based on it. Besides this, many approaches not using gait energy images directly, suggest similar aggregation of other basic features. For example, the recognition can be made by the gait entropy images (GEnI) [2], where the entropy of each pixel is calculated instead of silhouette averaging, or by the frame difference energy image (FDEI) [4] reflecting the differences between silhouettes in consecutive frames of video. The visualization of binary silhouette mask and the images of gait energy and gait entropy is shown in Fig. 1.

Despite the equal simplicity and naturalness of all these methods, it is the images of energy that are used and developed so far. Similarly to ordinary black-and-white images, GEI can be used for further features calculation, such as histograms of oriented gradients (HOG-descriptors) [7, 16] or histograms of optical flow (HOF-descriptors) [14, 32], or for constructing more complex classification algorithms that use the specifics of the gait recognition problem.

So, one of the most successful multi-view approaches are two non-deep methods that use GEI as basic features. The first one is the Bayesian approach suggested by Lee in [15]. The authors propose to consider the gait energy images as random matrices being the sum of the gait identity and noise independent of it, corresponding to different conditions (such as viewpoint, different clothes or the presence of carried things), and it is assumed that both terms are normal random variables. Considering the joint distribution of two gait representations under the assumption that the corresponding classes coincide or differ reduces the problem to the optimization problem for the covariance matrices that can be solved using the EM algorithm. In the second approach, [19] it is proposed to generalize the method of linear discriminant analysis [3] by carrying out a multi-view discriminant analysis. For gait features computed for each viewing angle, a separate embedding is learned to minimize the intraclass variation and maximize the interclass one.

The idea of reducing the intraclass distances and increasing interclass ones is also applied in a later work [18], where a unified framework is proposed for learning the metrics of joint intensity of a pair of images and the spatial metric. The sequential optimization of both metrics leads to a model that surpasses the basic discriminant analysis models in recognition quality.

The described approaches are intuitively clear and mathematically simple, which, additionally to high recognition results, gives them an advantage over many other more complex methods. A common drawback of methods using GEI for multi-view recognition is the need to calculate the gait energy image for each viewing angle present in the database. Therefore, for each frame of the video you need to know the shooting angle, which is not always possible in real data.

2.2. Human Pose

Another important source of information, in addition to silhouettes, is the human skeleton. Many rec-

ognition methods are based on human pose investigation: the position of the joints and the main parts of the body and their motion in the video while walking. Approaches based on posture range from fully structural (considering the kinematic characteristics of the pose) to more complex, combining the kinematic and spatial-temporal characteristics. The work [1], in which the recognition takes place both by gait and by appearance, can be included in the first group. The main characteristics of the gait used in this approach are absolute and relative distances between the joints, as well as features based on the displacement of key points of the figure between frames. The approach proposed in [30] also explores the human skeleton, but the authors introduce a more complex mathematical model, considering a family of smooth Poisson distance functions for constructing a skeleton variance image (SVI).

Several other works also propose the models based on the position of human body parts, but use them in conjunction with features of a different nature. For example, the method [8] combines a kinematic approach with a spatial one, considering both the trajectories of the joints movement and the change in silhouette shape over time as dynamic features of gait. Nevertheless, if the information about the shooting angle is known, it can be effectively used by applying the view transformation model, as suggested in several approaches [21, 22]. For the gait features of one person corresponding to different angles, a transformation is trained that transforms one into another. Due to such transformations, descriptors corresponding to different views can be embedded in a common subspace which makes the classification more accurate.

2.3. Body Points Trajectories

Another non-deep approach showing the high quality of recognition was proposed in [6], where the trajectories of the movement of points of a human figure are considered and the Fisher motion descriptors are built on them, which are classified by the support vector machine.

3. NEURAL NETWORK APPROACHES

Despite the abundance of structural non-deep approaches, convolutional neural networks (CNN) have a strong position in all tasks of computer vision, including gait recognition. Over the past few years, many neural network identification methods for gait have been proposed, differing both technically (by choosing network architectures, loss functions, training methods) and ideologically by the method of data processing and extracting the primary features supplied to the network input. Due to the fact that the shape and the appearance of a person can vary depending on the wearable clothing and lighting, it is important for the model to pay attention not so much to external parameters as to the motion of the person's figure. Therefore, most of the methods classify video not directly by frames, but calculate all sorts of dynamic characteristics of the gait and recognize the person from them. One of these characteristics, providing information about the motion, is the optical flow, the vector field of the visible motion of the scene points. Its advantage is that the model trained on such data does not pay attention to the color, brightness or contrast of video frames. The effect on recognition is exerted only by the movement of individual points of the figure, and this is precisely what constitutes the gait of a person. In several papers [5, 25] that appeared almost simultaneously, it is proposed to consider blocks of optical flow maps similarly to the temporal component of the classical two-stream model [24] for action recognition. For several consecutive pairs of adjacent frames, the optical flow is calculated and a block of several flow maps is built. For accuracy improvement, a patch containing a human figure in all the frames is cut out of this block, and a neural network is trained on such patches. At the testing stage, the network is used to extract neural features, that can then be classified by any machine learning algorithm, for example, by the support vector machine (SVM) or the nearest neighbor method (kNN). The approach proposed in [26] develops the ideas of [25], but refuses blocks of consecutive frames. Instead, the movement of points near important body parts (chosen experimentally) is investigated in more detail. During the preprocessing, the human pose is evaluated and the optical flow is considered around the human feet, as well as separately in the upper and lower parts of the body (above and below the hips, respectively). The research shows that examining the flow on a larger scale around more "influential" parts provides a significant increase in the quality of recognition.

The second and most popular source of information, used for neural network training, is the binary masks of silhouettes, which were already discussed when considering non-deep methods. In the simplest case [37], the convolutional architecture is trained to predict the individual by separate silhouettes. Similar to the previous methods, the network is further used to extract features, and the aggregation of individual frame descriptors over the entire video occurs by selecting the maximum response over the gait cycle. This method is the simplest of all deep approaches, since when the silhouette masks of people are available no additional preprocessing is required. The length of gait cycle is determined by considering the autocorrelation of a sequence of binary images. Due to the fact that two frames that differ by the full gait period should look similar, the correlation of such frames is larger than that of any other pair, which helps to calculate the cycle length. Another method using the silhouettes themselves is proposed in [28]. The two-step algorithm firstly determines the shooting angle of the video, and then it predicts the person basing on the

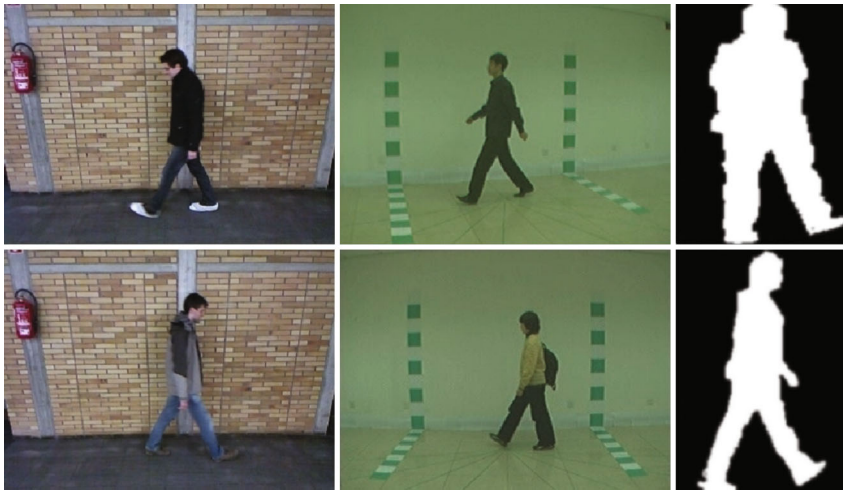

Fig. 2. The examples of frames from three databases: TUM-GAID (left), CASIA Gait Dataset B (middle) and OULP (right).

initial data and model for the angle found (its own for each one). To consider not only the spatial, but also the temporal characteristics of the motion, not separate masks, but blocks of several consecutive silhouettes are fed to the input of the network. In addition, the variability between frames is taken into account due to the network architectures in both subtasks: the authors use three-dimensional networks in which convolutions are performed not only in spatial domain, but also in temporal one.

It is also worth to highlight a variety of methods combining the "manual" feature extraction and deep learning. The GEI mentioned above are often used as input for networks. Models based on them range from the simplest ones [23], where a small network predicts a person from the gait energy images calculated for different angles, to more complex ones, such as [31], where the similarity of a GEI images is determined and various methods of comparing neural network gait features obtained from energy images are being investigated. In [27, 36], also using Siamese architectures with two or three streams, it is determined which gait images are close and belong to the same person, and which images belong to different people.

It is worth noting that most of the architectures and approaches popular in recent years are successfully used for gait recognition. For example, autoencoders used for image generation and representation learning are also applicable for identification. The authors of [34] propose to solve the problems of variability of views and carried items using a variety of autoencoders, each performing its own transformation similar to view transformation models [21]. Thus, passing through a sequence of "coding" layers, the image is transformed to a side view, which is easier for recognition. A similar approach is proposed in [33], however, generative adversarial models (GAN) are used to convert the GEI to the "basic" view (lateral viewing angle, no bag and coat). The [10] method is also based on adversarial models, but it solves the verification problem by evaluating and transforming the shooting angles to calculate features specific to each view. In addition, the authors build the period energy image (PEI), averaging the silhouettes over short time intervals within the gait cycle. This approach gives a noticeable increase in quality compared to the gait energy images used in most of the methods.

Additionally to the classical forward propagation convolutional networks, recurrent neural networks can be used for gait recognition as well as for other video analysis tasks. Recurrent architectures allow to calculate informative dynamic gait features even for very simple data (for example, silhouettes in individual frames [29]). In a more advanced approach, recurrent layers are applied to the human skeleton, namely, heatmaps for the joints obtained on previous convolutional layers from individual frames. Such a model makes a prediction based on a change in a person's pose, not relying directly on the figure and silhouette of a person, which makes it more general.

Many of the discussed approaches are evaluated on the same datasets under the same conditions, in this

Table 1. Comparison of recognition results on TUM-GAID dataset

Method	Accuracy
Sokolova, OF blocks [25]	97.5%
Sokolova, pose-based [26]	99.8%
Castro, SNN + SVM [6]	98.0%
Marín-Jiménez [20]	98.9%
Castro, Fisher descriptors [6]	99.2%
Zhang [37]	97.7%

Table 2. Comparison of recognition results on TUM-GAID videos taken in different days

Method	Accuracy
Castro, CNN + SVM [5]	59.4%
Marín-Jiménez [20]	63.6%
Castro, Fisher descriptors [6]	60.4%

Table 3. Comparison of recognition results on OU-ISIR database

Method	0	10	20	30
Zhang [37]	94.1%	71.6%	21.8%	2.9%
Shiraga [23]	94.9%	93.9%	90.5%	80.65%
Li [15]	98.3%	98.2%	97.3%	94.6%
Mansur [19]	96.8%	96.3%	94.2%	90.3%

Table 4. Comparison of cross-validation results on OU-ISIR database

Method	0	10	20	30
Sokolova [26]	98.4%	98.2%	97.1%	94.1%
Shiraga [23]	96.5%	95.8%	92.5%	84.9%
Wu [31]	98.9%	95.5%	92.4%	85.3%
Takemura [27]	99.3%	99.2%	98.6%	96.9%
He [10]	–	96.7%	93.2%	82.4%

Table 5. Comparison of average results for three viewing angles from CASIA database

Methods	54	90	126
Sokolova [26]	77.8%	68.8%	74.7%
Wu [31]	77.8%	64.9%	76.1%
Feng [9]	52.2%	60.0%	61.9%
Yu, SPAE [34]	63.3%	62.1%	66.3%
Yu, GaitGAN [33]	64.5%	58.2%	65.7%

review we compare some of the described methods and highlight the most successful solutions.

4. GAIT DATASETS

Currently, the most widely used complex datasets for gait recognition are TUM-GAID [12], OU-ISIR Large Population Dataset (OULP) [13] and CASIA Gait Dataset B [35]. Examples of video frames from these databases can be found in Fig. 2.

The first database is used for side-view recognition, all videos there are shot at an angle of 90, it is not very big (305 people, 10 videos for each one), but it consists of full-color videos, which makes it applicable to a large number of approaches. In addition, in this database there are videos taken with six month interval, which makes it possible to check the stability of the algorithms to temporal gait changes. Two other sets are collected for multi-view recognition. While CASIA is a relatively small base in terms of the number of people, but with a very large variation of views (11 different shooting angles from 0 to 180 degrees for 124 people), the OULP set consists of video sequences for more than 4.000 people taken with two cameras, and the shooting angle varies smoothly from 55 to 85. The data from this collection is distributed in the form of silhouette masks, therefore not all the described methods are applicable to this database. Many of the considered methods are estimated on these datasets, therefore we will give a comparison of approaches on them as well.

5. RESULTS COMPARISON

The performance of the algorithms on the described databases is evaluated as follows. First, the model is trained on the part of the data (usually this is all the video for a certain subset of people), and then tested on another part of the dataset consisting of video for other people. For TUM and OU-ISIR databases, the division into training and test subjects is provided by the authors of the collections. In experiments with CASIA, the model is trained on the first 24 people and then tested on the remaining 100. Recognition accuracy which is the proportion of correctly classified videos is usually considered as the metrics of quality.

Table 1 shows a comparison of recognition results based on TUM-GAID. The neural network approach [26] achieves the best quality, however, until its introduction, the deep methods could not surpass the [6] method, which does not use neural networks, for a long time. As will be seen below, in the problem of multi-view recognition, the struggle between deep and non-deep methods still continues.

It is also interesting to consider the temporal stability of the algorithms. It turns out that the quality of identification is badly deteriorated, if it takes a long time between the first time person has being captured by the camera and the moment of the test shooting and recognition. Table 2 shows the results of recognition based on TUM-GAID, when a half a year passes between the "training" and "test" videos. The accuracy of each of the presented algorithms falls by about 40%, which means that the features learned by these algorithms are poorly transferable in time.

For multi-view databases, the comparison is usually made for various pairs of angles: the data taken at some "test" angle is classified by the model fitted using a different, "training" shooting angle. To compare different algorithms based on OU-ISIR, two popular testing protocols are used. One of them, as already mentioned, was provided by the authors: 1.912

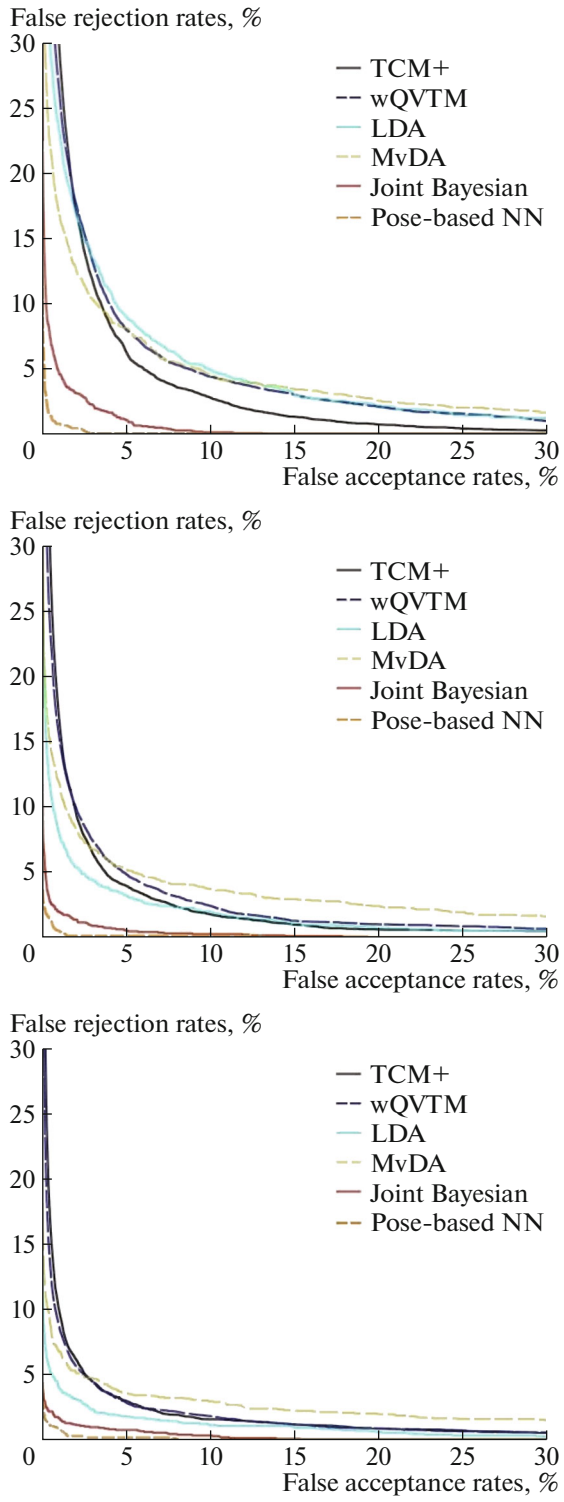

Fig. 3. Comparison of ROC-curves for different methods of cross-view verification task on OULP database for 85 gallery view and 55, 65, and 75 probe view, respectively.

people were selected from the collection; they are divided in half into training and test samples in five ways, after which the quality of the models trained on these partitions is averaged. The second protocol implements cross-validation, and the models are based on data for 3.844 people for whom videos captured by both cameras are present in the database. For convenience, comparison results are usually aggregated by considering the differences between the "training" and "test" angles. Table 3 shows the average accuracy of the methods for each of the 4 possible values of the angle difference.

The simplest method [37] turns out to be inconsistent when shooting angles differ a lot, but other approaches show a fairly high recognition quality. The best results in such experiments are also achieved by the method that does not use neural networks. However, when more data is available for training and testing, neural network methods are very successful. Table 4 shows the results of the comparison of algorithms when using data for almost 4 thousand people. Due to the larger size of the training set, methods using such a testing protocol achieve higher accuracy. The absence of open implementations and the general test protocol makes it impossible to compare all the methods and find the best one, but even the available results show that today deep and non-deep approaches continue to evolve and show almost equal quality.

For the CASIA database, the recognition quality for various angles is usually aggregated as well. Following the common approach, in the Table 5 we present the average values of recognition accuracy for probe angles of 54, 90 and 126 (the remaining 10 viewing angles are used as gallery ones in each experiment).

Additionally to the classical classification problem, in which it is necessary to determine which person from the database is shown in test video, the gait recognition task is often formulated in a form of a verification task. For a pair of video sequences with a moving person, it is required to determine whether there are different people or the same one. The verification task is interesting not only on its own, but also as a supplement to the identification in case a person is captured by the cameras for the first time and is not yet in the index. Even a person who is not included to the database will be somehow classified by an identifier, and evaluating the confidence of the model is a complex task. One of the possible approaches to the solution is additional verification of a pair consisting of a test video and a video with a candidate subject. For the task of verification, all described methods for gait features extraction can be used; however, instead of classifying or finding the nearest object, in the last stage, the similarity of a pair of descriptors is estimated and compared with a certain threshold. Relatively similar descriptors are considered to belong to the same person. To evaluate the quality in such a task, a ROC-curve is usually constructed and the area under this curve is computed. It is interesting to note that despite the fact that the same gait recognition problem is solved and the same descriptors are calculated, the approaches that show the highest results in the identi-

fication task may not be the most successful in verification. Figure 3 shows the ROC-curves for several methods of multi-view recognition on OU-ISIR database provided by their authors. The curve that corresponds to the [26] approach, which is a bit less accurate identifier than the Bayesian method [15], is the lowest on all graphs, which indicates that this algorithm more accurately determines whether people in the video coincide. This is one more confirmation that a single perfect method still does not exist and different approaches turn out to be better under different conditions and assumptions.

The results of the comparison of multi-view recognition methods show that different descriptors are competitive with each other in the information content. Methods based on the silhouettes solve the problem with almost the same accuracy as the approaches that consider the posture and movement of points in the frames, and sometimes even better.

6. CONCLUSION

Despite all the many features used and the diversity of the proposed models and training methods, the problem of gait recognition still does not lose relevance: the existing solutions have not yet reached the perfect accuracy of identification. The representation of motion is influenced by a large number of different conditions, and the datasets usable for this task are limited compared to other computer vision problems, for which millions of images of faces or tens of thousands of figures for reidentification are collected. The databases collected at the moment are not yet able to take into account all possible variations of gait, which prevents the creation of a perfect model.

REFERENCES

1. Arseev, S., Konushin, A., and Liutov, V., Human recognition by appearance and gait, *Programming and Computer Software,* 2018, pp. 258–265.
2. Bashir, K., Xiang, T., and S, G., Gait recognition using gait entropy image, *Proceedings of 3rd International Conference on Crime Detection and Prevention,* 2009, pp. 1–6.
3. Belhumeur, P.N., Hespanha, J.a.P., and Kriegman, D.J., Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.,* 1997, vol. 19, no. 7, pp. 711–720.
4. Chen, C., Liang, J., Zhao, H. Hu, H., and Tian, J., Frame difference energy image for gait recognition with incomplete silhouettes, *Pattern Recognit. Lett.,* 2009, vol. 30, no. 11, pp 977–984.
5. Castro, F.M., Marín-Jiménez, M.J., Guil, N., and Pérez de la Blanca, N., Automatic learning of gait signatures for people identification, *Advances in Computational Intelligence,* 2017. pp. 257–270.
6. Castro, F.M., Marín-Jiménez, M., and Medina Carnicer, R., Pyramidal Fisher Motion for multiview gait recognition, *22nd International Conference on Pattern Recognition,* 2014, pp. 692–1697.
7. Dalal, N. and Triggs, B., Histograms of oriented gradients for human detection, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),* 2005, vol. 01, pp. 886–893.
8. Deng, M., Wang, C., Cheng, F., and Zeng, W., Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning, *Pattern Recognit.,* 2017, vol. 67, pp. 186–200.
9. Feng, Y., Li, Y., and Luo, J., Learning effective gait features using LSTM, *International Conference on Pattern Recognition,* 2016, pp. 325–330.
10. He, Y., Zhang, J., Shan, H., and Wang, L., Multi-task gans for view-specific feature learning in gait recognition, *IEEE TIFS,* 2019, vol. 14, no. 1, pp. 102–113.
11. Han, J. and Bhanu, B., Individual recognition using gait energy image, *IEEE Trans. Pattern Anal. Mach. Intell.,* 2006, vol. 28, pp. 316–322.
12. Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., and Rigoll, G., The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits, *J. Visual Com. Image Repres.,* 2014, vol. 25, no. 1, pp.195–206.
13. Iwama, H., Okumura, M., Makihara, Y., and Yagi, Y., The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition, *IEEE Trans. on Information Forensics and Security,* 2012, **7**, Issue 5, pp. 1511–1521.
14. Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B., Learning realistic human actions from movies, *IEEE Conference on Computer Vision & Pattern Recognition (CVPR 2008),* 2008, pp. 1–8.
15. Li, C., Sun, S., Chen, X., and Min, X., Cross-view gait recognition using joint Bayesian, *Proc. SPIE 10420, Ninth International Conference on Digital Image Processing (ICDIP 2017),* 2017.
16. Liu, Y., Zhang, J., Wang, C., and Wang, L., Multiple HOG templates for gait recognition, *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012. pp. 2930–2933.
17. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., and Yagi, Y., Gait recognition using a view transformation model in the frequency domain, *Computer Vision – ECCV 2006,* 2006, pp. 151–163.
18. Makihara, Y., Suzuki, A., Muramatsu, D., Li, X., and Yagi, Y., Joint intensity and spatial metric learning for robust gait recognition, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2017, pp. 6786–6796.
19. Mansur, A., Makihara, Y., Muramatsu, D., and Yagi, Y., Cross-view gait recognition using view-dependent discriminative analysis, *2014 IEEE/IAPR International Joint Conference on Biometrics (IJCB 2014),* 2014.
20. Marín-Jiménez, M., Castro, F., Guil, N., de la Torre, F., and Medina Carnicer, R., Deep multi-task learning for gait-based biometrics, *IEEE International Conference on Image Processing (ICIP),* 2017.
21. Muramatsu, D., Makihara, Y., and Yagi, Y., View transformation model incorporating quality measures

for cross-view gait recognition, *IEEE Transactions on Cybernetics,* 2015, vol. 46.

22. Muramatsu, D., Makihara, Y., and Yagi, Y., Cross-view gait recognition by fusion of multiple transformation consistency measures, *IET Biometrics,* 2015, vol. 4.
23. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., and Yagi, Y., GEINet: View-invariant gait recognition using a convolutional neural network, *2016 International Conference on Biometrics (ICB),* 2016, pp. 1–8.
24. Simonyan, K. and Zisserman, A., Two-stream convolutional networks for action recognition in videos, *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14),* 2014, vol. 1, pp. 568–576.
25. Sokolova, A. and Konushin, A., Gait recognition based on convolutional neural networks, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 2017, vol. **XLII-2**/W4, pp. 207–212.
26. Sokolova, A. and Konushin, A., Pose-based deep gait recognition, *IET Biometrics,* 2018.
27. Takemura, N., Makihara, Y., and Muramatsu, D., On input/output architectures for convolutional neural network-based cross-view gait recognition, *IEEE Trans. Circuits Syst. Video Technol.,* 2017, vol. 1, p. 1.
28. Thapar, D., Nigam, A., Aggarwal, D., and Agarwal, P., VGR-net: A view invariant gait recognition network, *IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA),* 2018, pp. 1–8.
29. Tong, S., Fu, Y., Ling, H., and Zhang, E., Gait identification by joint spatial-temporal feature, *Biometric Recognition,* 2017, pp. 457–465.
30. Whytock, T., Belyaev, A., and Robertson, N.M., Dynamic distance-based shape features for gait recognition, *J. Math. Imaging and Vision,* 2014, vol. 50, no. 3, pp. 314–326.
31. Wu, Z., Huang, Y., Wang, L., Wang, X., and Tan, T., A comprehensive study on cross-view gait based human identification with deep cnns, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 2016, p. 39.
32. Yang, Y., Tu, D., and Li, G., Gait recognition using flow histogram energy image, *22nd International Conference on Pattern Recognition,* 2014, pp. 444–449.
33. Yu, S., Chen, H., Reyes, E.B.G., and Poh, N., GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Network, *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2017, pp. 532–539.
34. Yu, S., Chen, H., Wang, Q., Shen, L., and Huang, Y., Invariant feature extraction for gait recognition using only one uniform model, *Neurocomputing,* 2017, vol. 239, pp. 81–93.
35. Yu, S., Tan, D., and Tan, T., A Framework for evaluating the Effect of view angle, clothing and carrying condition on gait recognition, *Proc. of the 18'th International Conference on Pattern Recognition (ICPR),* 2006, vol. 4, pp. 441–444.
36. Zhang, C., Liu, W., Ma, H., and Fu, H., Siamese neural network based gait recognition for human identification, *IEEE International Conference on Acoustics, Speech and Signal Processing,* 2016, pp. 2832–2836.
37. Zhang, X., Sun, S., Li, C., Zhao, X., and Hu, Y., Deepgait: A learning deep convolutional representation for gait recognition, *Biometric Recognition,* 2017, pp. 447–456.