CCF Transactions on Pervasive Computing and Interaction
https://doi.org/10.1007/s42486-020-00031-5

REGULAR PAPER

NMF based image sequence analysis and its application in gait recognition

Enyuan Cao[1] · Kun Cao[2] · Kuining Feng[1] · Jiawei Wang[3]

Received: 25 September 2019 / Accepted: 27 May 2020

Abstract
Human identification is an important part of the intelligence system, and among them, gait recognition is more suitable in pervasive intelligence, due to its capability in identifying with low resolution images captured at a distance without subject cooperation. Existing human gait recognition methods are either too simple that cannot cope with complex scenes, or too complex to provide service at the right time and right where. In this paper, we propose a simple but efficient method for gait recognition. It not only has a high recognition accuracy, but also has a strong adaptability to different conditions, which means it can better adapt to the real environment. After data pre-processing, the method leverages nonnegative matrix factorization (NMF) to pick up high order image features. Then, it sends frames in the order of time sequence to a long short term memory (LSTM) neural network for performing classification. Experimental results show that this method can achieve high recognition accuracy, and has a strong adaptability in cross-view and multi-clothes conditions.

Keywords Gait recognition · NMF · LSTM

1 Introduction

Pervasive computing (Becker et al. 2019) is a kind of computing technology that integrates with the environment. It is a kind of computing that exists anywhere. Since a few years ago, artificial intelligence has become very popular, and has made significant progress. However, applying theoretical techniques to the real-world is not as simple as imagined. The problems such as fragmentation, high application costs, and gaps between theoretical and actual scenarios are exposed. Combining pervasive computing and artificial intelligence, also referred to as pervasive intelligence is a good solution to these problems. With the ultimate purpose of providing the services at right-time, right-where and by right-means, pervasive intelligence requires higher speed, higher accuracy and higher flexibility to suit different conditions (Wang et al. 2017, 2018; Yang et al. 2018).

✉ Enyuan Cao
caoey7@gmail.com

[1] University of Electronic Science and Technology of China, Chengdu, China

[2] East China Normal University, Shanghai, China

[3] St. Francis Xavier University, Antigonish, Canada

Human identification is an important part of universal intelligence. There are multiple methods in human identification, such as facial recognition, iris recognition, and fingerprint recognition. However, these methods cannot reliably identify individuals if they are not cooperative or at a distance, which is common in real-world applications. Compared to those identification methods listed above, the identification techniques developed on human gait have been considered to be a better choice due to its capability in identifying with low resolution images captured at a distance without subject cooperation (Johansson 1973; Bashir et al. 2010). Therefore, gait recognition is more suitable for providing services at the right time, in the right place, and in the right way due to its better adaptability.

Existing gait recognition approaches can be divided into two categories: silhouette-based approaches and model-based approaches. Model-based approaches try to construct a three dimensional model for the human, and then analyzes the movement of the model (Muramatsu et al. 2015; Gianaria and Grangetto 2019; Lam et al. 2011). A lot of recent research is about this area, and they have a relatively higher accuracy (Wolf et al. 2016; Liu etr al. 2016). However, these methods are very complex and high-dimensional, which would be unsuitable to provide service at the right time and right where.

Published online: 23 June 2020

The other category is Silhouette based approaches. These approaches have a relatively smaller computational complexity since they represent a series of images in a gait cycle by one single image. Traditional methods that used to represent gait information include Gait Energy Image (GEI) (Han and Bhanu 2005), Gait Entropy Image (GEnI) (Bashir et al. 2009), Enhanced Gait Energy Image (EGEI) (Chunli et al. 2010), Masked GEI (MGEI) (Bashir et al. 2010) and etc. GEI is the most classic of these methods. It is the average of a series of gait silhouette images in a gait cycle, simple and effective. GEnI, EGEI and MGEI are all based on GEI but highlight the areas of the movement in different ways to highlight the movement process. However, all these methods above will lose important information when averaging the silhouette, because they all try to put all the information from many images into a single image based on the calculated average only (Babaee and Rigoll 2017).

In summary, these existing gait recognition methods are either too simple to consider enough information to improve recognition accuracy, or too complex to be applied in real-time or deployed in devices with limited computing resources. There is no perfect method that is accurate enough and has relatively low computational cost, which can be done in local or embedded systems.

This paper proposes an improved silhouette-based method, which can achieve higher recognition accuracy and strong adaptability to the environment, with the computational complexity at the same level as other silhouette-based approaches. The first stage uses nonnegative matrix factorization (NMF) to learn image features, and the second stage uses long short term memory (LSTM) for classification. Experimental results show that the proposed method is an efficient and effective gait recognition method that is accurate enough and has a strong adaptability in cross-view and multi-clothes conditions.

The rest of the paper is organized as follows. Section 2 introduces the theoretical basis of NMF. Section 3 introduces the structure of our method. Section 4 introduces the advantages of our method. Section 5 shows the experimental results and comparison with other methods. The conclusion is given in Sect. 6.

2 Theoretical basis of nonnegative matrix factorization

Nonnegative matrix factorization (NMF) is a classic method to pick up high order features of images, and it works well in facial recognition area (Lee and Seung 1999, 2001). The core theory is that every matrix $V_{(F\times N)}$ can be represented by the production of two smaller matrices, as shown in Eq. (1).

$$V_{(F\times N)} \approx W_{(F\times K)} \times H_{(K\times N)} \tag{1}$$

V is the original matrix. W and H are smaller matrices generated by factorising V. F, N and K are the sizes of these matrices. Through multiple self-iterations, the product of two smaller matrices will be approximately equal to the original input matrix. Under the NMF rule, the two smaller matrices should contain features and weights, respectively. That is, the original matrix is divided into features and their corresponding weights, as shown in Fig. 1. The iterative process of nonnegative matrix factorization is shown below.

$$W_{ia} \leftarrow W_{ia} \Sigma_u \frac{V_{iu}}{WH_{iu}} H_{au} \tag{2}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\Sigma_j W_{ja}} \tag{3}$$

$$H_{au} \leftarrow H_{au} \Sigma_i W_{ia} \frac{V_{iu}}{WH_{iu}} \tag{4}$$

where u, a, i and j are auxiliary numbers used to represent the position of the elements in the matrices. The iterative process begins with the initial conditions of W and H. The iteration of the update rule in the above equations will gradually converge to the result $V \approx WH$. The initial conditions of W and H are non-negative, and the update rules maintain the

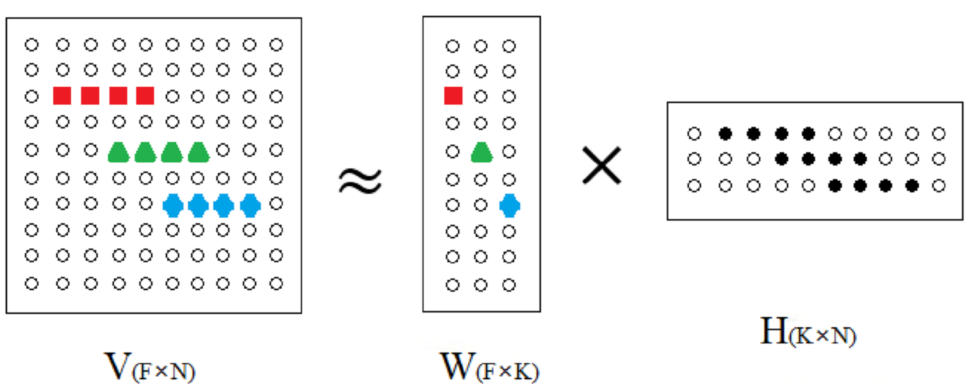


Fig. 1 The structure of NMF. In the picture, W matrix contains the features. Suppose that in H matrix, white points are zeros and black points are ones. The V matrix is W matrix times H matrix

non-negativity of W and H (Lee and Seung 1999). After the factorization, two matrixes W and H will be generated. W is the set of high order features, and H is the set of weights. The original matrix V will be represented by the production of W and H after this process.

3 NMF based sequence analysis method

The method can be divided into three steps. The first step is data pre-processing. Because the NMF method cannot distinguish the location of features, the silhouette of human have to be selected out by finding the gravity center. The second step uses NMF to pick up high order features of each single image. When using this method on images, the data should be extracted from the images to the matrix first.

The third step is nonlinear classification. Because each image is already represented by a series of weights, in this step, only those weights need to be classified. We choose the LSTM network for classification, because LSTM is good at analyzing timing information (i.e., the relationship between the current input and the previous input). The flow chart of the method is shown in Fig. 2.

3.1 Image pre-processing

The original image of the human gait cannot be sent directly to the NMF for processing. On the one hand, the features learned by NMF have a certain position, and if the position changes, it will be recognized as a completely new feature. On the other hand, the original image is usually much larger than the human contour, which means that large areas on the image are useless. Thus, in order to improve the calculating speed and reduce the influence of the position of the individual on the image, the image should be cut before being processed.

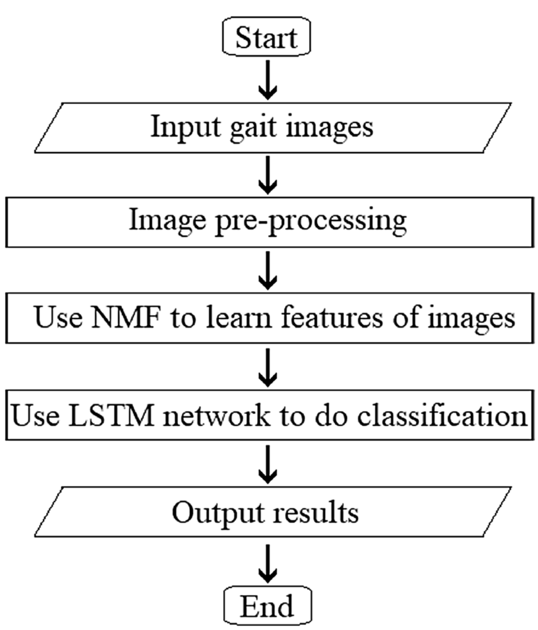


Fig. 2 The flow chart of our method

To this end, we try to find the center of gravity. The position of the gravity center in physics is defined as the sum of the production of mass for each point and its corresponding position over the total mass of the system. The process of finding the gravity center of a multidimensional object is to find the gravity center in each dimension separately, using the following equations

$$x = \frac{\Sigma m_i \times x_i}{\Sigma m_i} \tag{5}$$

$$y = \frac{\Sigma m_i \times y_i}{\Sigma m_i} \tag{6}$$

where x and y are coordinate values corresponding to the position of gravity center in the 2D image coordinate. m_i is the mass at x_i or y_i point. For an image, mass m_i means the value of the pixel. Equation (5) aims at finding the position of the gravity center on x axis and equation (6) targets at finding that on y axis. After finding the center of gravity, we select an area with 80×150 pixels centered on the center of gravity. This size of the area perfectly fits the silhouette of human in each image. The sample image is shown in Fig. 3.

3.2 Learning high order features

When using NMF for pattern recognition, each image should be placed in a column of the V matrix. To this end, each image has to be separated into columns of pixels, and arranged vertically to make up a big column, as shown in Fig. 4. Thus, in H matrix, each image is also represented by a column. The factorization process for gait images is shown in Fig. 5. These features are two-dimensional images with the same size as the original image, and each feature contains a part of the body outline. It can be clearly seen that the original images are represented by features and their corresponding weights, and the representation is approximately equal to their original images.

3.3 Analyzing movement of high order features

Since each image is represented by a series of weights, in this step, these weights will be sent to the classifier. Gait recognition cannot ignore time sequence information, and LSTM is good at processing time sequence information, so we chose a long-term short-term memory (LSTM) network to perform this step. For each input, first record the current

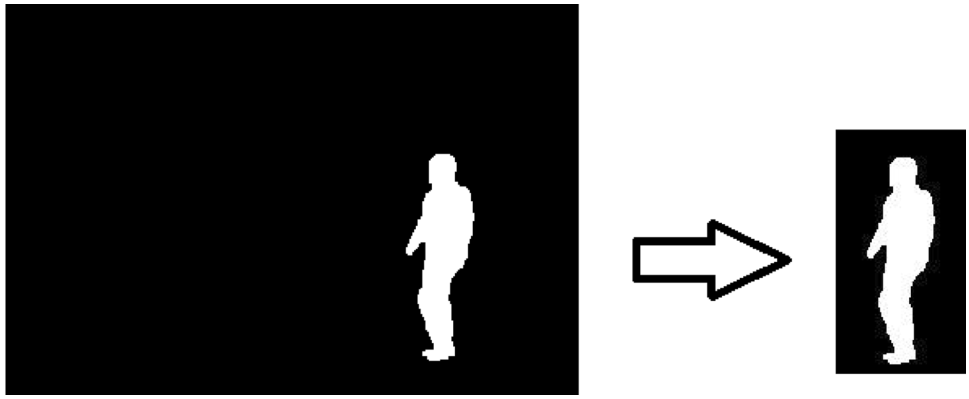


Fig. 3 The image after the cut process. The human will always be at the center and the size is perfectly cover the whole silhouette

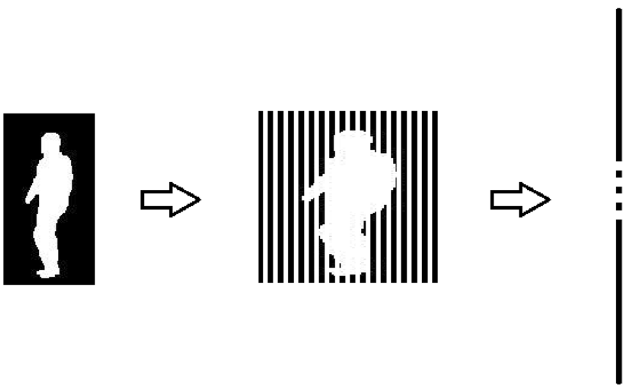


Fig. 4 Each image has to be separated into columns of pixels and arranged vertically to make up a big column. This column will be placed in V matrix to be factorized

state, and then send it to the network again with the next input, as shown in Fig. 6.

We choose to perform time series analysis after extracting the features, because the time series information obtained in this way is the movement of high-order features, not pixels. If you only analyze the pixel-by-pixel motion, it will be very susceptible to clothing and the viewing angle. Because in the gait contour map, pixel information is only limited in a few areas on the contour. The advantage of this method of using high-order feature motion information is that the information is more based on the whole image, and is not easily affected by slight disturbances or noise.

4 Analysis

Our method has two significant advantages, including better adaptability to the cross view conditions and smaller influence from clothes, as detailed below.

4.1 Better adaptability in cross-view condition

Generally, in silhouette-based methods, the training data is divided into several view degrees, as shown in Fig. 7. The correct recognition rate is very high when the testing angle is the same as the probe angle. However, most of the time, the testing angles are not perfect. As the testing angle moves away from the probe angle, the recognition rate will decrease. Therefore, most of these silhouette-based methods need to prepare several probe views as training data to make sure the correction rate will always within the acceptable region. In that case, the method has to have stronger adaptability with contiguous views to make sure the result will not be influenced much if the testing view lies between two probe views (Zifeng et al. 2015, 2016).

Compared with GEI or many other silhouette-based methods, our method has a significant advantage in dealing with this situation, due to the consideration of sequence information. In traditional gait recognition methods, for instance, GEI, the recognition process completely depends on the details of images. However, a slight change in the angle of view will make the silhouette completely different. Thus, if the view angle changes a little, the result will be completely different. With the use of LSTM, the higher-order features of movement are learned, so the influence of the change on silhouette will be less. The progress in cross-view conditions can lead to better stability for this algorithm, because the prediction result will be smoother when the individuals are tested by different direction, and the training process requires fewer probe view data.

4.2 Better adaptability in multi-clothing condition

In the traditional pattern recognition method, the result is easily affected by the occlusion of the individual. For example, the same person wearing a T-shirt or jacket will have a completely different silhouette, as shown in Fig. 8. This does not have much influence on facial or iris recognition, because the individuals are supposed to cooperate and their

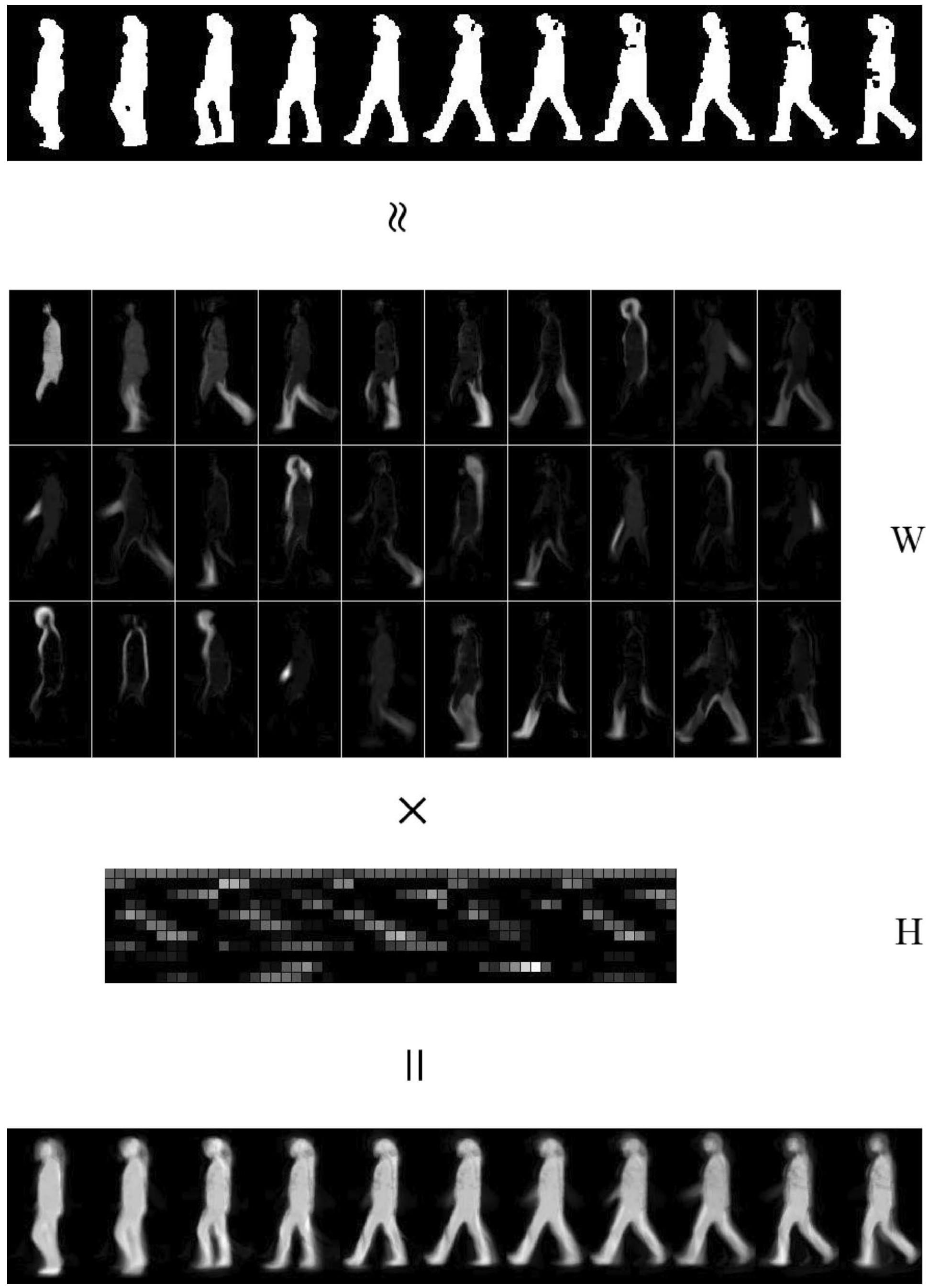


Fig. 5 Illustration of gait image factorization. The W matrix contains feature information such as legs, arms, and bodies. The H matrix contains weights information, in which each element is the weight of its corresponding feature. Thus, each image can be represented by a column in H. The original image is shown in the top of this figure. It is clear that the production of W and H is approximately equal to the original image

face or eye will be clearly exposed. However, humans will wear different clothes every day, and generally, individuals are not necessary to cooperate in gait recognition. Therefore, if the identification process is completely depended on the details of silhouette, the results will be affected a lot, because most of these details are about clothes. The recognition process will become recognizing a person wearing certain clothes. If the movement information is considered, the clothes will have less influence on the result, and the classification will depend more on the movement process of the body.

5 Experiments

We compared many kinds of methods on the three datasets of CASIA to thoroughly evaluate the performance of our method as a fast and efficient human identification method. The CASIA datasets are the largest datasets available for benchmarking gait recognition techniques. The platform to run the experiment is a common personal laptop, with CPU Intel Core i5 5200U. The advantages and disadvantages of different methods will be shown in this section.

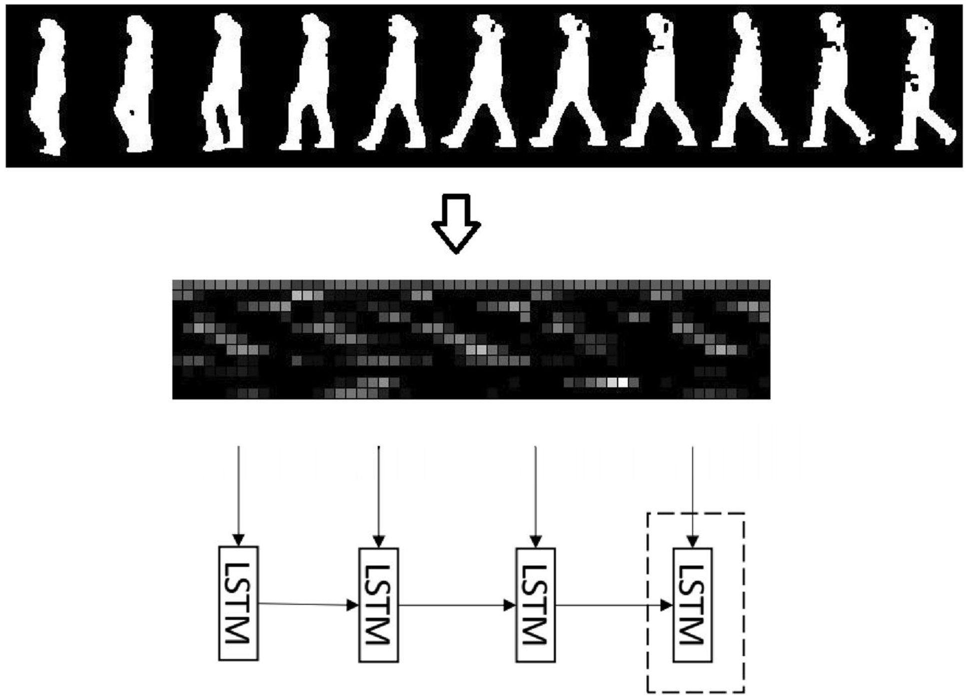


Fig. 6 The structure of LSTM classification process. After each image is represented by a column of weights, each column will be sent into LSTM recurrent neural network

Fig. 7 Different view angle on the same subject. The view angles changes from 0 to 180 degrees. The left image, middle image and right image show angles of 36 degree, 90 degrees and 144 degree, respectively

Fig. 8 Same subject but wearing different clothes. The left image, middle image and right image show normal condition, carrying bag and wearing coat, respectively

Fig. 9 Images that are taken from CASIA dataset A, dataset B and dataset C, respectively. The images from dataset B have higher quality while images from dataset C include more details

5.1 CASIA Datasets

The CASIA Gait Database is classic in gate recognition area. There are totally 3 sets of data, but were collected at different time. The sample images are shown in Fig. 9. This data collection contains comprehensive conditions, such as walking with a backpack or wearing a coat Wang et al. (2003). Dataset A is the earliest collected dataset containing the information of 20 subjects, each with 3 directions, 0 degrees, 45 degrees, and 90 degrees. Each direction has 4 series of data, which means there are 12 image sets for each subject. Each image set contains 50 to 70 images that are taken from a walking video and arranged by time sequence. Dataset B is the largest in number of images, and has a higher quality (the images are smoother). It contains 124 subjects, and each subject has 10 image set, including normal situation, carrying bag situation and wearing coat situation. Each set of images has 11 view angles. Dataset C is collected at night with an infrared (thermal) camera, which means the images from this dataset will be slightly different from the others, for example, the hairline and some other lines on the body are shown clearly.

The original images are about 240×350 pixels. Only 150×80 will be preserved, which means the length of columns in NMF matrix V is 12000. Under the normal case experiment, only the 90-degree direction is used, which is the same as the images in Fig. 10. Each person has 4 sets of images, containing many periods of step. In each dataset, a period of one step contains 11 to 13 images depending on the speed of walking. Thus, depending on the length of the step period, a number that can cover most of the step cycles will be selected. This number will be used as the length of LSTM training step.

5.2 Benchmark methods

In order to verify the improvement of our scheme under normal conditions and Multi-Clothing conditions, we performed a comparative experiment. We used two classic benchmark schemes to compare with our proposed scheme, one is the GEI scheme and the other is the traditional NMF scheme. In Cross-View experiments, because there are already sufficient experimental data in many related works, we will directly compare with other experiments.

GEI is a classic method of representing gait sequences. Assuming that $B_t(x, y)$ is used to represent the frame at time t in the preprocessed binarized gait silhouette image sequence, the gait energy image is defined as follows:

$$G(x, y) = \frac{1}{N} \Sigma_{t=1}^{N} B_t(x, y) \tag{7}$$

where N is the total number of frames in a complete step cycle, t is the number of frames (moment of time) in the sequence, and x and y are values in 2D image coordinates. For example, Fig. 10 shows a sequence of human gait silhouette images in a step cycle, where the rightmost image is the corresponding gait energy image. It can be clearly seen that GEI reflects the distribution of the contours in a complete

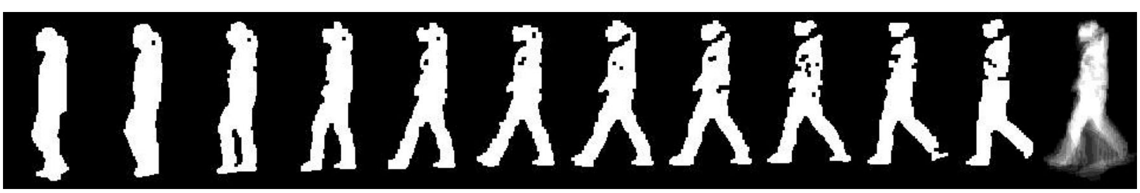


Fig. 10 An example of a sequence of human gait silhouette images in a step cycle. The rightmost image is the corresponding gait energy image

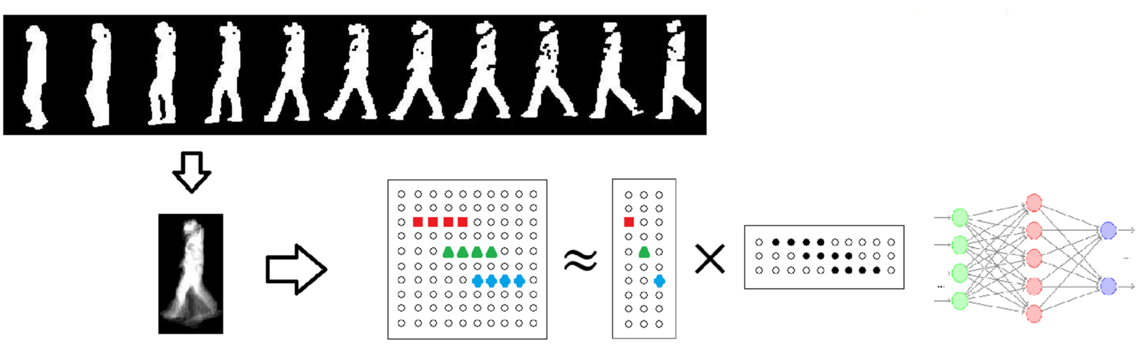


Fig. 11 Total structure of GEI method. This is sending the averaged images into NMF algorithm and using a fully-connected neural network to classify

gait cycle (Zhao et al. 2016). In order to keep the different algorithms in the same level of complexity, this result will also be sent to do NMF process, as shown in Fig. 11, which is the same as the traditional way.

Unlike method GEI, the traditional NMF method sends the weights directly into a fully-connected neural network, as shown in Fig. 12. It perform well in facial recognition; however, it does not take the relationship between images into consideration. In other words, its recognition process depends only on each single image.

5.3 Experimental results

For a fair comparison, we use 20 subjects from each dataset because there are only 20 subjects in dataset A. The size of the input image is $150 \times 80 = 12000$, and every 20 images are divided into a gait cycle. The total number of images for each object is different, so the number of gait cycles is also different, between 8 and 12. In the process of NMF, the number of features is set to be 40. In the LSTM process, the number of hidden neurons is set to be 128. During the experiment, the data set is randomly divided into 75% training set and 25% testing set. Use cross-validation to calculate the final accuracy.

Figure 13 presents the result that tested on dataset CASIA A, B, and C. The left one is under the normal condition, which means the probe view is the same as the test view, and there are no blocks for individual. The right one exhibits the result under condition that the individuals wearing coat and bags.

In every group, NMF-based sequence analysis methods are slightly better than GEI, and GEI is significantly better than traditional NMF. Under normal circumstances, the correction rate of GEI is 85 to 90% compared to traditional NMF, while the accuracy of GEI is 96%, which is much higher than the traditional NMF method. Using the same data and the same classification network has such a gap, which means that the outline of the person in a single image contains limited information. If you do not consider the entire gait cycle, then a lot of information will be ignored. GEI only considers the distribution area of pixels, but does not consider the moving process of pixels, so some information will still be lost. After using sequence analysis, the process of step-by-step image change was also considered, so the accuracy rate was further improved, reaching 98%. In the case of multiple garments, the recognition rate will drop significantly. One of the important reasons is that the experimental conditions are under extreme conditions, such as training on only one kind of normal clothes, and testing the jacket wearing another kind of clothes. In practical applications, the tester should include a variety of clothing to cover most situations.

In the Cross-View experiment, we compared our method with some other related work, including the

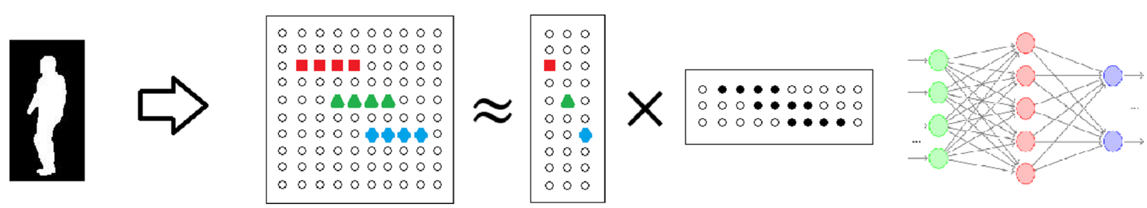


Fig. 12 The total process to do the identification. This method is, after the first NMF, connect the result directly into a fully-connected neural network

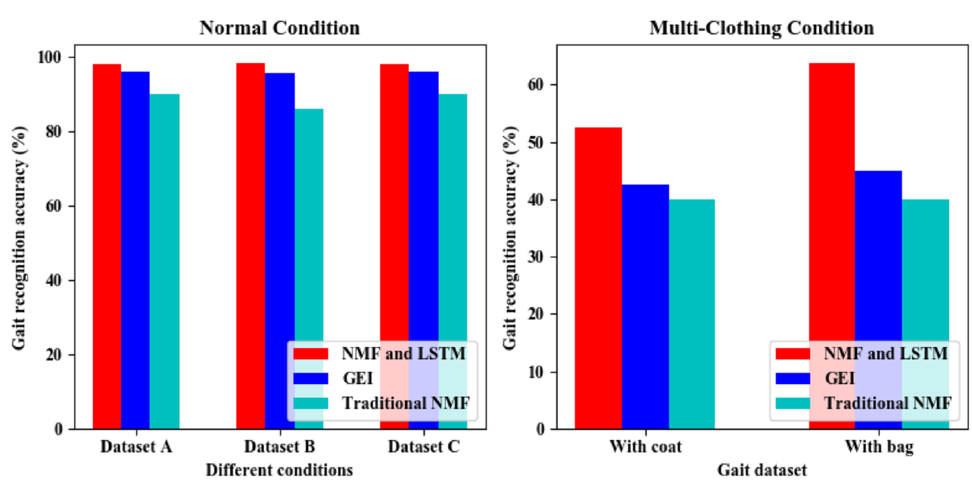


Fig. 13 Result of gait recognition using different methods on different datasets. The left one is under normal condition, and the right one is under multi-clothing condition

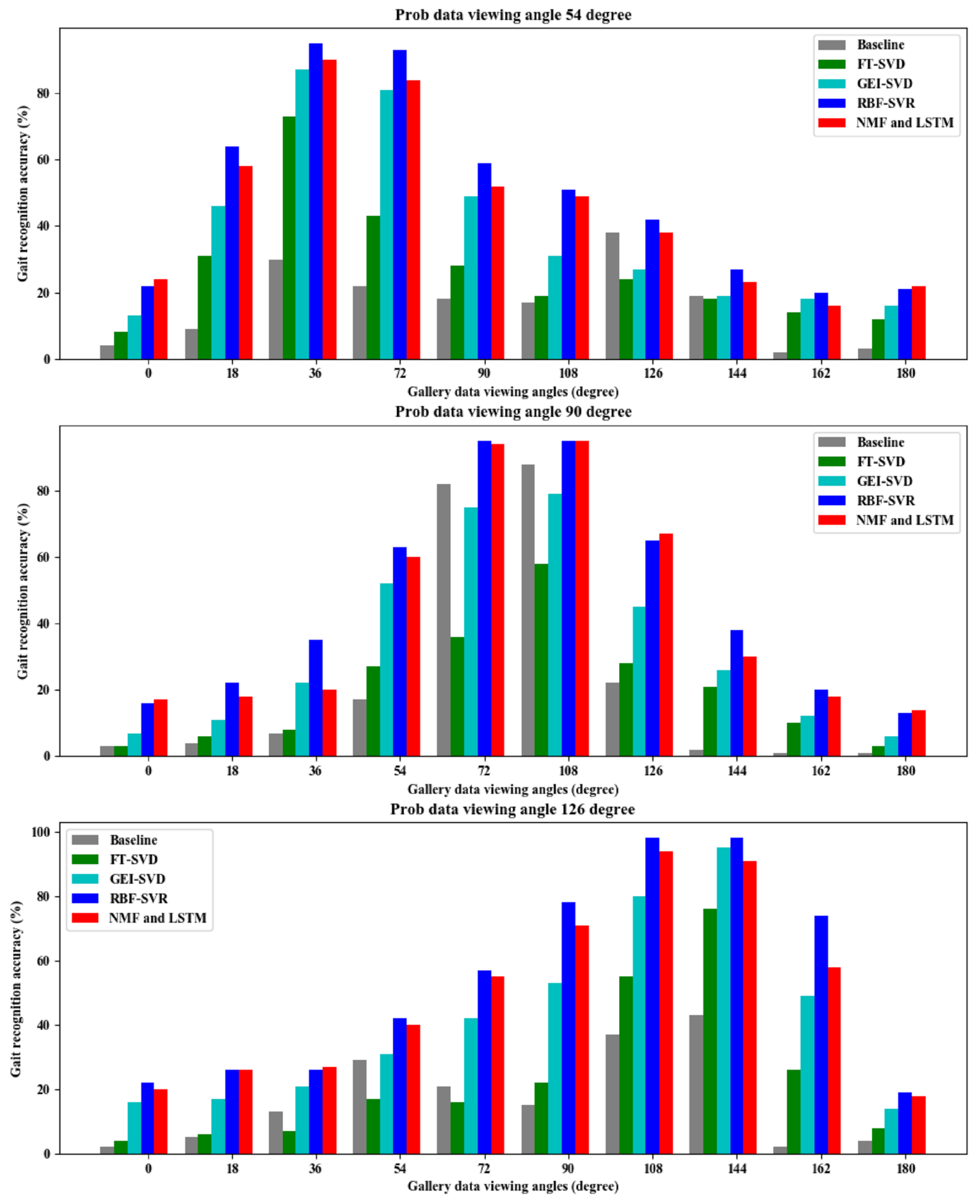


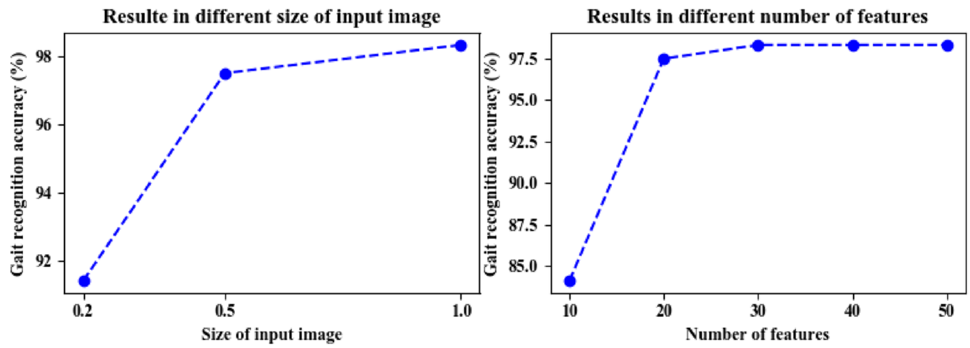
Fig. 14 Accuracy of cross-view gait recognition using our method and other competitive methods

Baseline method Yu et al. (2006), RT-SVD Makihara et al. (2006), GEI-SVD (Kusakunniran et al. 2009), and RBF-SVR (Kusakunniran et al. 2010). Figure 14 shows the results of the cross-view comparison. These methods are the most classic and competitive

It can be seen that for all these methods, the recognition rate under non-ideal conditions is not good. However, if more probabilistic views are used to construct the training data set, the net output may be relatively smooth. The performance of our method is similar to some powerful methods (such as RBF-SVR and GEI-SVD). However, those methods require more than 10 minutes of calculation time, which means that those methods will have very high requirements on the system and are not easily applied to the real world.

The main experimental setup is to use 40 features with a resolution of the original size. We also performed an experiment using different settings and different input sizes, as shown in Fig. 15 As expected, when using a different number of features and different image sizes, the accuracy of the method increases with the number of features and the image size increase. However, more features and larger images will also bring more computing cost, that is, longer computing time. Therefore, in practical applications, users should find a balance between accuracy and calculated cost.

Fig. 15 The accuracy changes with the input image size and the number of features. In the left picture, the size refers to the ratio to the original size. Here, the original size is 150×80, 0.5 and 0.2 means 75×40 and 30×18, respectively



6 Conclusion

In this paper, we propose a new gait recognition scheme. An NMF based feature learning method is used to extract features for every single frame. LSTM is then used to model the high level motion feature in the gait sequence. Experimental results show that this method can achieve high recognition accuracy and has strong adaptability in cross-view and multi-clothes conditions. In other words, this scheme also has strong adaptability to a non-ideal environment.

Since it does not require the cooperation of the subjects and enables longer-distance testing, gait recognition can be applied in more places than other human recognition methods. Our scheme can be easily applied to a very wide range of systems due to its significant advantages in terms of processing speed and low computational cost. In the actual environment, the view angle and personal clothing cannot be as ideal as the data set. Thus, the high adaptability of our scheme to non-ideal situations can further ensure the stability when applying it to general systems. Therefore, if gait recognition is to be applied to pervasive intelligent systems to provide services at right-time, right-where and right-means, our scheme will be a good choice.

The above is our entire scheme, but in addition to the above process, if this scheme is used in practical applications, a single image has images of multiple people. Not only our method, but all methods are facing the same problem, because all methods are using similar data sets. All gait data sets are used to make single-person images by identifying people and then binarizing. Therefore, a possible further improvement in the future is the data collection process, although this does not actually belong to the scope of human recognition, but is another similar topic, object detection.

References

Babaee, M., Rigoll, G.: View-invariant gait representation using joint bayesian regularized non-negative matrix factorization. In: Proceedings of the IEEE international conference on computer vision, pp. 2583–2589 (2017)

Bashir, K., Tao, X., Gong, S.: Gait recognition using gait entropy image pp. 2–2 (2009)

Bashir, K., Xiang, T., Gong, S.: Cross view gait recognition using correlation strength. In: Bmvc. pp. 1–11 (2010)

Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. Lett. **31**(13), 2052–2060 (2010)

Becker, C., Julien, C., Lalanda, P., Zambonelli, F.: Pervasive computing middleware: current trends and emerging challenges. CCF Trans. Pervasive Comput. Interact. **1**(1), 10–23 (2019)

Chunli, L., Kejun, W.: A behavior classification based on enhanced gait energy image. In: 2010 International Conference on Networking and Digital Society, volume 2, IEEE. pp. 589–592 (2010)

Gianaria, E., Grangetto, M.: Robust gait identification using kinect dynamic skeleton data. Multimed. Tools Appl. **78**(10), 13925–13948 (2019)

Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2005)

Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**(2), 201–211 (1973)

Kusakunniran, W., Wu, Q., Li, H., Zhang, J.: Multiple views gait recognition using view transformation model based on optimized gait energy image. In: 2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops, IEEE. pp. 1058–1064 (2009)

Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Support vector regression for multi-view gait recognition based on local motion feature selection. In: 2010 IEEE computer society conference on computer vision and pattern recognition, IEEE. pp. 974–981 (2010)

Lam, T.H.W., Cheung, K.H., Liu, J.N.K.: Gait flow image: a silhouette-based gait representation for human identification. Pattern Recognit. **44**(4), 973–987 (2011)

Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788 (1999)

Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems. pp. 556–562 (2001)

Liu, D.-X., Du, W., Wu, X., Wang, C., Qiao, Y.: Deep rehabilitation gait learning for modeling knee joints of lower-limb exoskeleton. In: 2016 IEEE international conference on robotics and biomimetics (ROBIO). IEEE. pp. 1058–1063. (2016)

Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: European conference on computer vision, Springer, pp. 151–163 (2006)

Muramatsu, D., Makihara, Y., Yagi, Y.: View transformation model incorporating quality measures for cross-view gait recognition. IEEE Trans. Cybern. **46**(7), 1602–1615 (2015)

Wang, L., Tan, T., Ning, H., Weiming, H.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1505–1518 (2003)

Wang, X., Yang, L .T., Liu, H., Deen, M.J.: A big data-as-a-service framework: State-of-the-art and perspectives. IEEE Trans. Big Data **4**(3), 325–340 (2017)

Wang, X., Yang, L.T., Chen, X., Deen, M.J., Jin, J.: Improved multi-order distributed hosvd with its incremental computing for smart city services. In: IEEE transactions on sustainable computing (2018)

Wolf, T., Babaee, M., Rigoll, G.: Multi-view gait recognition using 3d convolutional neural networks. In: 2016 IEEE international conference on image processing (ICIP). IEEE. pp. 4165–4169. (2016)

Yang, L.T., Wang, X., Chen, X., Wang, L., Ranjan, R., Chen, X., Deen, M.J.: A multi-order distributed hosvd with its incremental computing for big services in cyber-physical-social systems. In: IEEE transactions on big data (2018)

Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th International Conference on Pattern Recognition (ICPR'06), volume 4, IEEE. pp. 441–444 (2006)

Zhao, X., Jiang, Y., Stathaki, T., Zhang, H.: Gait recognition method for arbitrary straight walking paths using appearance conversion machine. Neurocomputing **173**, 530–540 (2016)

Zifeng, W., Huang, Y., Wang, L.: Learning representative deep features for image set analysis. IEEE Trans. Multimed. **17**(11), 1960–1968 (2015)

Zifeng, W., Huang, Y., Wang, L., Wang, X., Tan, T.: A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 209–226 (2016)

Enyuan Cao received his B.S. degree in Electronic Engineering from University of Electronic Science and Technology of China and University of Glasgow. His current research interests include pattern recognition, machine learning and computer graphics.

Kun Cao is currently pursuing the Ph.D. degree in Computer Science with the School of Computer Science and Technology, East China Normal University, Shanghai, China. His research interests are in the areas of Internet of things, real-time embedded systems, and cyber physical systems.

Kuining Feng received the B.S. degree in Electronic and Information Engineering from University of Electronic Science and Technology of China and University of Glasgow. He is currently a graduate students at University of Electronic Science and Technology of China. His research interests are Parallel and Distributed Computing.

Jiawei Wang undergraduate student of the Department of Mathematics, Statistics and Computer Science, St. Francis Xavier University.