# Memory-based Gait Recognition

Dan Liu
liudan11060126@gmail.com
Mao Ye*
cvlab.uestc@gmail.com
Xudong Li
lixudong268@gmail.com
Feng Zhang
xizero00@gmail.com
Lan Lin
linlan0921@hotmail.com

School of Computer Science and Engineering,
Center for Robotics,
Key Laboratory for NeuroInformation of Ministry of Education,
University of Electronic Science and Technology of China,
Chengdu 611731, P.R. China

**Abstract**

Gait recognition is an interesting and challenging task aiming to classify the subjects based on the way they walk, which is subject to various covariates including carrying, clothing, surface and view angle. In this paper, we propose to utilize the memory mechanism to effectively alleviate the aforementioned problems. Specifically, we extract the 2D location information of human joints as the gait features via the migratory articulated human detection. Inspired by the mechanism of brain sequence processing, we input the gait feature sequence into the memory-based gait recognition (MGR) network, which achieves the process of memory and identification of the gait sequence. Our proposed MGR is robust to the noise that maybe exist in the gait sequence features. Besides, MGR is able to learn on the data with long range temporal dependencies. The experimental results on the CASIA A and CASIA B gait datasets verify the feasibility and effectiveness of the proposed method.

# 1 Introduction

Human recognition and classification has been one of the most active research topics in computer vision. In the past decade, it facilitates a wide range of applications such as video surveillance, access control and criminal investigation. Biometrics is intended to address such a need by taking good advantage of the physiological or behavioral characteristics of people. The traditional biometric features used currently include face, iris, signature, voice, fingerprint, etc. However, the difference between these biometric features and gait feature is that the latter has many unique characteristics such as non-invasive, capture at distance and non-perceivable [7], and it is also less likely to be obscured than other biometric features. In recent years, Gait analysis has attracted a lot of attention from many international and domestic research institutes and scholars. Gait analysis mainly consists of two fields: gait recognition and gait classification. The former is to identify object's ID in the specific space


* corresponding author.
Pages 82.1-82.12
DOI: https://dx.doi.org/10.5244/C.30.82

while the latter includes gender [12], age and action recognitions [22]. Researching on gait recognition has a long history, and many gait recognition algorithms have been developed. The existing methods for human gait recognition can be divided broadly into two categories: model-based and appearance-based.

Model-based methods [3, 10, 24, 26, 27] attempt to explicitly model the human body or motion by employing the static and dynamic body parameters [19]. For example, the stride length and speed are the dynamic parameters while the size ratios of various body parts belong to the static one. Generally the model-based methods are to execute model matching in each frame of a walking sequence. Because it is necessary to model and track the subjects body or motion, these approaches are usually computationally intensive.

Compared to the model-based approaches, the appearance-based methods [1, 5, 11, 18, 21] usually extract some different types of features such as the whole silhouettes, silhouette width vectors and Fourier descriptors. They require much less computation cost. Moreover, the dynamic information can usually improve recognition performance than static counterpart. Therefore, many researchers prefer to introduce new features to obtain good performance. However, these methods are usually subject to viewpoint change and scale in performance.

## 1.1 Motivations

In the research of gait recognition, there still exists many challenges. Various covariates influence the performance of gait recognition including the presence of shadows, clothing variations, carrying conditions, mutative views, etc. It is very significant to find the features that are relatively robust to the various covariates. In this paper, in the first stage, we employ migratory articulated human detection method [23] to extract the positions of the joints as the gait features. Further, many methods extract the features frame by frame until the end of a complete gait sequence, and statically joint these features as the whole gait characteristics. But our brain, in contrast, processes a gait video frame dynamically based on the previous frames, in which all frames in a gait sequence are linear and climactic relation rather than paratactic and equal.

Inspired by the mechanism of memory and prediction in our brains [9], we propose a memory-based gait recognition method (MGR) for human gait recognition. Compared to the traditional neural network, the memory neuron network (MNN) simulates the human brain and stores the objects in the weights of neural connections. Besides, by the large-scale parallel computing, MNN can repair the incomplete and tainted data (the extracted gait feature is dirty). In this paper, we take advantage of the MGR to realize the memory and recognition process of the gait sequences. It maybe provide a fresh orientation for solving gait recognition problem.

## 1.2 Main Contributions

The main contributions of the paper are summarized as follows:

- Based on the articulated human detection algorithm, we skillfully utilize the labeled Image Parse dataset and training strategy to acquire the 2D positions of joints as the gait features.

- A novel and straightforward memory-based gait recognition method (MGR) is presented, which accords with human's sequence cognition process. It is the first time

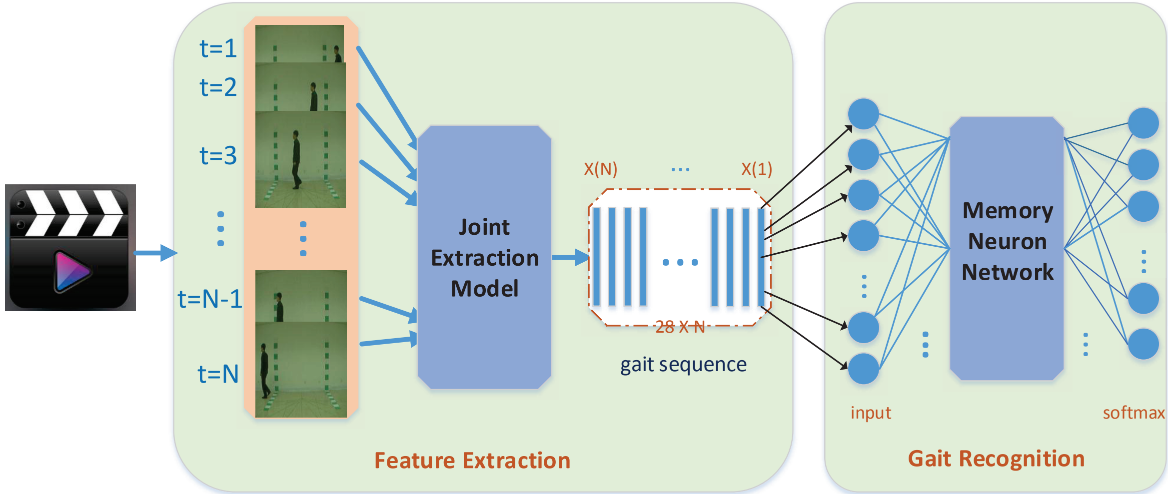



Figure 1: The Memory-based gait recognition framework. Obviously, the process is divided into two stages: feature extraction and gait recognition. $N$ denotes the length of a gait sequence.

that we utilize the memory neuron network to address the gait recognition issue.

The remainder of this paper is organized as follows: Sec. 2 describes the overview of the proposed approach. Sec. 3 introduces the algorithm process in detail. Then we shows and analyzes the experimental results in Sec. 4. In the end, Sec. 5 concludes the paper.

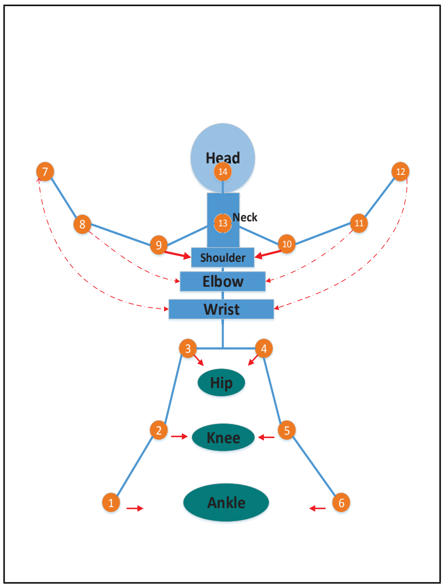
# 2 Method Overview

The approach presented in this paper chooses the position information of joints as gait features to distinguish different subjects. Our feature extraction is accomplished based on the articulated human detection and human pose estimation algorithm [23]. Fig.1 shows the overall framework of the method. $N$ denotes the length of a gait sequence. The whole process is divided into two phases: feature extraction and gait recognition. In the feature extraction process, we put the gait image sequences into the migratory articulated human detecting algorithm and obtain the human articulations. Since there are 14 joints involved in the research, the length of the extracted 2D joint positions in an image is 28. So far, a gait picture sequence has been transformed to a gait feature sequence. According to our MGR algorithm, in the gait recognition stage, at every moment one frame data of the gait feature sequence is input. At the end of the sequence, the MGR outputs a probability vector, and the index of the maximal number in the vector shows the subject's ID.

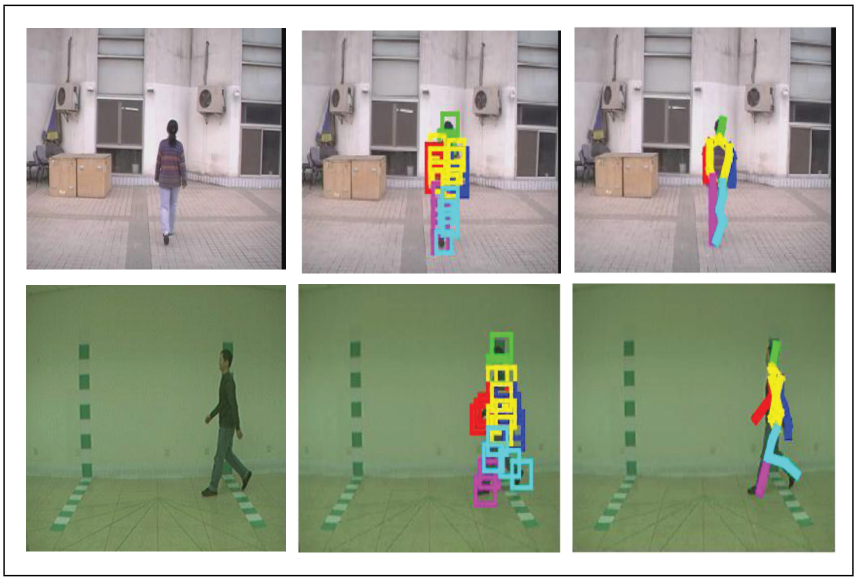
# 3 MGR Method

## 3.1 Joint Extraction

Human detection and tracking is the first step to gait analysis. Compared to the latter gait recognition, it is not a main part of our work. Therefore, we give a concise introduction to human detection and feature extraction. In this paper, we take advantage of the articulated human detection and pose estimation approach by Yang and Ramanan [23] which conflates the human detection and pose estimation issues. Because of the lack of labeled joint position




(a)



(b)

Figure 2: (a) A visualization of 14 body key points; (b) Two joint extraction samples. The left column is the input image. The middle and right column are part bounding boxes along with skeletons computed from the bounding boxes.

data and part type labels in our experimental data, we first train a model on the Image Parse dataset [16] (the labels are available on this dataset). In order to adapt the fully trained model to the experimental gait datasets, we manually produce labels for a small number of positive samples, and generate a mass of samples' background pictures as negative samples. Then we take advantage of these positive and negative data obtained from the experimental gait datasets so as to adjust the detection and pose estimation model. Compared with the model that is trained just by one time (the Parse dataset and our new added data are regarded as a whole dataset to train the model), the former has better performance for joint extraction on our experimental datasets. We show some skeleton extraction results in Fig.2(b). The two samples come from different datasets, and the last column is corresponding skeleton image results.

As shown in Fig.2(a), for a subject, there are 14 critical points extracted by the pose estimation model, and they are left ankle, right ankle, left knee, right knee, left hip, right hip, left wrist, right wrist, left elbow, right elbow, left shoulder, right shoulder, neck and head respectively. The numbers on the key points depict their order in the final feature vector. Since the joint feature does not include the depth and other information, it is a 28 dimensional vector($14 \times 2$). Obviously, the dimension is far lower than other gait features.

## 3.2 Memory Neuron Network Architecture

The most important part of MGR is the memory neuron network (MNN). In this paper, we choose to utilize the long short-term memory (LSTM). In general, recurrent neuron network (RNN) can use their feedback connections to store representations of recent input events in the form of activations. However, when minimal time lag between inputs and corresponding object outputs is long, it does not work well at all. Hochreiter and Schmidhuber [6] presented a novel recurrent network architecture LSTM, which has the ability to successfully learn on data with long range temporal dependencies. According to the brain processing mechanism, there exists the considerable time lag between the gait inputs and their corresponding outputs. In addition, the LSTM can repair the damaged data to some extent. Therefore, it is a good

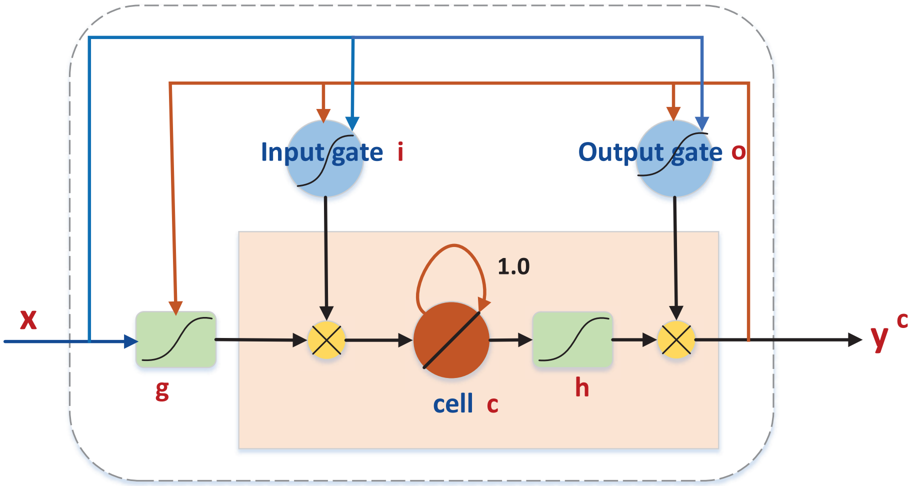



Figure 3: The structure of a LSTM cell neuron.

choice to make use of the LSTM to cope with the gait recognition issue in this paper.

By the above mentioned joint extraction phase, the gait image sequence has been converted to the gait feature sequence. According to the gait recognition process in human brain, at each time step, the input to the network is a vector denoting the 2D positions of the skeleton joints in a frame, and the recognition result is output at the end of a gait sequence. In other words, during the inference of LSTM, all hidden neurons need be activated at any time, while the output neurons are only activated at the end of a sequence. Since we use softmax activation function to produce the predictions, the number of subjects in the datasets equals to the number of output units.

In the LSTM, the memory cell is complicated. Each memory cell is built around a central linear unit with a fixed self-connection of weight 1. It is also called the Constant Error Carousel (CEC), which ensures the gradient can pass across many time steps without vanishing or exploding. As shown in Fig.3, it is the traditional LSTM cell neuron structure, which contains an input gate $i$, a cell $c$, an output gate $o$ and a cell output response $y^c$. The input gate governs the information flowing into the cell. The output gate controls how much information from the cell is passed to the next.

Without loss of generality, for the LSTM neurons at time $t$ in LSTM layer, as input $x(t)$ is a vector including 28 elements (there are 28 input units in MNN), and the recursive computation of activations of the units is:

$$\begin{aligned} i(t) &= \mathbf{f}(W_{xi}x(t) + W_{hi}h(t-1) + W_{ci}c(t-1) + b_i) \\ o(t) &= \mathbf{f}(W_{xo}x(t) + W_{ho}h(t-1) + W_{co}c(t) + b_o) \\ c_{in}(t) &= \mathbf{g}(W_{xc}x(t) + W_{hc}h(t-1) + b_c) \\ y^c(t) &= o(t) \odot \mathbf{h}(c(t)) \end{aligned} \tag{1}$$

where $c_{in}$ is the input of memory cells. $\odot$ denotes element-wise product. $W_{pq}$ is the weight matrix between $p$ and $q$. For example, $W_{xi}$ is the weight matrix from the input $x$ to the input gate $i$. $h$ is the hidden unit activation and $b_\alpha$ denotes the bias term, $\alpha \in \{i, c, o\}$. For the symbol $\mathbf{f}(\cdot), \mathbf{g}(\cdot), \mathbf{h}(\cdot)$, they are different logistic sigmoid functions used in this paper:

$$\mathbf{f}(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

$$\mathbf{g}(x) = \frac{4}{1+e^{-x}} - 2 \tag{3}$$

$$\mathbf{h}(x) = \frac{2}{1+e^{-x}} - 1 \tag{4}$$

As the main part, the internal states of memory cells $c$ are calculated as follows:

$$\begin{aligned} &c(t) = 0; \quad t = 0 \\ &c(t) = c(t-1) + i(t) \odot c_{in}(t) \quad t > 0 \end{aligned} \tag{5}$$

For the cell output $y^c$, it will be regarded as the input of the softmax layer at the time $t = N$. Since gait recognition is a multi-class problem, we can get the output probability vector $y$ from the last layer in LSTM:

$$\begin{aligned} y(t) &= (p(c_1|X), p(c_2|X), \cdots p(c_k|X))^{\mathrm{T}}; \quad k = 1, 2, \cdots C \\ y(t) &= \frac{e^{W_{cy}y^c(t)+W_{hy}h(t)+b_y}}{\mathbf{sum}(e^{W_{cy}y^c(t)+W_{hy}h(t)+b_y})} \end{aligned} \tag{6}$$

where $X$ is a gait sequence. $P(c_k|X)$ is the probability of the gait sequence belonging to $c_k$ class. $C$ denotes the number of classes. $W_{cy}$ and $b_y$ symbols have the same meaning as that in Eq.1, and $\mathbf{sum}(\cdot)$ denotes the summation of all elements in a vector. The index of the maximum number in the probability vector represents the subject's ID.

### 3.3 Decision and Training Rules

The proposed MGR method is described in Fig.1. In the joint extraction process, the gait image sequences have been transformed to the gait feature sequences. Since 14 joints are extracted in a gait image, the input layer consists of 28 units. Note that, without any preprocessing, our MGR only depends on the raw and noisy gait features. In order to eliminate the differences between different dimensional data and speed up the convergence of the network, the gait features are linearly normalized between 0 and 1.

In the MGR, the LSTM hidden layers are fully connected to the previous layer, and they are full recurrent connections between the hidden units. For the output layer, the size equals to the number of subjects. The softmax activation function is used in the output layer to give a response at the end of a gait sequence. The output is a probability vector, in which each value indicates the probability of the input gait sequence belonging to a certain category.

In the training process of LSTM, the back propagation (BP) is used for the output units. The back propagation through time (BPTT) is employed for the output gates. For the input gates and the memory cells, the real time recurrent learning (RTRL) is applied. In this paper, the learning rate is 5e−2. Given the previous gait recognition algorithm's analysis, the update complexity of every weight in the LSTM is essentially that of BPTT, namely, $O(1)$. It is highly competitive in comparison to other approaches [6].

## 4 Experiment

### 4.1 Dataset

In this paper, the proposed approach is tested on the CASIA gait datasets (Dataset A and B) [20]. The CASIA A dataset includes 20 subjects in an outdoor environment. Each subject

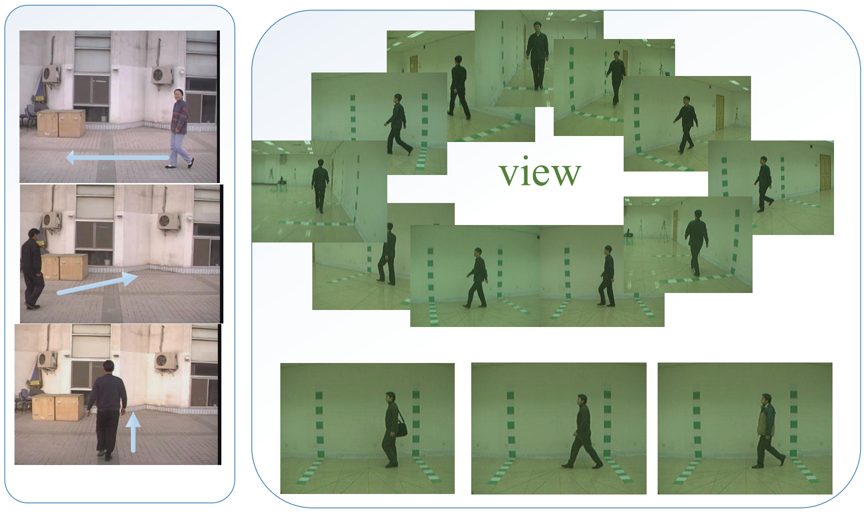



Figure 4: Examples of the CASIA A and CASIA B datasets. The left part shows three subjects with different views on the CASIA A dataset. Some samples are exhibited in the right side on the CASIA B dataset, in which 11 different views are showed in the top right part, and the bottom right indicates three different covariate condition changes.

walks along a straight line path at free cadences in three different views with respect to the image plane, and the three views respectively are laterally (0°), obliquely (45°), and frontally (90°). For each viewing angle every object has four gait sequences. Therefore the CASIA A dataset includes a total of 240 gait sequences ($20 \times 4 \times 3$). The length of each gait sequence is about 90 frames in average, and there are approximately 20000 pictures in the dataset.

The CASIA B dataset consists of the data from 124 subjects in an indoor environment with a simple background, including 93 males and 31 females among all subjects. The data are captured under three different covariate condition changes at free cadences: carrying, clothing, and view angle. Note that the view changes are much bigger in CASIA B dataset than in CASIA A dataset. For each subject, there are 11 different view angles from frontal view(0°) to back view (180°). And for each view, one person has 10 gait sequences: six normal sequences (the person does not carry a bag or wears a coat), two carrying-bag sequences and two wearing-coat sequences. There are a total of 13640 video sequences in the database, with 2-3 gait cycles in every sequence. The frame size is 320-by-240 pixels, and the frame rate is 25 fps. Fig.4 shows some data samples of CASIA A and CASIA B datasets.

## 4.2 Experiments on CASIA A

Since there are 20 subjects in the CASIA A dataset, we construct the MGR including 28 input units and 20 output units. For each view, there are only 80 gait sequences. Here, we choose the leave-one-out cross-validation rule to estimate the algorithm's performance. In other words, each time we use one gait sequence as a test sample and the remainder as training set. In order to augment the training data, we randomly produce 10 numbers in the range of the length of a gait sequence, and remove the 10 corresponding gait frame vectors. The process is repeated five times to achieve the goal of data augmentation.

We complete three experiments with three views ($0^\circ, 45^\circ, 90^\circ$) on CASIA A dataset. For the MGR, by massive experiments, we design one hidden layer containing 15 memory blocks with 8 memory cells in each block in the 0° view experiment, while arranging 8 memory blocks with 15 memory cells in every block in the 45° view experiment and setting

12 blocks with 8 cells in a block in the 90° view experiment. In addition, we find the additional bias has side effects for the performance of gait recognition.

In order to make a direct comparison, the motion-based gait recognition in monocular sequences [2], spatiotemporal correlation of silhouette images algorithm [14] and Wang's method [20] are compared with our method (the experimental setup is same). Tab.1 shows the recognition rates of these algorithms in 0° view. Our method achieves a recognition rate of 82.50% in 0° view, and exceeds 10% when compared to the worse result. Moreover, we get some enlightenment from [8]: in the LSTM's applications, the longer sequences do not improve the algorithm performance in some cases. Therefore, we reduce the length of sequences to half of the original and the length of each gait sequence is about 45 in average.

As can be seen from the Tab.2, the results of before length reduction and after length reduction are compared with Wang's three results [20]. The gait recognition rate is further improved. Before length reduction, our method has achieved 86.25%'s accuracy rate in average. By the length reduction process, in the 0° view, the correct identification rate is increased from 82.50% to 85.00%. The 45° view and 90° view gait recognition rates are improved to 87.50% and 95.00% respectively. The mean gait classification rate reaches 89.17%. However, it is worth noting that some methods [25] have achieved the recognition rate of 97% nearly on the CASIA A dataset. Since they fuse many feature vectors, its computational cost is more expensive than ours. In contrast, our suggested algorithm is time-saving and easy to understand.

| Methods | BenAbdelkader [2] | Phillips [14] | Wang [20] | Ours |
|---|---|---|---|---|
| Rec R | 72.50 | 78.75 | 75.00 | **82.50** |

Table 1: Algorithms comparisona on the CASIA A dataset(0°-view).

| Methods | 0°-view | 45°-view | 90°-view | avg |
|---|---|---|---|---|
| Wang1 [20] | 65.00 | 63.75 | 77.50 | 68.75 |
| Wang2 [20] | 65.00 | 66.25 | 85.00 | 72.08 |
| Wang3 [20] | 75.00 | 81.25 | 93.75 | 83.33 |
| Original results | 82.50 | 83.75 | 92.50 | 86.25 |
| Length reduction | **85.00** | **87.50** | **95.00** | **89.17** |

Table 2: The comparisons of some algorithms on the CASIA A (0°,45°,90°) dataset. Wang1, Wang2 and Wang3 indicate that the different classifiers and similarity measures are used in the same method.

## 4.3 Experiments on CASIA B

In the CASIA B dataset, there exists 124 subjects. Therefore, 124 neuron units form the output layer. In order to simplify the experiments, we only consider the data in 90° view. Because there are different covariate condition changes (carrying, clothing, and view angle) are included on the CASIA B dataset, we conduct three experiments to evaluate our approach under condition changes. There are 10 sequences available for each subject in 90° view.

The data selection rule and data information of three experiments are summarized as follows:

- **Exp1** Exp1 focuses only on carrying conditions. Two normal sequences out of six are randomly selected along with the two carrying sequences. There are a total of 496 ($124 \times 4$) sequences.

- **Exp2** Exp2 is different from Exp1. Exp2 pays attention to clothing changes. Two out of six normal sequences and the two coating sequences are chosen. There also exists a total of 496 ($124 \times 4$) sequences.

- **Exp3** Exp3 explores both covariate conditions together. Two out of six normal sequences, the two carrying and two clothing sequences form the dataset, in which 744 ($124 \times 6$) gait samples are included.

For every experiment's dataset, it is randomly and equally split into two subsets, one for training and the other for testing. Then, the test set is further divided into two sets. The whole experimental settings are same with [13]. In this paper, we use an approximate 2-fold cross-validation rule because of the testing on 1/4 of the data rather than 1/2. In addition, we also augment the training data by quadrupling the original data, and make comparisons between before the sequences' length reduction and after length reduction.

As shown in Tab.3, the recognition rate in Exp3 is relatively low. Because there exist three condition changes in Exp3, it is comprehensible to get the result. The best gait recognition rate obtained by our proposed method exceeds more than 16.00% when compared to Martin [13]. Before length reduction our algorithm has obtained better performance than Martin's, which improves the correct rate from 67.65% to 82.88%. The recognition rate of the suggested method is further improved by length reduction. The mean correct classification rate reaches 83.69%.

| Methods | Exp1 | Exp2 | Exp3 | Avg |
|---|---|---|---|---|
| Martin [13] | 70.16 | 74.19 | 58.60 | 67.65 |
| Our1 | 83.06 | 85.48 | 80.11 | 82.88 |
| Our2 | **83.87** | **85.48** | **81.72** | **83.69** |

Table 3: Algorithms comparisons on the CASIA B dataset on Exp1, Exp2 and Exp3.

Compared to the state-of-the-art algorithms, ours does not achieve the best gait recognition rate. However, remarkably the extracted 2D joint positions are time-saving and force-saving but noisy. The proposed approach provides a new baseline method for the gait recognition issue. Though many algorithms [4, 15, 17] also evaluate their performance on the CASIA B dataset, the experimental settings are different and we cannot directly make comparisons.

For Exp1, Exp2, and Exp3, we only perform them in $90^\circ$ view. We find the fact during the experiments: the extracted 2D articulation position features can completely distinguish the views of gait sequences (**100%** in accuracy). Therefore, when we have constructed models in every view, it is not necessary for us to structure a new model in cross-view cases. Moreover, the average gait recognition rate is constant with cross-view cases. Fig.5 displays the general process.

# 5 Conclusions

With the increasing demands of visual surveillance systems, human identification at a distance has received a lot of attention. As a biometric for recognition, gait has been proved to

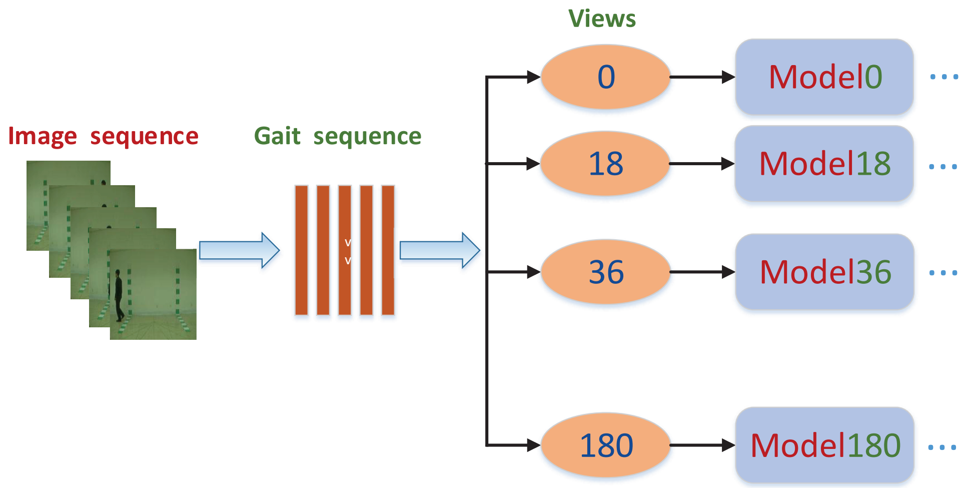



Figure 5: The processing procedure of the gait recognition issue based on multi-views.

be a potential behavioral feature. In this paper, we present a straightforward and effective memory-based method for automatic gait recognition. We train the extraction model on the basis of the labeled Parse dataset. It takes less time and energy for us to adjust the joint extraction model and get the gait joint positions on experimental datasets. Though the extracted 2D joint positions are seriously noisy, we utilize the MNN to alleviate the problem, which can repair the damaged data based on the attractors. It is the first time that we use the memory mechanism to solve the gait recognition issue. The network configuration is simple but the proposed method still achieves the relatively satisfactory and comparable results.

It is worth noting that the input features of MNN are human body joint tracking output, which may not be available under a more realistic setting to some extent. It must use the joint extraction model to get the necessary features and this is also a limitation of this paper. Thankfully, the process of joint extraction is not complicated.

# Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61375038) and Applied Basic Research Programs of Sichuan Science and Technology Department (2016JY0088).

# References

[1] Ricardo M. Araujo, Gustavo Graña, and Virginia Andersson. Towards skeleton biometric identification using the microsoft kinect sensor. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 21–26. ACM, 2013.

[2] Chiraz BenAbdelkader, Ross Cutler, Harsh Nanda, and Larry Davis. *Audio- and Video-Based Biometric Person Authentication: Third International Conference, AVBPA 2001 Halmstad, Sweden, June 6–8, 2001 Proceedings*, chapter EigenGait: Motion-Based Recognition of People Using Image Self-Similarity, pages 284–294. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[3] Nikolaos V. Boulgouris and Zhiwei X. Chi. Human gait recognition based on matching of body components. *Pattern Recognition*, 40(6):1763 – 1770, 2007.

[4] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11):977 – 984, 2009.

[5] Ju Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, Feb 2006.

[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9 (8), Nov 1997.

[7] Guochang Huang and Yunhong Wang. *Computer Vision – ACCV 2007: 8th Asian Conference on Computer Vision, Tokyo, Japan, November 18-22, 2007, Proceedings, Part I*, chapter Gender Classification Based on Fusion of Multi-view Gait Sequences, pages 462–471. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[8] Armand Joulin and Tomas Mikolov. Inferring algorithmic patterns with stack-augmented recurrent nets. In *Advances in Neural Information Processing Systems 28*, pages 190–198. Curran Associates, Inc., 2015.

[9] Christof Koch and Joel L. Davis. *Large-scale Neuronal Theories of The Brain*. MIT press, 1994.

[10] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):145–155, March 2008.

[11] J. Liu and N. Zheng. Gait history image: A novel temporal template for gait recognition. In *2007 IEEE International Conference on Multimedia and Expo*, pages 663–666, July 2007.

[12] J. Lu, G. Wang, and P. Moulin. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE Transactions on Information Forensics and Security*, 9(1):51–61, Jan 2014.

[13] Raul Martin-Felez and Tao Xiang. *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I*, chapter Gait Recognition by Ranking, pages 328–341. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[14] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. Baseline results for the challenge problem of humanid using gait analysis. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 130–135, May 2002.

[15] Yan qiu Liu and Xu Wang. Human gait recognition for multiple views. *Procedia Engineering*, 15:1832 – 1836, 2011.

[16] Deva Ramanan. Learning to parse images of articulated bodies. *Advances in neural information processing systems*, pages 1129–1136, 2006.

[17] Imad Rida, Ahmed Bouridane, Gian Luca Marcialis, and Pierluigi Tuveri. *Image Analysis and Processing — ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part II*, chapter Improved Human Gait Recognition, pages 119–129. Springer International Publishing, Cham, 2015.

[18] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, Feb 2005.

[19] Faezeh Tafazzoli, George Bebis, Sushil Louis, and Muhammad Hussain. Genetic feature selection for gait recognition. *Journal of Electronic Imaging*, 24(1):013036, 2015.

[20] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, Dec 2003.

[21] J. Wu. Automated recognition of human gait pattern using manifold learning algorithm. In *Natural Computation (ICNC), 2012 Eighth International Conference on*, pages 199–202, May 2012.

[22] Bruce Xiaohan Nie, Caiming Xiong, and Song-Chun Zhu. Joint action recognition and pose estimation from video. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[23] Y. Yang and D. Ramanan. Articulated human detection with flexible mixtures of parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2878–2890, Dec 2013.

[24] J. H. Yoo, D. Hwang, K. Y. Moon, and M. S. Nixon. Automated human recognition by gait using neural network. In *2008 First Workshops on Image Processing Theory, Tools and Applications*, pages 1–6, Nov 2008.

[25] Chuanlei Zhang, Shanwen Zhang, Jucheng Yang, and Weijun Cheng. *Biometric Recognition: 10th Chinese Conference, CCBR 2015, Tianjin, China, November 13-15, 2015, Proceedings*, chapter Gait Recognition Based on Energy Accumulation Images, pages 456–463. Springer International Publishing, Cham, 2015.

[26] X. Zhang and G. Fan. Dual gait generative models for human motion estimation from a single camera. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4):1034–1049, Aug 2010.

[27] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan. Robust view transformation model for gait recognition. In *2011 18th IEEE International Conference on Image Processing*, pages 2073–2076, Sept 2011.