

Chapter 6

Cross-View Gait Recognition Using Deep Learning Approach



Jyoti Bharti and Lalit Lohiya

Abstract Walking speed is used principally to assess the status of human health. Recent gait recognition systems often experience difficulties, including variations in the viewing angle and enormous variations in the intraclass. For decades, computer vision-based approaches are in great demand and more effective in clinical gait analysis. Used referenced dataset is under well-controlled conditions. Gait movement rate is the predominant human biomechanical determinant. This paper proposes a scientific way of determining the pattern of gait movement at a particular speed of healthy people. Based on deep learning technology, our proposed model comprises of a fully convolutional neural network accompanied by batch normalization and max pooling. Proposed models on large-scale gait datasets were being extensively tested and have been compared with the benchmark approaches. The proposed model's overall average accuracy is about 93.0%. Where various clothing and transport conditions have been encountered, the method is also robust for such conditions.

6.1 Introduction

Gait is one of the most popular biometric aspects of humans because it can be bona fide excluding subject cooperation at a distance from a lens. In the gait recognition model, the speed difference between matching pairs plays an important role and the gait mode of walking or running makes it more challenging. The distinctive features of the human body, such as handwriting, voice, and gait, have been highly studied for many years to make excellent progress in biometrics [1]. Every individual has a particular way of walking that happens because of facilitated and worked together activities of the skeletal muscles and sensory system. This makes the biometric step a ground-breaking marker to decide obsessive behavior caused by actual injury, maturation, or related issues. These inside and outside variables legitimately influence the movement and activity of the body and result in walk impedance [1]. Biomechanical examples of human movement are by and large speed-dependent, the adequacy of

J. Bharti (✉) · L. Lohiya
Maulana Azad National Institute of Technology, Bhopal, India
e-mail: jyotibharti@manit.ac.in

explicit development normally levels with the development speed (stride velocity is a determining factor in the nature of phases) [1, 2].

In an average walk examination, patients perform stride preliminaries at their agreeable speed and their step designs are regularly contrasted, and a reference design is from a regulating information base. Nonetheless, the impact of stride speed is commonly not represented when the step example of obsessive people is contrasted and sound ones who do not walk at an equal speed [2].

A repeated pattern consisting of steps and strides is the gait loop. A step phase began with one foot's initial contact and finishes with the other foot's initial contact. The stride phase begins with one foot's initial contact and finishes with the same foot's next initial contact, which comprises of two steps. There are two main phases in the gait cycle: stance phase comprises of 60% of the entire gait cycle and swing phase comprises of 40 percent of the full gait cycle. The step is the time when the foot is in touch with the ground and the weight is borne by the limb. The swing process is the time during which the reference foot is not in touch with the surface and swings in the air, indicating the weight of the body is borne by the contralateral limb [1, 3–5].

6.2 Related Work

6.2.1 Gait Recognition System

Human authentication based on gait has significant importance and a wide scope of research. The science of recognizing or identifying people by their physical characteristics is known as biometrics. One form of biometric technology is gait recognition [5]. The authors are working on gait recognition biometric technology to analyze the movements of body parts such as the foot, the knee, the shoulder, and so on. There are various gait acquisition technologies such as cameras, wearable sensors, and non-wearable sensors. Figure 6.1 demonstrates the gait recognition system.

Any recognition system is developed based on the training phase and the testing phase. There are two approaches, such as authentication and recognition. Automatic video processing was the basis for the first gait recognition [3, 6, 7]. It generates a mathematical model of motion. This method is the most common and requires the study of video samples of walking subjects, joint trajectories, and angles. In order to perform detection, the second approach uses a radar system to monitor the gait period that is then compared with the other samples [7].



Fig. 6.1 Gait recognition system

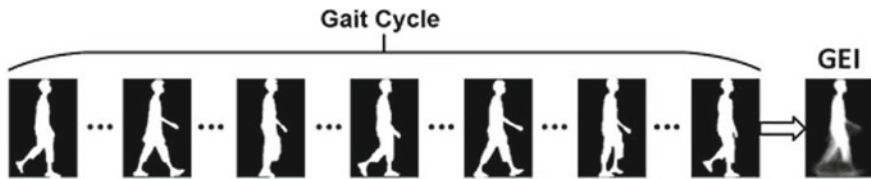


Fig. 6.2 Gait energy image extraction

A potential solution for this issue would be to gather a few walking preliminaries at different walking rates to construct a reference information base for essentially any conceivable step speed [8–11]. The tedious idea of such an assortment of knowledge, however, would be cost-restrictive and unviable. Scientists have suggested regression techniques as a reachable alternative for anticipating phase limits based on exploratory facts to resolve this barrier. The expectation data depends purely on the normal, slow, and fast walking speeds for stable subjects but only 10% time frame phase period when the complete step cycle was considered. [1, 2, 7, 12, 13]

The gait energy image (GEI) is an enhanced spatiotemporal feature to analyze a subject's gait and posture characteristics. Binary silhouette sequence normalization was used instead of direct silhouette sequence to provide robustness. GEI can be stated as,

$$\text{GEI} = \frac{1}{S_f} \sum_{s=1}^{S_f} I_o(u, v, s) \quad (6.1)$$

where S_f states the total number of silhouette images in the gait cycle and s is an image number at an instant. $I_o(u, v)$ denotes the silhouette frame at an instant. Figure 6.2 shows the extracted GEI image from full gait cycle.

6.3 Proposed Work

6.3.1 Implementation Details of Proposed Model and Model Summary

CASIA B Gait Dataset contains silhouette sequence for each of the view angle. This silhouette sequence of images has been converted to gait energy image. So, there are in total 13,640 images. 11 angle view * 10 categories * 124 subjects = 13,640 GEIs. To evaluate the proposed method, all the sequences have been converted to a gait energy image (GEI), then set of GEIs are fed to gait recognition proposed model. Figure 6.3 shows the various GEIs under three conditions and 11 view angles.

One layer consists of a 2D convolutional layer followed by batch normalization and max pooling in the proposed deep learning model for cross-view gait recognition.



Fig. 6.3 GEI examples from an individual’s CASIA B dataset under various conditions from view 0° to 180° with an 18° interval. Top row: regular walking, middle row: coat-wearing, and bottom row: bag-wearing



Fig. 6.4 Proposed model

As such, three stacks of layers are available in total. Figure 6.4 shows the design of the model proposed. Table 6.1 shows description of the model proposed.

6.3.1.1 Convolutional 2D Layer

A convolutional layer is a part of the deep neural network that can intake an input image, attribute importance to the different aspects/objects in the frame by learning weights and biases, and distinguish one from the other. The preprocessing needed for the convolutional neural network is much lower than for other classification algorithms. The size of input image is (240×240) , and size of convolutional kernel is (7×7) . The input image is downsampled, and output of convolutional 2D layer is about (234×234) . Output of convolution layer is nothing but the feature map.

Table 6.1 Summary of proposed model (BN—batch normalization, MP—max pooling, and Conv—convolution layer)

Layers	Output size	No. of filters	Filter size	Padding	Params
Conv 1	$234 \times 234 \times 16$	16	7×7	1	2368
BN	$234 \times 234 \times 16$	–	–	–	64
MP 1	$117 \times 117 \times 16$	16	2×2	2	0
Conv 2	$111 \times 111 \times 16$	16	7×7	1	12,560
BN	$111 \times 111 \times 16$	–	–	–	64
MP 2	$55 \times 55 \times 16$	16	2×2	2	0
Conv 3	$49 \times 49 \times 16$	16	7×7	1	12,560
BN	$49 \times 49 \times 16$	–	–	–	64
MP 3	$24 \times 24 \times 16$	16	2×2	2	0
Conv 4	$18 \times 18 \times 64$	64	7×7	1	50,240

6.3.1.2 Batch Normalization

Batch normalization helps each network layer to learn much more independently of other layers. Higher learning values can be used because batch normalization assures that no activation has actually went high or really low. It reduces overfitting due to a slight regulation effect. To stabilize a neural network, it normalizes the performance of a previous activation unit by deducting batch average and dividing by batch confidence interval.

6.3.1.3 Pooling 2D

An issue with a feature map output is that they are sensitive to input feature position. To fix this sensitivity, sample the function maps is one approach. Pooling layers offer an approach to sample feature maps by summarizing features in the feature map kernel. Pooling is needed to detect feature in feature maps. We used max pooling. Results are pooled function maps highlighting the kernel's most present feature. The kernel size is (2×2) .

6.3.1.4 Dropout

To simulate having a large number of distinct network architectures, a single model can also be used by arbitrarily lowering out nodes during training. This is known as dropout and provides a very inexpensive and effective regularization impact for reducing overfitting and enhancing generalization error in deep neural networks.

There are two sets of GEIs known as gallery set and probe set. While training the model gallery set is used and for evaluation probe set is used. As there are 11 views, there will be 11 models trained and tested.

6.4 Experiments and Result Analysis

6.4.1 CASIA-B Gait Dataset

This is a large multiview gait database; the information was gathered from 11 views and consists of 124 subjects. There are three varieties of variations clothing and carrying conditions including wide range of view angles and also consist of extracted silhouettes.

With 124 subjects, including both genders, 93 males and 31 females in an indoor environment, the data was gathered at 25 fps and a size of frame is 320×240 . Therefore, there are 13,640 sequences in total. Figure 6.5 shows how data was collected from different angles [1–6, 14]. CASIA-B dataset contains 124 subjects; data is captured from 11 angle views from 0° to 180° with 18° interval. There are in total

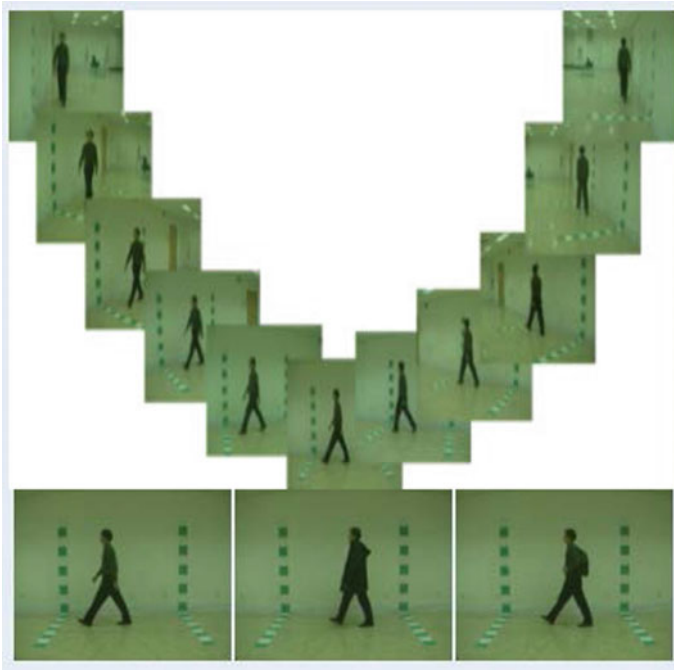


Fig. 6.5 11 angle view of CASIA-B dataset

10 categories such as normal walking with 6 subsets, walking while carrying bags with 2 subsets, and walking while wearing coat with 2 subsets.

6.4.2 Experimental Environment

To perform our experiment, we have used Python 3.7, Jupyter Notebook, and Anaconda Environment. TensorFlow 2.0 is used. The dataset used is CASIA-B dataset.

6.4.3 Result Analysis on CASIA-B Dataset

Proposed model is trained at 11 angle views, and each trained model is evaluated against different viewing angles probe set. Table 6.2 shows the experimental results on probe set.

It should be noted that the proposed method achieves the best recognition accuracies at almost all angles. This is the result of gait energy image over the silhouette sequence. For evaluation of the proposed model, we compared our model with GEI template methods, as input to model is GEIs. This are the models—Gait-Net [5], L-CRF [15], LB [16], RLTD [17], CPM [12], and DV-GEIs [14]. Table 6.3 shows the comparison of the benchmark approaches and state-of-the-art models with the proposed approach with respect to some of the probe evaluation CASIA-B gait dataset.

Table 6.2 Shows cross-view gait recognition accuracy on CASIA-B dataset

Gallery	Probe										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
0°	–	95.71	94.86	94.25	95.97	95.98	96.03	95.98	95.98	95.99	96.03
18°	95.27	–	94.38	95.50	95.98	95.94	95.98	95.94	95.89	95.76	95.72
36°	95.51	95.69	–	91.92	91.51	90.08	92.39	94.73	94.79	95.97	95.63
54°	93.50	92.24	89.29	–	77.47	73.39	75.78	85.42	90.84	94.45	93.63
72°	95.62	95.57	95.70	95.63	–	94.56	93.65	94.86	95.57	95.86	95.60
90°	96.00	95.90	96.01	95.76	96.50	–	95.99	95.78	95.99	96.02	95.95
108°	95.93	95.87	96.09	96.25	96.64	96.68	–	96.17	96.11	96.02	95.97
126°	96.04	95.73	95.86	96.07	96.30	96.44	96.51	–	96.15	95.99	96.06
144°	93.29	89.35	89.71	91.23	90.13	88.09	85.84	87.40	–	90.67	90.29
162°	93.80	92.73	89.15	84.83	84.35	83.15	81.94	86.97	89.26	–	94.17
180°	94.35	90.64	88.35	86.80	85.74	85.37	84.27	86.56	90.72	95.37	–

Table 6.3 Comparison with under different walking view angles on CASIA-B by accuracies

Probe	Gallery	Proposed model	GaitNet [5]	L-CRF [38]	LB [37]	RLTDA [17]
54°	36°	91.92	91.6	93.8	92.7	80.8
54°	72°	95.63	90.0	91.2	90.4	71.5
90°	72°	94.56	95.6	94.4	93.3	75.3
90°	108°	95.99	87.4	89.2	88.9	76.5
126°	108°	96.17	90.1	92.5	93.3	66.5
126°	144°	87.40	93.8	88.1	86.0	72.3
	Mean	93.61	91.4	91.5	90.8	73.8

Table 6.4 Comparison of recognition cross-view accuracy on CASIA-B dataset

Methods	Average accuracy (%)
CPM [12]	24.1
GEI-SVR [13]	42.2
CMCC [18]	43.9
ViDP [19]	45.4
LB [16]	56.9
L-CRF [15]	67.8
Gait-Net [5]	81.8
DV-GEIs [14]	83.4
Proposed model	93.0

It should be noted from Table 6.2. The average accuracy of the proposed model is around 93.0%, which outperforms the GaitNet with 12.2% and DV-GEI with 9.6%. Compared the output with state-of-the-art methods like L-CRF, use GEIs instead of.

Silhouette sequence as an input to the model, this is another proof that our gait characteristics strongly illustrate all major gait variations. Table 6.4 shows the comparison of cross-view recognition accuracy on CASIA-B dataset.

6.5 Conclusion

The deep learning approach is proposed in this paper. The proposed method can efficiently extract spatial and temporal parameters. In this model, we fed gait energy image (GEIs) as input to the model rather than gait silhouette series. This method is efficient than conventional methods. The gait energy image keeps the dynamic and static facts of a gait sequence, although time is not properly considered, this is the one limitation of the proposed method. The proposed model is robust in variations including obstructive objects, clothes, and viewing angles. Eventually, gait awareness develops stronger. Our model was tested on a broad CASIA-B dataset. Max pooling

layer is used to adapt the spatial information, improving the mapping efficiency. Recognition results obtained demonstrated our model's dominance over well-known approaches.

References

1. Sepas-Moghaddam, A., Etemad, A.: View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Trans. Biometrics Behavior Identity Sci.*
2. Sepas-Moghaddam, A., Ghorbani, S., Troje, N.F., Etemad, A.: Gait recognition using multi-scale partial representation transformation with capsules (2020). arXiv preprint [arXiv:2010.09084](https://arxiv.org/abs/2010.09084)
3. Lin, B., Zhang, S., Yu, X., Chu, Z., Zhang, H.: Learning effective representations from global and local features for cross-view gait recognition (2020). arXiv preprint [arXiv:2011.01461](https://arxiv.org/abs/2011.01461)
4. Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., He, Z.: GaitPart: temporal part-based model for gait recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14225–14233 (2020)
5. Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J., Wang, N.: Gait recognition via disentangled representation learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4710–4719 (2019)
6. Chao, H., He, Y., Zhang, J., Feng, J.: Gaitset: regarding gait as a set for cross-view gait recognition. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8126–8133 (2019)
7. Moissenet, F., Leboeuf, F., Armand, S.: Lower limb sagittal gait kinematics can be predicted based on walking speed, gender, age and BMI. *Sci. Rep.* **9**(1), 1–12 (2019)
8. Nandy, A., Chakraborty, R., Chakraborty, P.: Cloth invariant gait recognition using pooled segmented statistical features. *Neurocomputing* **191**, 117–140 (2016)
9. Anusha, R., Jaidhar, C.D.: Human gait recognition based on histogram of oriented gradients and Haralick texture descriptor. *Multimedia Tools Appl.* 1–22 (2020)
10. Hasan, M.M., Mustafa, H.F.: Multi-level feature fusion for robust pose-based gait recognition using RNN. *Int. J. Comput. Sci. Inform. Secur. (IJCSIS)* **18**(1) (2020)
11. Janković, M., Savić, A., Novičić, M., Popović, M.: Deep learning approaches for human activity recognition using wearable technology. *Medicinski podmladak* **69**(3), 14–24 (2018)
12. Chen, X., Weng, J., Wei, Lu., Jiaming, Xu.: Multi-gait recognition based on attribute discovery. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(7), 1697–1710 (2017)
13. 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*, pp. 974–981. IEEE (2010)
14. Liao, R., An, W., Yu, S., Li, Z., Huang, Y.: Dense-view GEIs set: view space covering for gait recognition based on dense-view GAN (2020). arXiv preprint [arXiv:2009.12516](https://arxiv.org/abs/2009.12516)
15. Recognition based on attribute discovery. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **40**(7), 1697–1710 (2018)
16. Wu, Z., 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 (2017)
17. Haifeng, Hu.: Enhanced gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition. *IEEE Trans. Circ. Syst. Video Technol.* **23**(7), 1274–1286 (2013)
18. Kusakunniran, W.: Recognizing gaits on spatio-temporal feature domain. *IEEE Trans. Inf. Forensics Secur.* **9**(9), 1416–1423 (2014)

19. Hu, M., Wang, Y., Zhang, Z., Little, J.J., Huang, Di.: View-invariant discriminative projection for multi-view gait-based human identification. *IEEE Trans. Inf. Forensics Secur.* **8**(12), 2034–2045 (2013)