Cross View and Cross Walking Gait Recognition Using a Convolutional Neural Network

Sonam Nahar(✉), Sagar Narsingani, and Yash Patel

Pandit Deendayal Energy University, Gandhinagar, Gujarat, India
{sonam.nahar,sagar.nce19,yash.pce19}@sot.pdpu.ac.in

Abstract. In this paper, we propose a gait recognition method using a convolutional neural network (CNN). A CNN architecture is designed and trained to learn an efficient representation with which walking patterns i.e., gait can be disentangled from the visual appearance of the subjects caused by covariate factors such as variation in view angles, clothing and carrying conditions. Since dynamic areas contain the most informative part of the human gait and are insensitive to changes in various covariate conditions, we feed the gait entropy images as input to CNN model to capture mostly the motion information. The learned gait features from CNN are then fed into a K-NN classifier to identify individuals based on their unique gait patterns. Experiments are carried out for cross-view and cross-walking gait recognition using the CASIA-B dataset. Our experimental results demonstrate the effectiveness of the proposed method.

Keywords: Gait Recognition · CNN · Cross View · Cross Walking

1 Introduction

'Gait' is defined as the way people walk and is used as a behavioral biometric [9]. The gait pattern can be captured and perceived from a distance in an unconstrained background. It does not require subject cooperation and can operate without interrupting or interfering with the subject's activity unlike other kinds of biometric features such as face, ear, iris, and fingerprint. Gait recognition is a relatively new and rapidly evolving field in computer vision and biometric research community [11,13]. However, for vision-based gait recognition, one of the biggest challenges is to disentangle the covariate factors which can alter gait appearances drastically and make the recognition process difficult. It is essential that gait recognition is more robust against these covariates. In this work, we consider the covariates such as view angles, walking with carrying a bag and walking with wearing a coat, and propose a novel gait feature extraction method which can automatically select gait features invariant to these covariates for subject recognition.

H. Kaur et al. (Eds.): CVIP 2023, CCIS 2009, pp. 112–123, 2024.
https://doi.org/10.1007/978-3-031-58181-6_10

In general, silhouettes are used to represent the human gait. However, gait silhouettes are sensitive to changes in the appearance of the subject and to overcome this problem, several state-of-art gait recognition methods have proposed gait representations which are invariant to different covariate conditions [1,4,6,7,15]. Even though these strategies have provided relatively satisfactory results in the past years, they are usually constrained to hand-crafted features and have limited capacity for learning intrinsic patterns in data.

Recently, deep learning has become popular in gait recognition task because of it's ability to learn the gait features from the large amount of gait data [11]. In particular, convolutional neural networks (CNN) has been mostly used in gait recognition, especially as feature extractor [2,12,17]. CNNs use convolutional layers to extract local features, pooling layers to reduce dimensionality, and fully connected layers to classify images based on learned features. Our work is similar to the method presented in [12]. In the method entitled as GEINet [12], a CNN was trained in order to learn the gait features, given gait energy images as input. For recognition, a simple Euclidean distance measure was used between the gallery and probe features. While authors in [12] reported that their method outperformed on the benchmark dataset for cross view gait recognition, we find few limitations: (1) they used gait energy images as input to the CNN because GEI can efficiently capture both the static (e.g., head, torso) and dynamic parts (e.g., lower parts of legs and arms) of the human silhouette. However, since GEI mainly contain body shape information, they are sensitive to changes in various covariate conditions, and hence is not an appropriate representation to feed in a CNN for learning robust gait features, (2) The method has shown the results in cross view settings only. Gait recognition should also be tested in cross walking scenarios where the same person walks with different carrying and clothing conditions, (3) they have chosen the CNN architecture based on empirical results, no systematic hyperparameter tuning was performed. (4) The gallery and probe gait features extracted from CNN were used as templates and the gait recognition problem was solved by measuring the distance between templates directly. However, direct template matching is susceptible to noise.

We therefore propose a gait recognition method using CNN that learns suitable representations with which walking patterns i.e., gait can be disentangled from the visual appearance of the subjects and subsequently used for recognition, and also demonstrate its effectiveness in the settings of cross view and cross walking using the benchmark dataset [18]. Here, cross-view means the probe gait sequences are with different view angles as angles in gallery sequences. In cross-walking setting, the subjects have walking sequences either with a coat or with a bag, while subjects in the gallery are under the normal walking condition. The key contributions of the paper are summarized as follows:

1. We propose to use gait entropy images [1] to be fed as input to our CNN model to capture mostly the motion information because gait entropy image captures the dynamic areas (motion) of the human body as dynamic areas contain the most informative part of the human gait and are insensitive to changes in various covariate conditions.

2. We design our CNN model consisting of two sets of convolutions, batch normalization and pooling layers followed by two fully connected layers and employ a method of cross-validation for tuning the hyperparameters such as number of filters, size of filters, number of epochs, dropout rate and batch size.
3. With the learned gait features in gallery and probe set, K-NN classifier is used to recognize the subject uniquely.
4. We present extensive experimental results for cross-view and cross-walking gait recognition using the CASIA-B benchmark dataset [18].

The rest of the paper is organized as follows: in Sect. 2, related work is reviewed. The proposed gait recognition method is detailed in Sect. 3. Experimental results and conclusion are presented in Sect. 4 and Sect. 5, respectively.

2 Related Work

In general, model-free gait recognition approaches represent gait based on silhouettes extracted from the human walking sequences. Gait energy image (GEI) is proven to be an effective gait representation which is obtained by averaging the silhouette value pixel-by-pixel over the gait period [4]. However, GEI loses information in a gait sequence which affects performance due to changes caused by covariate conditions such as clothing, carrying conditions and view variations. To mitigate these affects, gait entropy image is proposed [1] which focus on dynamic regions and is computed as pixel-by-pixel entropy of the GEI. In another variant of GEI, to preserve the temporal information from loss, Chrono-Gait Image (CGI) is proposed which is based on multi-channel temporal encoding scheme [15]. A gait flow image (GFI) focuses more directly on the dynamic components, where the optical flow lengths observed on the silhouette contour are averaged over the gait period [6]. Frequency domain gait features are proposed by considering the periodic property of gait [7]. [5]. In such frequency-based methods cross-view projections are learned with which one can normalize gait features from one view to another, and hence one can compare the normalized gait features extracted from any two videos in order to compute their similarity.

The state-of-the-art methods based on such gait representations have shown promising results in cross-view and cross-walking scenarios [11,13]. However, these methods use gait features which are hand-crafted and have limited capacity for learning intrinsic patterns in data. In addition, these image-based gait features are transformed into a feature vector and the techniques such as linear discriminant analysis (LDA) [4], primal rank support vector machines [3], multi-view discriminant analysis (MvDA) [8] are applied to extract the relevant gait features that are invariant to various covariate conditions. However, each dimension in the feature vector corresponds to each pixel for further classification/recognition, and the spatial proximity in the gait image is not captured which results in overtraining.

Recently, gait recognition methods based on deep learning have dominated the state of the art in the field through the ability to automatically learn discriminative gait representations. In particular, a CNN is mostly used because it considers spatial proximity in the image using a convolutional operation and improve the recognition accuracy significantly. The work termed as GEI-Net [12] learns gait representations from GEI via convolutional neural networks directly. Authors in [17] proposed a deep convolutional neural network based cross-view and cross-walk gait recognition framework which learns similarities between pairs of GEIs. Another recent work is GaitSet [2] makes an assumption that appearance of a silhouette contains its position information and thus regards gait as a set to extract temporal information using CNN.

In addition to CNN, recently several other deep architectures [11] have also been designed to solve the gait recognition problems, for example, deep belief networks (DBN), long short-term memory (LSTM) - a type of recurrent neural network (RNN) architecture, deep autoencoders (DAE), generative adversarial networks (GAN), capsule network, and hybrid networks that combines one or more of these architectures. Though, these methods have shown excellent results in the challenging cross-view and cross-walking scenarios, they form a very complex network and require huge amount of labeled data during training. In contrast, we present a simple CNN architecture which can learn efficient gait features using a moderate size of training data.

3 Proposed Method

3.1 Generation of Gait Entropy Images

Given a human walking sequence, a silhouette is extracted from each frame using the background subtraction [10]. The height of the silhouettes is then normalized followed by the center alignment. Gait cycles are then estimated using the method of autocorrelation presented in [7]. The gait cycle is defined as the time interval between the same repetitive events of walking that generally starts when one foot is in contact with the ground. Since, the walking of a person is periodic, it is sufficient to consider only one gait cycle from the whole gait sequence.

Given a gait cycle of size-normalized and center aligned silhouettes, a gait entropy image (GEnI) is computed by calculating the Shannon entropy for each pixel in the silhouette images over a complete gait cycle as given in [1]:

$$GEnI = I(x, y) = \sum_{k=1}^{K} p_k(x, y) \log_2 p_k(x, y), \tag{1}$$

where x, y are the pixel coordinates and $p_k(x, y)$ is the probability that the pixel takes on the k^{th} value in a complete gait cycle. In our case, the silhouettes are binary images and we thus have K = 2. Figure 1 shows some examples of gait gntropy images along with gait energy images from the CASIA-B dataset. One can clearly see that dynamic areas such as legs and arms are represented by higher intensity values whilst the static areas such as head, torso have low values in the GEnIs. This is because silhouette pixel values in the dynamic areas

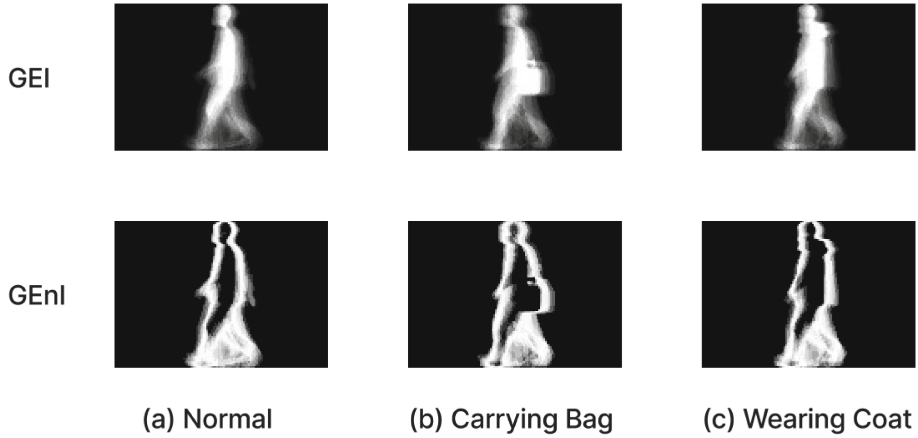

Fig. 1. Examples of Gait Energy Images (GEI) and Gait Entropy Images (GEnI) from the CASIA-B dataset [18] with different walking conditions. Columns (a) Normal Walking, (b) Carrying a bag, and (c) Wearing a coat.

are more uncertain and thus more informative, leading to higher entropy values. It can also be observed that the effect of appearance changes caused by carrying a bag and wearing a coat is significant in GEIs whereas it is clearly reduced in GEnIs and shown only in the outer contour of the human body. In our proposed method, we feed GEnI as input to our CNN architecture.

3.2 Convolutional Neural Network Architecture

Our CNN architecture consists of eight layers, where the leading six layers are the two sets of convolutions, batch normalization and pooling layers. The last two layers are the fully connected (FC) layers where the first FC layer has 1024 units followed by another FC layer consisting of M units where SoftMax function is applied at each output unit. Let us assume that there are M number of subjects in the training set and each subject is denoted by number ranging from 1, 2, ..., M.

More specifically, the i^{th} unit of the last layer ideally outputs 1 for the input belonging to subject i, otherwise, it outputs 0. Figure 2 shows our CNN architecture. Dropout for regularization and ReLU activation function for non-linearity is applied at every layer of the network except at the last layer. We consider number of filters, size of filters, number of epochs, dropout rate and batch size as set of hyperparameters for our CNN model and choose their optimal values using cross validation. The configuration for each convolution and pooling layer with optimal set of hyperparameters is illustrated in Table 1. We consider this configuration for learning relevant gait features for the test data.

Though our CNN architecture is not very deep, it learns robust gait representations based on input GEnI. This is due to the fact that input gait data

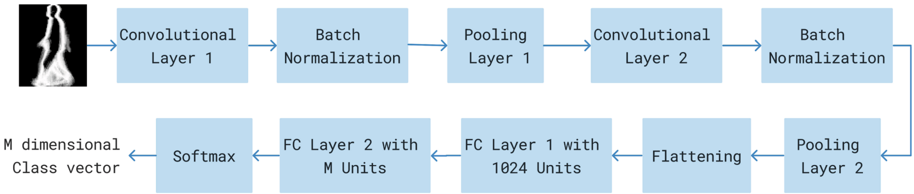

Fig. 2. The CNN Architecture

Table 1. CNN Configuration with Optimal Hyperparameters

Hyperparameters		
Conv Layer 1	# Filters	40
	Size of Filters	3×3
	Stride	2
Max Pooling Layer 1	Size of Filters	2×2
	Stride	2
Conv Layer 2	# Filters	32
	Size of Filters	5×5
	Stride	3
Max Pooling Layer 2	Size of Filters	3×3
	Stride	2
Number of Epochs	100	
Learning Rate	0.001	
Dropout Rate	40%	
Batch Size	64	

in form of silhouettes (for e.g., GEnI) do not present considerable complexity in terms of texture information. Hence, a shallow CNN architecture is sufficient for encoding gait data. This is contrary to many other domains such as face [14] or activity [16] recognition, where very deep networks are used to learn highly discriminative features. Moreover, through preliminary experiments, we have confirmed that additional convolutional layers following two existing sets of convolution, batch normalization and pooling in our network do not help in improving the significant gait recognition accuracy.

3.3 Feature Learning

Let the training set contains N number of gait sequences belonging to M number of subjects. We first compute a gait entropy image for every gait sequence and obtain a set of training GEnIs as $\{I_1, I_2, \ldots, I_N\}$ with their corresponding ground truth label vectors as $\{y_1, y_2, \ldots, y_N\}$. For an input I_i belonging to subject j, it's label is denoted as M dimensional vector, $y_i = [y_{i1}, y_{i2}, \ldots, y_{iM}]$ where $y_{ij} = 1$ and the other entries are equal to 0.

Given an input image I_i and set of weighting parameters w, we define the output $\hat{z}_i$ of last layer as a function of I_i and w using forward propagation as:

$$\hat{z}_i = f(I_i, w), \tag{2}$$

where $\hat{z}_i$ is denoted as a M dimensional vector. We subsequently apply the SoftMax function on every element of $\hat{z}_i$ and obtain the final class probability vector $\hat{y}_i$ where each j^{th} element in $\hat{y}_i$ is computed using the softmax function as follows,

$$\hat{y}_{ij} = \frac{e^{\hat{z}_{ij}}}{\sum_{j=1}^{M} e^{\hat{z}_{ij}}} \quad for\ j = 1, 2, \ldots, M \tag{3}$$

Given a training set: $\{I_1, I_2, \ldots, I_N\}$, we train our CNN network by minimizing the following cross-entropy loss function L using the stochastic gradient descent algorithm:

$$L(w) = \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log \hat{y}_{ij} \tag{4}$$

In this learning phase, we obtain an optimal set of weighting parameters w.

3.4 Gait Recognition

A GEnI is fed as an input to the trained network and the output computed at the fully connected layer (FC1), which is the layer immediately preceding the last year, is considered as a feature vector to represent the gait in our work. Since the FC1 layer consists of 1024 units, we obtain 1024-dimensional feature vector. For a given gallery and probe gait sequences with their corresponding gait entropy images, we extract the gait features from our trained CNN architecture and use K-NN classifier for gait recognition. The K-NN classifier is trained using the features of gait sequences in gallery set and the subject/class is recognized for every gait sequence in probe set.

4 Experimental Results

4.1 Datasets and Test Protocol

We use CASIA-B dataset [18] for our experiments. CASIA-B is one of the most widely used and publicly available gait database which contains gait sequences of 124 subjects with 11 different views ranging from 0° to 180° (with 18° increments). The dataset considers three different walking conditions namely normal walking (NM), walking with a coat (CL), and walking with a bag (BG), respectively with 6, 2, and 2 gait sequences per subject per view.

In order to test the performance of our proposed method, we use the subject independent testing protocol. In subject independent testing protocol, the subjects in the training set are disjoint from the subjects in test set. We choose randomly 74 subjects for training and the remaining 50 subjects for testing. Training set consists of $11 * 10 = 110$ gait sequences per subject i.e., total $110 * 74 = 8140$ sequences. We compute the gait entropy image for every gait sequence in the training set and obtain a fixed size GEnIs of 88×128 pixels. Example GEnIs extracted from this dataset can be found in Fig. 1. Since our training set is moderate in size and hence to reduce the problem of overfitting, we perform the cross validation by randomly splitting the training set into train (70%) and validation (30%) set, and choose the hyperparameters that reduce the variance error, for instance, we consider those set of hyperparameters where both the train and validation errors are low and the difference between these errors are also low as well. With the training GEnIs and best set of hyperparameters (see Table 1), we train our CNN model and learn the parameters (w) of the model. Since our model is trained using normal walking, carrying bag and wearing coat gait sequences captured using different view angles, it can learn the reliable gait features invariant to these covariates.

The test set consisting of 50 subjects is further divided into a gallery set including the first four gait sequences from the NM gait data and the probe set consists of the rest of the sequences namely, the remaining 2 NM (probe subset-NM), 2 CL (probe subset-CL), and 2 BG (probe subset-BG) sequences, per each subject per each view. This testing scenario is named as cross-walking because here the subjects in the gallery are under the normal walking conditions while probe contains the walking sequences with coat or bag. We perform the experiments for gait recognition in cross walking along with cross view settings where the probe gait sequences are with different view angles from the view angles in gallery sequences. Note that we have used TensorFlow for training the CNN and all experiments were conducted on machine with 11th Gen Intel(R) Core(TM) i7-1165G7, 2.80 GHz frequency, SSD - 512 MB, RAM - 8 GB and System Type - 64-bit operating system.

4.2 Results

We learn the gallery and probe features using the trained CNN model and use K-NN classifier (with K = 1) to compare the probe features with the gallery ones in order to identify most similar gait patterns and label them as being from the same subject. Results are measured and presented using rank-1 recognition accuracy. Table 2 shows the results for cross view scenario where both the gallery and probe gait sequences belong to same walking conditions (4 NM in gallery and 2 NM in probe) with different view angles. Table 3 and Table 4 demonstrate the cross-view recognition accuracies when probe contains bag and coat sequences, respectively.

Cross view and cross walking gait recognition is challenging. Our method has obtained promising results on the NM probe subset, when the cross-view angle is not larger than 18° as shown in Table 2. The results are better with probe angles

Table 2. Cross-View recognition accuracies when both gallery and probe contains normal walking sequences.

		Probe Angles (Normal Walking: NM # 5-6)										
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Angles: NM #1-4	0°	86.73	57.9	31	17	9	4.04	6.06	13	12	19	28
	18°	66.32	84	69	20	19	12.12	8.08	11	15	28.9	37
	36°	27.55	51	76	66	16	6.06	5.05	8	15	26	21
	54°	13.26	21	36	76	30	19.2	17.2	17	18	15	12
	72°	8.16	13	16	23	97	71.71	49.5	43	28	14	7
	90°	6.12	12	9	10	81	96.97	91.91	53	19	9	10
	108°	6.12	4	7	10	46	90.9	93.93	74	31	12	6
	126°	6.12	9	9	14	57.9	71.71	86.86	94	67	21	11
	144°	9.18	12	18	20	41	33.33	43.43	78	93	59	24
	162°	17.34	23	19	18	21	16.17	13.13	19	55	96	60
	180°	24.48	18	13	6	9	5.05	3.03	9	15	44	97

Table 3. Cross-View recognition accuracies when gallery contains normal and probe with bag sequences.

		Probe Angles (Bag: BG #1-2)										
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Angles: NM #1-4	0°	54.08	51.02	26.26	10.20	7.14	3.06	1.02	5	8	10.1	16
	18°	26.53	53.06	28.28	9.18	9.18	4.08	5.10	6	6	12.12	18
	36°	16.32	34.69	43.43	23.46	8.16	3.06	3.06	5	10	12.12	12
	54°	12.24	17.34	21.21	42.85	19	12.24	13.26	14	14	16.16	11
	72°	6.12	12.24	10.10	13.26	68.36	37.75	23.46	21	16	11.11	10
	90°	5.10	9.18	7.07	11.22	52.04	69.38	59.18	31	17	3.03	8
	108°	4.08	6.12	5.05	7.14	38.77	67.34	62.24	39	19	10.10	6
	126°	8,16	3.06	8.08	8.16	35.71	51.02	48.97	61	30	15.15	11
	144°	10.20	9.18	11.11	13.26	25.51	20.40	21.42	37	66	28.28	16
	162°	14.28	16.32	12.12	16.32	21.42	6.12	3.06	6	28	66.67	40
	180°	15.30	19.38	16.16	8.16	5.10	5.10	3.06	6	5	31.31	67

72°, 90°, 108° and 126° because most of the gait information is visible in these viewpoints. Especially, with cross view of 90° and 108° angles, we obtain more than 90% accuracy on the NM probe set. Even though the silhouettes in frontal and back viewpoints such as 0°, 18°, 162° and 180° carry little gait information, our method works well with more than 60% accuracy with the NM probe subset and the cross-view angle less than 18° as shown in Table 2. As for the BG probe subset, our method still performs well (see Table 3). However, the performance degrades for the CL probe subset even when the cross-view angle is not larger than 18° as depicted on Table 4. The cause behind these results is that carrying a bag only affects a small part of a gait silhouette, while wearing a coat can greatly change the appearance. Another possible reason for the performance degradation may be the lack of training data. Due to larger appearance variations, the CL

Table 4. Cross-View recognition accuracies when gallery contains normal and probe with coat sequences.

		Probe Angles (Coat: CL #1-2)										
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Angles: NM #1-4	0°	24.48	19.38	13.26	10.10	10	2	4	5.05	8.08	12.24	7
	18°	23.46	24.48	21.42	11.11	14	6	6	11.11	10.10	14.28	8
	36°	8.16	18.36	29.59	22.22	15	6	6	6.06	7.07	13.26	8
	54°	7.14	12.24	20.40	36.36	12	9	8	12.12	9.09	17.34	11
	72°	4.08	11.22	7.14	14.14	43	42	16	26.26	14.14	8.16	5
	90°	5.10	5.10	7.07	24	41	43	22.22	13.13	4.08	1	3.06
	108°	3.06	1.02	3.06	10.10	14	40	46	25.25	17.17	7.14	4
	126°	3.06	7.14	5.10	8.08	17	28	32	37.37	30.3	12.24	4
	144°	3.06	11.22	3.06	20,20	19	19	17	26.26	40.4	19.38	15
	162°	6.12	15.30	11.22	13.13	8	11	8	13.13	20.2	43.87	19
	180°	7.14	9.18	6.12	4.04	4	1	3	7.07	9.09	19.38	36

subset is harder than the NM and BG subsets. Networks may easily over-fit if the training set is not big enough. For all the three probe subsets: NM, BG, and CL, our method needs further improvements when the cross-view angle is larger than 18°.

Table 5. Comparison of cross-view methods on CASIA-B in terms of Rank-1 mean accuracy (%). Here, '-' denotes that results are not reported.

Gallery (NM #1-4): 0°–180°													
Probe		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
NM #5-6	GEI	9.14	14.87	15.03	16.07	23.75	22.5	23.91	17.2	12.44	13.49	9.02	16.13
	GEnI	12.45	16.12	13.1	12.38	15.02	23.62	23.21	21.10	15.20	15.57	16.18	16.72
	CMCC [5]	46.3	–	–	52.4	–	48.3	–	56.9	–	–	–	-
	GEINet [12]	45.8	57.6	67.1	66.9	56.3	48.3	58.3	68.4	69.4	59	46.5	58.5
	CNN-LB [17]	79.1	88.4	95.7	92.8	89.1	87	89.3	92.1	94.4	89.4	75.4	88.4
	GaitSet [2]	90.8	97.9	99.4	96.9	93.6	91.7	95	97.8	98.9	96.8	85.8	95
	Proposed	24.67	27.72	25.73	22.73	36.99	38.84	38.02	38.09	33.45	30.35	28.45	31.37
BG #1-2	GEI	6.44	9.91	11.61	9.32	15.32	12.21	11.65	11.97	8.26	8.94	6.37	10.18
	GEnI	7.64	9.82	7.31	7.08	9.81	13.63	12.37	12.59	7.27	7.26	7.97	9.34
	CMCC [5]	–	–	–	–	–	–	–	–	–	–	–	-
	GEINet [12]	–	–	–	–	–	–	–	–	–	–	–	-
	CNN-LB [17]	64.2	80.6	82.7	76.9	64.8	63.1	68	76.9	82.2	75.4	61.3	72.4
	GaitSet [2]	83.8	91.2	91.8	88.8	83.3	81	84.1	90	92.2	94.4	79	87.2
	Proposed	17.11	21.05	17.17	14.84	26.4	25.41	22.17	21	19.91	19.65	19.55	20.39
CL #1-2	GEI	2.57	4.27	6.23	6.32	6.28	7.01	6.6	6.97	5.18	4.62	2.54	5.33
	GEnI	3.02	5.01	4.61	4.52	6.44	8.24	8.06	8.64	6.36	5.33	4.42	5.88
	CMCC [5]	–	–	–	–	–	–	–	–	–	–	–	-
	GEINet [12]	–	–	–	–	–	–	–	–	–	–	–	-
	CNN-LB [17]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54
	GaitSet [2]	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50	70.4
	Proposed	8.62	12.24	11.59	15.77	17.55	17.91	5.45	16.62	15.43	15.3	10.91	13.4

The comparison of our proposed method with those in the literature is presented in Table 5. Note that each of these identification accuracies is the average score for a given probe view angle with different gallery view angles (0°–180°). These methods are listed because they are the most recent and are evaluated under cross-view and cross-walking scenarios, and their scores are directly taken from the original papers. We also include two traditional methods (non deep) such as GEI and GEnI for fair comparison. These methods use GEI and GEnI as features, respectively and perform gait recognition in cross walking and cross view scenarios by direct template matching (Euclidean distance) between gallery and probe set. There are little results reported under cross-walking where probe set belongs to BG and CL. Our method performs better than the traditional methods such as GEI and GEnI with significant margins. This shows the advantage of using the learned features from CNN as compared to handcrafted features in gait recognition. In the case of experimental setting where both the gallery and probe set contains the NM sequences, our method shows comparable performance when compared to the methods CMCC [5] and GEINET [12], especially, when the probe angle is 90°. The proposed method does not show significant improvements when compared to the recent deep learning based gait recognition methods, GaitSet [2] and CNN-LB [17] because these recent methods use the view transformation models and are trained using a large dataset. Overall, our method outperforms the traditional gait recognition method in cross view setting under the 3 different cross walking scenarios such as probe with NM, BG and CL, respectively.

5 Conclusion

This paper presents a gait recognition method using CNN. Given gait entropy images, the CNN is trained to learn the gait features invariant to viewpoints, clothing and carrying conditions. Finally, a K-NN classifier is used to compare the probe features with the gallery ones in order to identify most similar gait patterns. The experimental results show the effectiveness of our approach in cross-view and cross-walking scenarios and comparison with state of the art traditional and deep learning-based gait recognition methods.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), pp. 1–6 (2009)
2. Chao, H., He, Y., Zhang, J., Feng, J.: Gaitset: regarding gait as a set for cross-view gait recognition. CoRR abs/1811.06186 (2018)
3. Chen, X., Xu, J.: Uncooperative gait recognition. Pattern Recogn. **53**(C), 116–129 (2016)
4. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)

5. Kusakunniran, W., Wu, Q., Zhang, J., Li, H., Wang, L.: Recognizing gaits across views through correlated motion co-clustering. IEEE Trans. Image Process. **23**(2), 696–709 (2014)
6. Lam, T., Cheung, K., Liu, J.: Gait flow image: a silhouette-based gait representation for human identification. Pattern Recogn. **44**, 973–987 (2011)
7. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 151–163. Springer, Heidelberg (2006). https://doi.org/10.1007/11744078_12
8. Mansur, A., Makihara, Y., Muramatsu, D., Yagi, Y.: Cross-view gait recognition using view-dependent discriminative analysis. In: IEEE International Joint Conference on Biometrics, pp. 1–8 (2014)
9. Murray, M.: Gait as a total pattern of movement (1967)
10. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)
11. Sepas-Moghaddam, A., Etemad, A.: Deep gait recognition: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **45**(1), 264–284 (2023)
12. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y.: Geinet: view-invariant gait recognition using a convolutional neural network. In: 2016 International Conference on Biometrics (ICB), pp. 1–8 (2016)
13. Singh, J.P., Jain, S., Arora, S., Singh, U.P.: Vision-based gait recognition: a survey. IEEE Access **6**, 70497–70527 (2018)
14. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
15. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
16. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: a survey. CoRR abs/1707.03502 (2017)
17. 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)
18. 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 2006), vol. 4, pp. 441–444 (2006)