IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 26, 2020, accepted December 8, 2020, date of publication December 24, 2020, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047266

An Efficient Gait Recognition Method for Known and Unknown Covariate Conditions

MARYAM BUKHARI[1], KHALID BASHIR BAJWA[2], SAIRA GILLANI[3], MUAZZAM MAQSOOD[1], MEHR YAHYA DURRANI[1], IRFAN MEHMOOD[4], (Member, IEEE), HASSAN UGAIL[4], (Member, IEEE), AND SEUNGMIN RHO[5]

[1]Department of Computer Science, COMSATS University Islamabad–Attock Campus, Attock 43600, Pakistan
[2]Department of Computer Science, Islamic University of Madinah, Madinah 42351, Saudi Arabia
[3]Department of Computer Science, Bahria University–Lahore, Lahore 54600, Pakistan
[4]Centre for Visual Computing, Faculty of Engineering and Informatics, University of Bradford, Bradford BD71DP, U.K.
[5]Department of Industrial Security, Chung-Ang University, Seoul 06974, South Korea

Corresponding authors: Muazzam Maqsood (muazzam.maqsood@cuiatk.edu.pk) and Seungmin Rho (smrho@cau.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1F1A1060668).

ABSTRACT Gait is a unique non-invasive biometric form that can be utilized to effectively recognize persons, even when they prove to be uncooperative. Computer-aided gait recognition systems usually use image sequences without considering covariates like clothing and possessions of carrier bags whilst on the move. Similarly, in gait recognition, there may exist unknown covariate conditions that may affect the training and testing conditions for a given individual. Consequently, common techniques for gait recognition and measurement require a degree of intervention leading to the introduction of unknown covariate conditions, and hence this significantly limits the practical use of the present gait recognition and analysis systems. To overcome these key issues, we propose a method of gait analysis accounting for both known and unknown covariate conditions. For this purpose, we propose two methods, i.e., a Convolutional Neural Network (CNN) based gait recognition and a discriminative features-based classification method for unknown covariate conditions. The first method can handle known covariate conditions efficiently while the second method focuses on identifying and selecting unique covariate invariant features from the gallery and probe sequences. The feature set utilized here includes Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Haralick texture features. Furthermore, we utilize the Fisher Linear Discriminant Analysis for dimensionality reduction and selecting the most discriminant features. Three classifiers, namely Random Forest, Support Vector Machine (SVM), and Multilayer Perceptron are used for gait recognition under strict unknown covariate conditions. We evaluated our results using CASIA and OUR-ISIR datasets for both clothing and speed variations. As a result, we report that on average we obtain an accuracy of 90.32% for the CASIA dataset with unknown covariates and similarly performed excellently on the ISIR dataset. Therefore, our proposed method outperforms existing methods for gait recognition under known and unknown covariate conditions.

INDEX TERMS Gait recognition, covariate conditions, discriminative feature learning, FLDA.

I. INTRODUCTION

Gait is a biometric trait that depicts and measures how people move. Over the decades, gait analysis has been successfully used in different domains, including biometrics and posture analysis for healthcare applications. It has also been used in human psychology where gait analysis using point lights employed for recognition of emotional patterns. The same idea was extended and ultimately resulted in the development of gait signatures through which the identification of individuals can be performed [1]. Borrowing from this, computer vision-based approaches have also used motion analysis and human movement modeling for person identification [2]. In the early days of gait recognition, the focus was to identify and classify the different movement patterns such as walking, jogging, and climbing. Gradually, the focus shifted towards human identification and has become an active area of research. As compared to other biometric traits such as

The associate editor coordinating the review of this manuscript and approving it for publication was Yeliz Karaca.

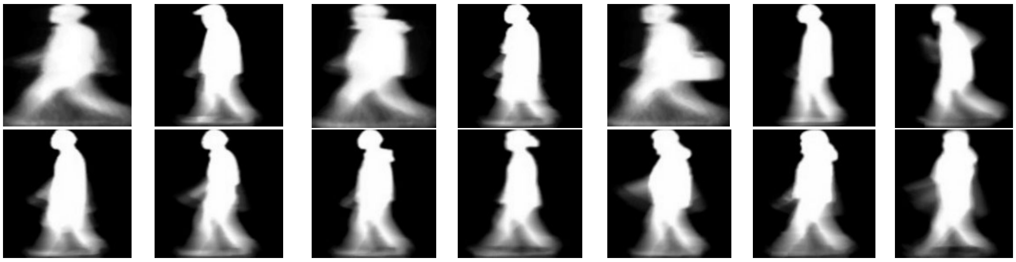

FIGURE 1. The Samples GEI Sequences from CASIA and ISIR Dataset, the first row corresponds to CASIA GEI and the second row corresponds to ISIR.

fingerprint and iris, gait recognition can work without the cooperation of a person. Moreover, it can work without interfering with a person's activity. This makes gait more suitable for different real-time applications like surveillance and long-distance security [3], [4].

Existing techniques employed for gait analysis are divided into model-based and appearance-based methods. The former requires high-resolution videos whereas the latter can deal with low-resolution imagery. Model-based approaches use the parameters of the body, appearance-based approaches on the other hand employ the features extracted directly from image sequences of gait. The simplicity of appearance-based methods and their robustness against noise make them more suitable for real-world scenarios. Appearance-based methods rely on silhouettes extracted from a gait sequence. Silhouettes contain important information about the stance and shape of the human body.

Gait representations used in appearance-based approaches include frequency-domain features, chrono-gait images, features extracted from silhouettes (Gait Energy Image (GEI)), and Gabor GEIs [5]. GEI is popular and creates a single grayscale image from the normalized binary frames over a complete gait cycle and is not susceptible to segmentation errors [6]. It is reported that, in the absence of covariates, direct matching with GEI templates exhibits excellent results [7]. However, in a real-world scenario, the absence of covariates is not always feasible, which makes gait recognition a challenging task. A covariate is a condition when a person appears with a carrying condition, i.e. bag or clothing condition like a coat or long coat, and the system is trained with only normal walk data. To handle this issue, various techniques are used to capture discriminant information from GEIs. One such scheme is proposed in [6], which uses Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for feature extraction. A similar approach is adopted in () [8] where Discriminant Locally Linear Embedding (DLLE) based framework is used for preserving the local structure. However, the main drawback of appearance-based approaches is that they are sensitive to covariate conditions.

The success of gait as a biometric is largely affected due to covariate factors. Some of these factors are clothing, camera viewpoint, carrying conditions, walking style, shoe wear, and walking surface. Some of the examples of clothing and carrying covariate conditions are shown in Figure 1. Currently, most of the gait analysis applications use gait sequences under normal conditions in the training phase and must deal with gait sequences under variable covariate conditions in the testing phase. Owing to this, the performance of these methods for gait recognition under covariate conditions remains unsatisfactory in real-world conditions. The unsatisfactory performance is related to the changes in the underlying representation caused by these conditions. It is evident from Figure. 1 that major changes are seen in portions of the representation that belong to non-moving regions. This leads to the observation that dynamic information is more important as compared to the static part of the representation. When models are trained with covariate conditions and testing is performed on similar covariate conditions, it is known as known covariates. While on the other hand, when models are trained only with simple GEI of a normal walk and tested on different covariate conditions, it is known as unknown covariate conditions.

GEI is a compact representation of a gait sequence representing it in a single image. It is considered a good candidate to extract gait features. Under real-world conditions, the covariate conditions are unknown for the gallery and probe set. However, the known covariate conditions are relatively easy to handle. From this line of research, we propose two methods for gait recognition- one for known covariate conditions and the second for unknown covariate conditions. The first method only takes GEI as input and CNN is used for gait recognition. The second method uses a unique set of features extracted from the ROIs extracted from GEI, which excludes clothing or carrying conditions. The feature set includes Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Haralick texture features. Fisher Linear Discriminant Analysis is used for dimensionality reduction and selecting the most discriminant features. Three classifiers- Random Forest, SVM, and Multilayer Perceptron are used for gait recognition. The objective of this proposed work is to extract discriminative features for unknown covariate conditions. The two standard datasets CASIA and OUR-ISIR are used to evaluate the performance of the proposed work. There are different and complex covariate conditions available in both these datasets, which include clothing and speed variations. The experiments include an extensive set of covariate possibilities for both clothing and speed variation to show the performance of the proposed work under difficult conditions. The results for both these datasets are good and outperforms existing published literature on covariate-based gait recognition. The proposed work has the following contributions:

- A CNN based method to efficiently handle known covariate conditions using only simple GEI
- A discriminative feature learning-based method to handle unknown covariate conditions
- The extraction and selection of discriminative features from ROIs to identify and select unique covariate invariant features from the gallery and probe sequences

The rest of the paper is organized as follows. In Section 2 we explain the related work, Section 3 presents the proposed methodology, Section 4 presents the experiments and results which is followed by a conclusion.

II. LITERATURE REVIEW

A. SPATIAL METRIC LEARNING BASED APPROACHES

These approaches learn a feature space from the original appearance features, which provides resistance against covariates and proves to be more robust. Methods in this category can be further subdivided into whole-based and part-based approaches [9]. In the whole-based approach, to counter against covariates, holistic appearance features are calculated in a discriminative space, an example of this is [5] where LDA is applied on synthesized as well as real GEI templates for the reduction in interclass variation to some extent. The use of a similar approach was advocated in [10] where an RSM framework is proposed to combine inductive biases.

Part-based approaches on the other hand try to divide the holistic appearance-based features into different body parts to enhance features important for gait recognition. This is an important aspect because variation in clothes and carrying status affects only certain parts of a gait representation leaving some of the other parts unaffected. The affected parts are the reason for reduced accuracy. In [11] anatomical knowledge is used, and the body is divided into eight sections. To counter the effects of variations, different weights are assigned to the unaffected and affected sections. A similar strategy is proposed in [12] where the representation of the human body is divided into equal parts and weights are assigned to each part based on similar features extracted from the gait.

B. INTENSITY TRANSFORMATION BASED APPROACHES

As the name suggests, intensity transformation changes the value of the intensity of the gait feature so that it provides resilience against covariate conditions by providing more discriminate values. This approach is exploited in [11] where GEnI is calculated, by using the Shannon entropy method, of the foreground probability of each pixel. GEnI is used for encoding the randomness of each pixel in the gait image within a complete gait cycle. This provides important motion information, instead of the static information, about the change in clothing and change in carrying status. Another such approach known as Masked GEI is proposed in [12]. It is yet another intensity transformation approach that by adopting a certain threshold value keeps the motion information at its original value but it zero-pads the static information (most background and foreground parts). Similarly [13] proposes a so-called gait energy response function that changes the intensities of the pixel thus eliminating the need for native transformation. The concept of joint intensity transformation is extended to include a pair of images instead of one image in [14]. In this approach, a linear SVM based framework is used to learn the intensity metric along with the spatial metric. The main issue with intensity transformation methods is the use of linear optimization for independent transformations.

C. DEEP LEARNING APPROACHES

Deep learning-based approaches have gained popularity in many applications including gait recognition [15]. A Convolutional Neural Network (CNN) takes input from raw silhouettes in each gait sequence. Temporal information along with skeleton data is obtained from the silhouettes with the help of deep graph learning in [16]. Another deep learning-based method is the GEINet which is an eight-layer CNN obtained through average silhouettes (GEI) [17]. They handled the gait recognition as a person classification problem from the same gaits. Similarly, [18] proposes multiple networks with pairs of images (query and enrollment images) which compares images at the start of the input layer. A comparison of input and output architectures for gait recognition using CNN is discussed in [19].

The proposed network compares two input images (GEIs) and determines whether the images are of the same person or not. To counter against multiple covariates an autoencoder is proposed [20] which removes invariant gait features. Generative Adversarial Networks (GAN) have also been used for handling variable covariate conditions in gait. GaitGAN [21] is an adversarial network that is used to generate feature maps removing covariates. The generation of motion features such as optical flow is proposed in [22]. A deep neural network is proposed which provides gait based gender identification aided by clothing and carrying status [23].

III. PROPOSED METHODOLOGY

There are two different methods for cooperative (known covariates) gait recognition and gait recognition under the unknown covariate condition presented in Figure 2 and Figure 3 respectively.

A. GAIT RECOGNITION WITH KNOWN COVARIATE CONDITIONS

1) GATE ENERGY IMAGE (GEI)

By using the method proposed in [12] human silhouettes are extracted from the given gait sequence.

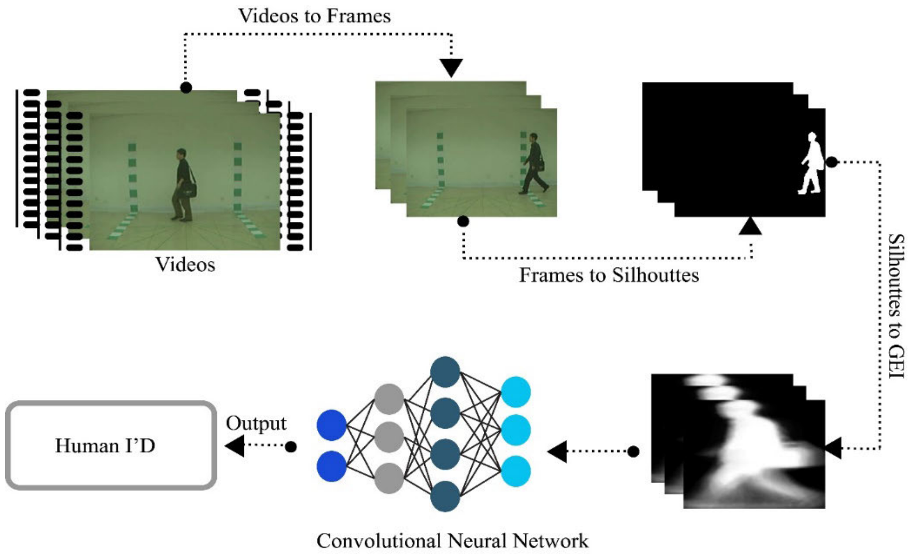

FIGURE 2. The overview of the proposed CNN based gait recognition under normal conditions.

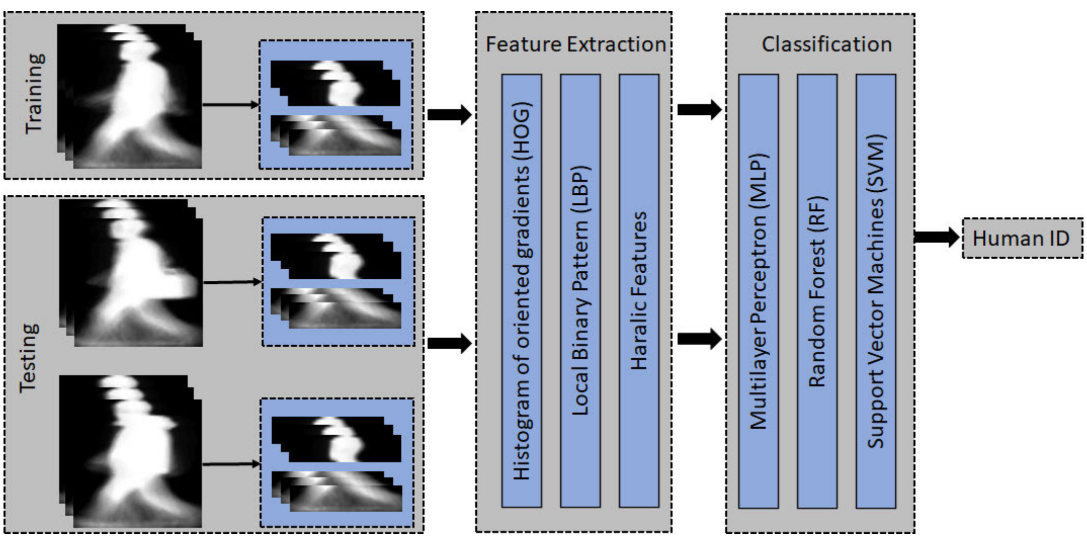

FIGURE 3. The overview of the proposed gait recognition under covariate conditions.

All the images are processed by applying size normalization and horizontal alignment. This is followed by estimation of gait cycle segmentation done by estimation of gait frequency and maximum entropy estimation technique. Finally, Gait Energy Image (GEI) is computed as shown as samples in Figure 1 through the following Equation 1,

$$\mathrm{GEI} = \mathrm{G}(\mathrm{x}, \mathrm{y}) = \frac{1}{\mathrm{T}} \sum\nolimits_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{I}(\mathrm{x}, \mathrm{y}, \mathrm{t}) \tag{1}$$

where T is the total number of frames per gait cycle shown in Figure 2, x and y are the pixel coordinate of the silhouette image I shown in Figure 2 and t correspond to frame number in a gait cycle. High-intensity areas provide information about the shape of the body and stance. Whereas the lower intensity areas describe the movement while walking [12]. The higher intensity areas are known as static areas and lower intensity ones are dynamic areas of a GEI. The dynamic parts have the most important information of a GEI as they are not susceptible to the change of human appearance by clothing and carrying condition. Which is generally the common covariate conditions. Thus, the dynamic areas are most important for human identification in the presence of variable covariate conditions. The static area also provides useful information for human identification (such as hairstyle, body structure). However, they are susceptible to change in covariate conditions.

2) CONVOLUTIONAL NEURAL NETWORKS (CNNS)

The grayscale GEI is given as an input with a 240∗240∗1 dimension to the first input layer as shown in Figure 4.

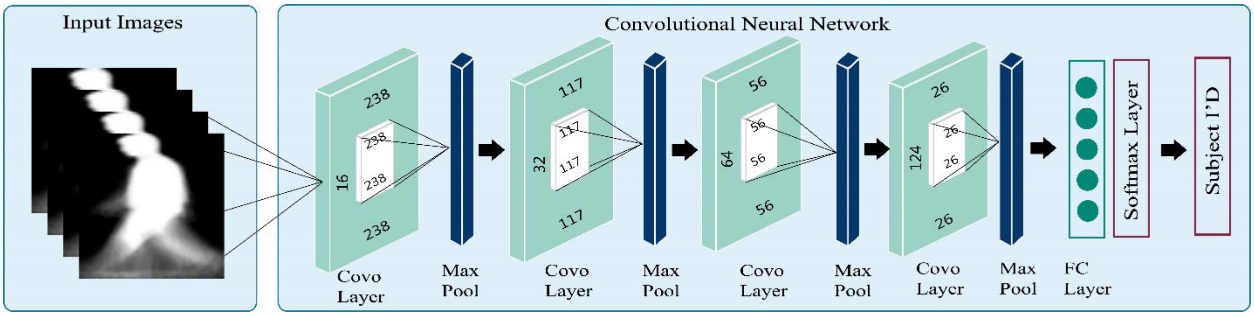

FIGURE 4. The CNN architecture used for cooperative gait recognition process.

We have utilized a total of 10 layers of CNN model with four convolutional layers. The weights of the convolutional filters are initialized through "Xavier Initialization".

The default weight initialization method used in our network is Glorot uniform initialization or "Xavier initialization" and these weights are optimized by the optimizer to best classify the GEI of every subject. In our CNN architecture shown in Figure 4, the filters are generated from the uniform distribution of [-limit, limit], where the limit is,

$$\text{Limit} = \frac{\sqrt{6}}{(\text{fan}_{\text{in}} + \text{fan}_{\text{out}})} \tag{2}$$

where fan_{in} is the number of inputs to layer and fan_{out} is the number of outputs to layer as shown in Figure 4. Therefore,

$$\text{W} = [low = limit, high = limit, size = (fan_{in}, fan_{out})].$$

The weights of the network are updated every iteration using an input batch size of 4. The optimization algorithm for optimizing weights is "Adam". Thus, the feature maps of these 1-4 convo layers are 16, 32, 64, and 124, respectively. These feature maps have resulted after a filter or kernel is applied to convolve an image. In each of the convolutional operation, a filter or kernel of size 3∗3 is applied with no zero paddings. We have used the Leaky ReLU activation function in our whole architecture shown in Figure 4, which is defined as per equation, $\text{F}(\text{x})\,\{x \text{ } if \text{ } x > 0 \text{ } otherwise \text{ } 0.01x$.

In this proposed study, we use 0.05x instead of 0.01x. To decrease spatial measurements of the input, we utilized the max-pool with a 2×2 window size. The fully connected layer essentially takes an info volume and outputs an n-dimensional vector where n is the number of classes that the program needs to browse. The last output equivalent to class labels generated by these FC layers. The fundamental convo layer implements filtration to the info images of 240 x 240 x 1. As shown in Figure 4, the output is taken from the 1st convo layer with all the filtration from the pooling layers sent to the 2nd convo layer as info and separated with $119 \times 119 \times 16$ measurements. Essentially, the convolutional yield from the subsequent layer is decreased through the pooling layer and is associated with the bit size $58 \times 58 \times 32$ in the 3rd convo layer. The 4th convo layer includes 124 number of feature maps of $28 \times 28 \times 64$ dimensions. Besides with SoftMax activation function, there are a total of 1024 neurons in this FC layer. The final layer is the classification layer. This layer uses SoftMax layers returned probabilities to each input to authorize any of the manually privileged classes and calculate the loss. The learning rate for CNN is 0.0001, the number of epochs is 30, and the kernel size is 3∗3. The complete details of network architecture are provided in Figure 4.

B. GAIT RECOGNITION UNDER UNKNOWN COVARIATE CONDITION

1) REGION OF INTEREST (ROI) EXTRACTION

The covariate conditions are difficult to handle which makes strict testing for gait recognition under unknown covariate conditions extremely difficult. To handle this issue, we extracted 2 ROIs from each GEI image to remove the regions with covariate (bags, coats, etc.) as shown in Figure 3, each GEI image has two ROIs. The reason to choose two ROIs is based on the regions that are least affected by clothing and carrying conditions. The part of the human body which is occluded by clothing or bag is removed in order to choose only the discriminative features.

2) FEATURE EXTRACTION

As shown in Figure 3 after extraction of ROIs, features are extracted using three different methods that are LBP, HOG, and Haralick Texture. The features returned by all these three methods from both ROIs are concatenated to a single feature vector of a GEI image.

a: LOCAL BINARY PATTERNS (LBP)

The overall texture information of the image including the spatial distribution is important, but the local texture may contain important information that is extracted using LBP. The technique is widely used and considered an efficient technique for denoting local patterns. LBP tags the pixels so that it can identify eight neighborhood pixels with respect to the center value of the image window. Based on the threshold value these pixels are assigned a binary number. In our case, the central pixel of each ROI of GEI image is compared with the neighboring pixels by using the following LBP equation as shown in Figure 3,

$$LBP_{(P,R)} = \sum_{p=0}^{p-1} S\left(g_p - g_c\right) 2^p. \quad (3)$$

In the above S(z) is the thresholding function, g_c and g_p are the grey level values of the center and its neighbor's pixels respectively. *P* is the total number of neighbors whereas *R* is the radius of the neighborhood.

b: HISTOGRAM OF ORIENTED GRADIENTS

It is observed recently that the performance of appearance-based methods for gait recognition techniques can be improved by applying HOG. HOG is a technique that portrays the direction of intensity gradients and it provides global descriptors. The following equation is used to compute the 1st order of gradients which are applied to extract horizontal and vertical magnitudes for each ROI of GEI image as shown in Figure 3,

$$F\_xdir = [\text{-}1\ 0\ 1]\, f\_y\ dir\ [\text{-}1\ 0\ 1]^T. \quad (4)$$

The combination of these horizontal and vertical gradient images is used to obtain gradient magnitude and orientation. Based on pixel intensity in the gradient orientation, a bin is selected, whereas the pixel intensity in gradient magnitude serves as the basis of the vote. This vote is cast by every pixel of the ROI to compose HOG. A histogram of gradients' direction is calculated for every pixel for an ROI. Their overall concentration is denoted as the HOG descriptor. To account for illumination and contrast the values of each ROI are normalized locally. This way the HOG descriptors are created for every ROI. This research work used HOG descriptors to characterize the shape and appearance of the subjects based on the distribution of local intensity gradients and directions.

c: HARALICK TEXTURE DESCRIPTOR

Haralick features consist of 14 statistical entities that are used for indicating certain texture properties from *P*. These are extracted and calculated at four directions by computing at 0°, 45°, 90°, and 135° using the GLCM based method [24]. The following equation is used for the calculation of px(g) and py(v) where *x* and *y* are the columns and row coordinates respectively of an entry in the co-occurrence matrix for each ROI of GEI image as shown in Figure 3.

$$P = \frac{p(g, v)}{\sum_{g=1}^{Ng} \sum_{v=1}^{Ng} p(g, v)} \quad (5)$$

$$Px(g) = \sum_{v=1}^{Ng} p(g, v) \quad \text{and } Py(v) = \sum_{g=1}^{Ng} p(g, v) \quad (6)$$

Moreover, Px+y(i) which is the probability of co-occurrence matrix coordinates sum to x+y is done through the following equation in the Haralick feature extraction method shown in Figure 3.

$$P_{x+y}(r) = \sum_{g=1}^{Ng} \sum_{v=1}^{Ng} p\,(g, v), \quad \text{where}$$
$$r = g + v \text{ with } r = 2, 3.....2N_g \quad (7)$$

We have considered only sum variance (fsv) in this research work. For calculation of fsv, one needs to compute (fsa) which is the sum average Haralick texture descriptor. This sum average of each ROI is computed by the following equation as shown in Figure 3.

$$f_{sa} = \sum_{r=2}^{2Ng} r p_{x+y}(r). \quad (8)$$

Finally, the sum variance is calculated as,

$$f_{sv} = \sum_{r=2}^{2Ng} (r - f_{sa})^2 p_{x+y}(r). \quad (9)$$

3) FEATURE REDUCTION USING FISHER LINEAR DISCRIMINANT ANALYSIS

The process of selecting the most discriminant features is known as feature selection. The success of any machine learning method is dependent on the selection of the most discriminant features. We have incorporated the dimensionality reduction method for this purpose. Therefore, the single feature vector of a very higher dimension of each GEI image is passed through for feature reduction. This dimensionality technique not only selects the most discriminant features it also reduces the dimensions of feature space. In the proposed approach we have incorporated FLDA for dimensionality reduction. FLDA is a supervised dimensionality reduction algorithm that uses class labels for the identification of most discriminant features. On the contrary, the unsupervised dimensionality reduction techniques such as PCA selects only those features which suit class labels. The goal of FLDA is a conversion from high dimension data to lower dimension data through the calculation of scattered matrices between and within-class labels. A transform matrix FLDA for the reduction of features of each subject can be obtained through the following Equation,

$$FLDA = argmax_w \frac{|W^T S_B W|}{|W^T S_W W|}. \quad (10)$$

4) CLASSIFICATION

After feature extraction from each GEI image, three different classifiers; Random Forest, Support Vector Machine, and Multilayer Perceptron were used for gait recognition under covariate conditions as shown in Figure 4 [25].

a: RANDOM FOREST

Random forest is one of the popular and supervised algorithms used for both classification and regression purposes. Its performance is very good as compared with other machine learning classifiers. It uses the ensemble technique by creating many decision trees on different data samples and finds the best solution by getting predictions of each decision tree. For classification purposes, it uses two popular techniques, Bagging, and random feature selection. In bagging, it takes bootstrap samples from the training data and then builds the trees. For each random tree, a process of top-down induction is followed to favor the diverseness of the ensemble process, and then by majority voting, a prediction is made. A part of the original features is taken to design each tree i-e n≪N where n is the subset of the complete feature set with size N. Later on, a tree is built by splitting these features randomly at each node. Each tree is of full-depth or the depth as required by the problem, and once the tree is built then no pruning process is followed. Then, in the end, a classification is made by doing voting among predictions of different trees. So, after the extraction of subject features from three different methods is given to a random forest for classification as shown in Figure 4. This proposed work uses default settings for this paper.

b: MULTILAYER PERCEPTRON

Artificial neural networks are machine learning classifiers that are designed to mimic the human brain. They have a wider range of applications such as pattern recognition, classification, and forecasting. Its architecture is formed by making connections among different artificial neurons called units or nodes. Each neuron carries some information in the network. An artificial neuron model receives a vector of $X = (x1, x2 \ldots, xn)$ of I input signals from an environment, or any other artificial neurons followed by some computation and activation functions to produce the results. They are categorized into a single layer and multilayer perceptron.

A single-layer perceptron has only input that is connected directly to the output layer while a multilayer perceptron has input, output, and one or more hidden layers. A multi-layer perceptron is a supervised machine learning algorithm, and it learns by adjusting the connection weights after calculating the error between model output and the expected result. The training procedure of the classifier continues until there exists a difference between an expected output and model output and it stops when the error rate between model output and the desired output is minimum or zero. This minimal difference shows that models learn a good mapping between input and desired output. Further, they are data-driven self-adaptive methods and can model any real-world problem. In our proposed work, we used a multi-layer perceptron-based ANN, as a classifier to classify different subjects from the perspective of gait recognition as shown in Figure 4. The learning rate used here is 0.0001, 2 hidden layers with 50 epochs.

c: SUPPORT VECTOR MACHINE (SVM)

Support vector machines are one of the other popular algorithms used for both classification and regression challenges. However, in most cases, it is used for classification purposes. The classification of data points in the dataset is done by finding a hyperplane in an N-dimensional space where N is the number of features. The SVM focuses on finding a hyperplane (an optimal hyperplane) that maximize the margins defined by support vectors where the margin is simply the distance between support vectors. Support vectors are essential training tuples that influence the orientation and position of the hyperplane. The equation for hyperplane as the set of points x satisfying for separating each subject,

$$f(x) = w.x + b = 0. \tag{11}$$

Here $W = \{w1, w2 \ldots wn\}$ is a weight vector and b are a scaler (bias). SVM can easily work with the input space of high dimensional. For a non-linear dataset, in which the data points are not linearly separable, the SVM needs a kernel function to map the original data to a higher dimension so that it can be linearly separable. There are many kernel functions with each have different performance on different types of data which includes linear, polynomial, Gaussian kernel. In our proposed work, for the feature vector X of every subject in the dataset, a linear kernel which is $K(xi, xj) = xiTxj$ is used. It involves mapping of the form,

$$\Phi : \mathrm{x} \rightarrow \varphi(\mathrm{x}), \tag{12}$$

where $\varphi(x)$ is x itself and K denotes the linear kernel function. Furthermore, in multi-class classification, it uses one-vs-all and one-vs-one strategy. In this proposed work, we used an SVM algorithm with a linear kernel function and a one-vs-one strategy to get our required results on gait recognition as shown in Figure 4. Here we used a linear kernel with a value of C = 1.0.

IV. EXPERIMENTAL SETUP AND RESULTS

A. DATASET

We consider the two datasets CASIA and OU-ISIR dataset in this research. The CASIA Gait Dataset [26] is provided by the Chinese Academy of Sciences (CASIA). It is divided into three parts, CASIA-A Gait Dataset, CASIA-B Gait Dataset, and CASIA-C. Similarly, OU-ISIR Treadmill Dataset [27] is an indoor gait dataset divided into two parts one part is focusing on speed variation called Treadmill dataset A-speed variation, and the other is part focuses on clothing variations called Treadmill Dataset B. For simple clothing and carrying conditions we consider, CASIA-B Gait Dataset consists of 124 subjects. Each subject has 6 normal walk sequences, two sequences with a bag, and two sequences with a coat. So, a total there are 10 sequences are available for each person. The other dataset is the OU-ISIR Treadmill dataset B consists of 68 subjects from a side view with 32 clothing variations. The list of clothing combinations is shown in Table 1. For speed invariant gait recognition, we consider CASIA-C Gait Dataset consists of 153 subjects.

TABLE 1. Clothing variations taken from OUR-ISIR dataset

Variation	Details of Covariate
Type 09	RP: Regular pants, FS: Full shirt
Type A	RP: Regular pants, PK: Parker
Type B	RP: Regular pants, LC: Long Coat, CS: Casquette cap
Type C	RP: Regular pants, DJ: Down jacket, MF: Muffler
Type 02	RP: Regular pants, HS: Half shirt

Each subject has 4 normal speed walk sequences, 2 slow walk sequences, 2 fast walk sequences, and 2 normal speed walk sequences with Bag. So, there are 10 sequences are available for each subject. All these videos are captured at night by the infrared (thermal) camera. All subjects are walking from left to right. The other dataset is OU-ISIR Treadmill dataset A consists of 34 subjects with 9 different speed variations varying between 2km/h and 10km/h with a 1Km/h interval. The subjects are walked between 2km/h to 7km/h and ran (or jogged) between 8km/h to 10km/h.

B. RESULTS AND DISCUSSION

All the experiments were carried out with Python on AMD processor A8-7410 APU with AMD Radeon R5 Graphics with 8GB RAM. We have used accuracy as an evaluation metric in this research work [25], [28]–[31]. The proposed methods are evaluated over 2 datasets under different cooperative and strict covariate conditions. The results are presented in two sections: gait recognition for cooperative persons and gait recognition under strict covariate conditions.

1) GAIT RECOGNITION RESULTS FOR KNOWN COOVARIATE CONDITIONS

This section presents results for gait recognition with cooperative persons (known covariates) which means no covariate conditions are used. In this research, the covariate conditions are only those where the gallery set is different than probe sets. The results presented in Table 2 show the gallery and probe set and either gallery set is the same as the probe set or the probe set is also a part of the gallery set. Table 2 shows the results for clothing and speed cooperative conditions from CASIA datasets. It is important here to mention that speed is not considered as a covariate condition and only bags and clothing variations are considered as strict covariates. This is because the shape of humans is not much changed due to speed variations.

TABLE 2. Results for CASIA dataset gait recognition under normal conditions

	CASIA		
	Gallery	Probe	Accuracy
1	Normal	Normal	97.98%
2	Normal + Coats	Coats	95.16%
3	Normal + Bags	Bags	91.10%
4	Normal + Coats	Normal + Coats	97.50%
5	Normal + Coats	Normal	100.00%
6	Normal + Bags	Normal	100.00%
7	Normal + Bags	Normal + Bags	95.50%
	Speed		
8	Normal Speed	Fast Speed	90.20%
9	Fast Speed	Normal Speed	83.66%
10	Normal Speed	Slow Speed	87.58%
11	Slow Speed	Normal Speed	85.62%
12	Normal Speed	Normal Speed	100%
13	Normal + Fast	Normal + Fast	100%
14	Normal + Slow	Normal + Slow	99.60%

The GEI images are directly given to the CNN algorithm and very good accuracies are achieved. This shows if no strict covariate conditions are considered then simple CNN is powerful enough to give satisfactory performance. This strengthens the argument that GEI performs well for gait recognition under normal conditions. However, the performance drop for covariate conditions is too high as it has been widely reported. Table 3 presents the results for the ISIR dataset for speed variations.

TABLE 3. Results for ISIR dataset gait recognition under normal conditions

	ISIR Speed		
	Gallery	Prob	Accuracy
1	2km,3km,4km,5km,6km,7km,8km,9km,10km	2km	100%
2	2km,3km,4km,5km,6km,7km,8km,9km,10km	3km	100%
3	2km,3km,4km,5km,6km,7km,8km,9km,10km	4km	97%
5	2km,3km,4km,5km,6km,7km,8km,9km,10km	5km	100.00%
6	2km,3km,4km,5km,6km,7km,8km,9km,10km	6km	100%
7	2km,3km,4km,5km,6km,7km,8km,9km,10km	7km	100%
8	2km,3km,4km,5km,6km,7km,8km,9km,10km	8km	100%
9	2km,3km,4km,5km,6km,7km,8km,9km,10km	9km	100%
10	2km,3km,4km,5km,6km,7km,8km,9km,10km	10km	100%
11	2km	3KM	100%
12	4km	5KM	88.23%
13	10km	8km	97.05%
14	7km	6km	91.17%
15	5km	7km	67.64%
16	5km	4km	94.11%

The experiment used just a part of the full dataset shows the effect of CNN when the covariate is not strictly followed. Here it is pertinent to mention that number of samples for clothing variations present in the ISIR dataset is low which makes it difficult to use it for training as well as for testing. Therefore, some experiments are only carried out for speed variations. The gallery set has different variations of speed present, but the probe set is also a part of the gallery set which makes it an unknown covariate experiment. The results are extremely good from experiment 1 to 10. The next 6 experiments are to make the experimental setup consistent with Table 2 where the gallery and probe sets are different for speed. The results show that overall results are very good here too.

Table 4 shows a comparative analysis of our proposed method with existing literature under no covariate conditions. The results prove that our method performs better than existing work and the important conclusion can be made that simple GEI with deep learning is enough to handle non-covariate conditions.

2) RESULTS FOR GAIT RECOGNITION FOR UNKNOWN COVARIATE CONDITIONS

This section presents results for covariate conditions where it is strictly maintained that the gallery and probe sets are not overlapped. The simplest approach is adopted to overcome the covariate condition problems which is to extract only the relevant and important ROIs from the GEI. This enables us to only focus on the common parts of the GEI of both gallery

TABLE 4. A comparative analysis of the proposed Deep learning Method with existing work under no covariate conditions

Authors	Technique	Covariate	Average Accuracy
Ahmed R. Hawas, et al. [32]	Deep Learning	No	93.5%
Rijun Liao et al. [33]	Deep Learning	No	63.78%
Munif Alotaibi et al. [34]	Deep Learning	No	90.43%
Xiaofang Wu et al. [35]	Deep Learning	No	95.96%
Proposed Method	Deep Learning	No	96.75%

TABLE 5. Experimental setup for covariate gait recognition

CASIA	
Gallery	Probe
Normal	Coats
Normal	Bags
ISIR	
Type09+TypeC	Type A
Type09+TypeC	Type B
Type09+TypeC	Type 02

TABLE 6. The highest values of each feature against both experimental conditions from casia for all classifiers

	CASIA								
	HOG			LBP			Haralick		
	SVM	RF	MLP	SVM	RF	MLP	SVM	RF	MLP
Normal vs Bags	65.72	47.58	67.33	67.74	54.83	64.90	61.29	56.45	54.83
Normal vs Coats	86.69	74.19	87.09	32.25	24.59	36.69	28.22	28.62	29.83

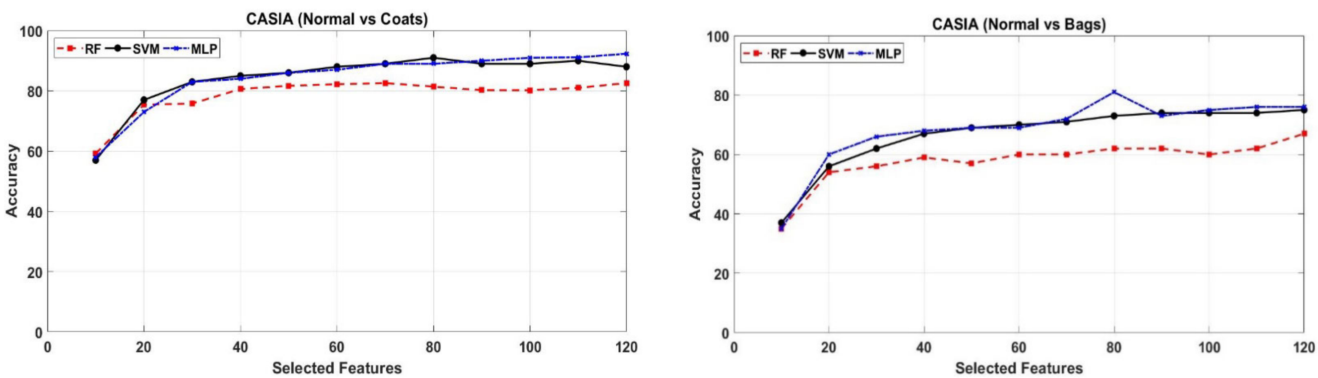

FIGURE 5. The feature-wise accuracy for CASIA Dataset under Covariate Conditions (Coats and Bags) using RF, SVM and MLP.

and probe sets. The details of the experiment are presented in Table 5. Table 5 shows that for the CASIA dataset, training is only performed on Normal GEIs while testing is carried out on the bag and coat GEIs separately. For the ISIR dataset, Type09 and Type C are used for training while testing is performed on Type A, Type B, and Type 02 separately. This experimental setup is used to ensure strict unknown covariate conditions.

Three classifiers were used to evaluate the performance of the features extracted. These classifiers include Random Forest, Multilayer Perceptron, and SVM. These classifiers are used because of their generalizability to different high dimensional data. The number of extracted features from ROIs is too high. Therefore, we applied Fisher linear discriminant analysis to then only used 120 features for the CASIA dataset and 60 features for ISIR experiments. In CASIA experiments, the classifiers were trained over the ROIs of normal persons and tested under covariate conditions of bags and coats. In the first experiment, where normal sequences are used for training and sequences with a person wearing a coat is presented in Figure 5. All three classifiers are trained and tested for 120 features and results show the MLP and SVM

TABLE 7. The highest values of each feature against all experimental conditions from ISIR for all classifiers

	ISIR								
	HOG			LBP			Haralick		
	SVM	RF	MLP	SVM	RF	MLP	SVM	RF	MLP
Type9+TypeC vs Type A	85.29	55.88	80.88	44.17	26.47	44.11	66.17	47.05	67.64
Type9+TypeC vs Type B	91.17	58.82	85.29	38.23	27.94	42.64	73.52	58.82	69.11
Type9+TypeC vs Type 02	42.30	58.82	69.23	34.61	26.92	34.61	42.36	30.76	50.0

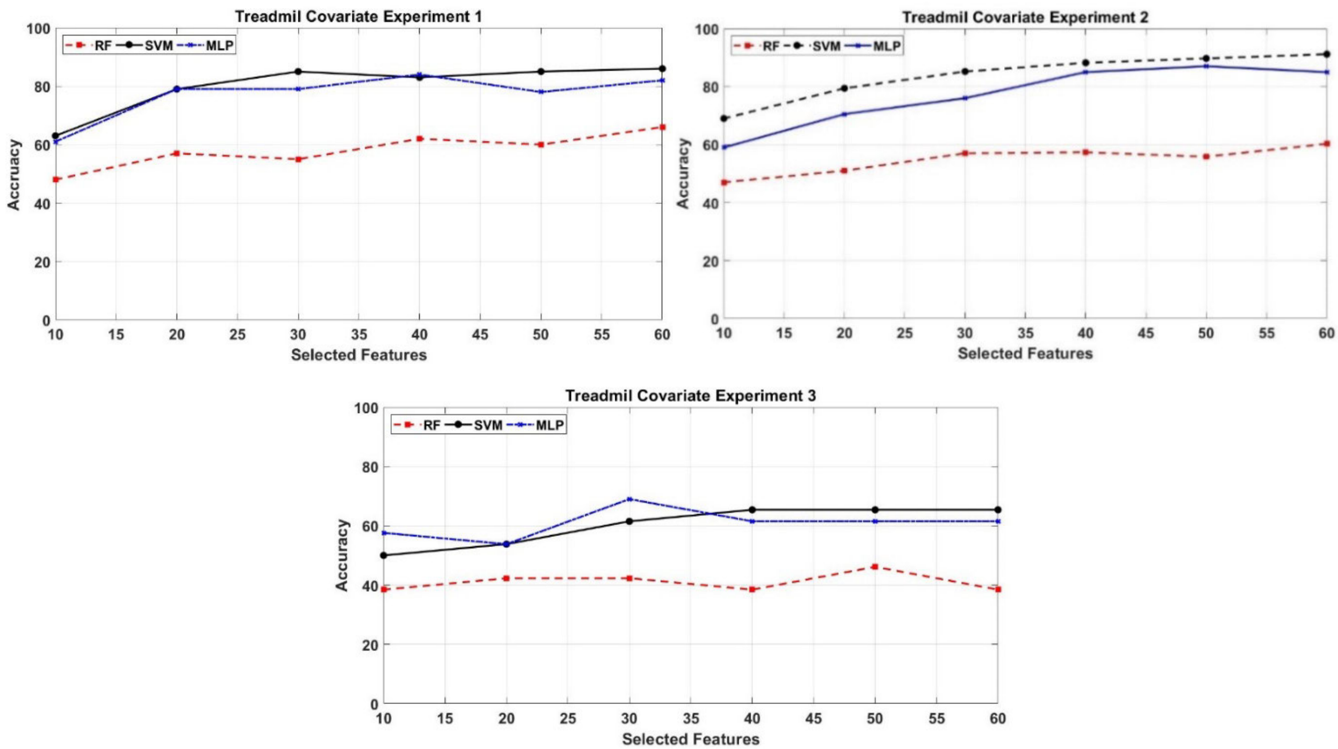

FIGURE 6. The feature-wise accuracy for ISIR Dataset Covariate Experiments 1,2 and 3 (from left to right and top to bottom) using RF, SVM and MLP.

perform better than RF and show an almost similar pattern for all number features. The results for the highest values of individual features against each experiment for all classifiers are presented in Table 6. This shows that the performance of the individual features is not that good as compared to when these features are combined.

The highest accuracy achieved for the experiment when a combined feature vector is used is 92% with 120 features using MLP which is very good considering the training was only done on normal ROI sequences. The second experiment was carried out for persons carrying bags. Here, again the normal GEI's ROIs were used for the training. An almost similar pattern of results was produced where MLP and SVM performed better than RF. The best results were achieved by 81% with MLP with 80 features as shown in Figure 5.

In our next experiment, we evaluated the performance of our proposed ROI based feature extraction technique with covariate conditions as shown in Table 6. The method was trained over Type 09 and Type C gallery set for all experiments and Type A, Type B, and Type 02 is used as probe set separately. The results for the highest values of individual features against each experiment for all classifiers are presented in Table 7. This shows that the performance of the individual features is not that good as compared to when these features are combined. Then three classifiers are used for training and testing over 60 combined features and accuracies are reported in Figure 6. The experiments show that SVM and MLP performed better than RF.

The highest accuracies achieved are 86%, 91.2%, and 69% for experiments 1, 2, and 3, respectively. The first two best accuracies were achieved by SVM and MLP performed best for experiment 3. We have compared our results with the methods designed for strict covariate conditions. The point we want to establish here is that if the gallery set includes

TABLE 8. Comparative analysis with state of the art work under covariate conditions

Sr. No.	Method	Normal	Bags	Coats	Average Accuracy	Covariate
1	Bashir et al. [12]	100.0%	78.3%	44.4%	74.2%	Yes
2	Rokanujjaman, M. [36]	97.61%	83.87 %	51.61%	77.69%	Yes
3	Gupta, S.K [37]	NA	86.2%	61.4%	73.8%	Yes
4	Hawas, A.R [32]	97.6%	45.3%	49.6%	64.1%	Yes
5	Shiqi Yu et al. [20]	95.97%	65.32%	42.74%	68.01%	Yes
6	Lingxiang Yao et al. [38]	NA	NA	38%	38%	Yes
7	Jingran Su et al. [39]	93.2%	72.8%	59.1%	75.03%	Yes
8	Guoheng Huang at al. [40]	91.9%	80.3%	72.3%	81.50%	Yes
9	Proposed Method	97.98%	81%	92%	90.32%	Yes

any of the prob set samples like a bag or coat sequence then good results can be easily achieved. We proved this in our first experiment, where we only used GEI with CNN and achieved very good results.

Furthermore, the comparison is carried out with techniques that only use the available data without using augmentation data to improve the results. It is evident from Table 8 that we were able to achieve very good mean results as compared to the latest and classical methods. The results show that we have achieved the best average results for unknown covariates. The comparison is carried only for strict unknow covariate conditions and with a single view (90^o). Furthermore, the comparison is only carried out for the CASIA dataset because this is usually considered as a benchmark dataset for covariate conditions.

In addition to this, we performed a limited number of experiments on the ISIR dataset just to show the performance of the proposed method. Therefore, the comparative analysis is not carried out for ISIR with existing work. However, the proposed method can be extended to apply to all experiments. This approach is efficient as compared to the scenarios where full GEI or its variants are directly provided to deep learning architectures to handle the covariate conditions. In that case, the relevant patterns (bags, jackets, hats, etc.) are also used by the deep learning architectures and become difficult for it to handle efficiently. In our proposed case, the proposed scenario is more realistic where unique covariate invariant features are selected and passed to CNN for learning which makes it easier to handle covariate conditions. The proposed architecture can be extended for real-time systems.

V. CONCLUSION

Gait recognition without the subject's cooperation remains one of the most challenging research areas in the field. The covariate conditions, including clothing and speed variations, are still difficult to handle in realistic experimental setups. The existing solutions perform poorly when subject cooperation is not possible, and there are changes in covariate conditions, making them unsuitable to deploy for practical purposes. The emergence of deep learning approaches has made computer vision tasks easier. However, there are certain scenarios where pre-processed data can further improve the performance of these deep learning methods. In this work, we have developed a gait recognition method that extracts features from ROIs of the gallery and probe gait GEI sequences. The unique covariate condition invariant feature-based gait sequences used with RF, SVM, and MLP performed very well for covariate conditions. The results demonstrate the overall superiority of our approach over the existing approaches. It is pertinent to mention that the feature selection method deals only with changes in different covariate conditions and has no effect on gait itself.

The proposed method handles covariate conditions by selecting the discriminative covariate invariant features and removes the occluded part of the body. The aim is to remove the body part, which is affected by covariate conditions, especially for bags and coats. The same technique can be used on other datasets with similar covariate conditions. The proposed method can be used to handle dynamic covariates like putting on a coat and taking out a coat as the occluded and affected part of the body remains the same for these conditions. The ROIs can still be used for unique covariate invariant features. In future, the ROI selection process can be improved for automatic candidate selection. The algorithm can be extended to design zero-shot learning-based algorithms to work in real-time data. The latest zero-shot training-based algorithms and proposed discriminative feature learning can be combined to handle covariate conditions in real-time.

ACKNOWLEDGMENT

(Maryam Bukhari and Khalid Bashir Bajwa are co-first authors.)

REFERENCES

[1] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "A survey of behavioral biometric gait recognition: Current success and future perspectives," *Arch. Comput. Methods Eng.*, to be published, doi: 10.1007/s11831-019-09375-3.

[2] H. M. Alawar, H. Ugail, and M. Kamala, "The relationship between 2D static features and 2D dynamic features used in gait recognition," *Proc. SPIE*, vol. 8712, May 2013, Art. no. 87120I.

[3] N. Lynnerup and P. K. Larsen, "Gait as evidence," *IET Biometrics*, vol. 3, no. 2, pp. 47–54, Jun. 2014.
[4] H. Iwama, D. Muramatsu, and Y. Makihara, "Gait verification system for criminal investigation," *Inf. Media Technol.*, vol. 8, no. 4, pp. 1187–1199, 2013.
[5] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Joint intensity transformer network for gait recognition robust against clothing and carrying status," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3102–3115, Dec. 2019.
[6] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
[7] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
[8] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 38, no. 2, pp. 342–352, Apr. 2008.
[9] N. V. Boulgouris and Z. X. Chi, "Human gait recognition based on matching of body components," *Pattern Recognit.*, vol. 40, no. 6, pp. 1763–1770, Jun. 2007.
[10] Y. Guan, C.-T. Li, and F. Roli, "On reducing the effect of covariate factors in gait recognition: A classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1521–1528, Jul. 2015.
[11] K. Bashir and T. S. X. Gong, "Gait recognition using gait entropy image," in *Proc. 3rd Int. Conf. Imag. Crime Detection Prevention*. London, U.K.: IET, 2009, pp. 1–6. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/ic.2009.0230
[12] K. Bashir, T. Xiang, and S. Gong, "Gait recognition without subject cooperation," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 2052–2060, 2010.
[13] X. Li, Y. Makihara, C. Xu, D. Muramatsu, and Y. Yagi, "Gait energy response function for clothing-invariant gait recognition," in *Proc. Asian Conf. Comput. Vis.* Taipei, Taiwan: Springer, 2016, pp. 257–272.
[14] Y. Makihara, A. Suzuki, D. Muramatsu, X. Li, and Y. Yagi, "Joint intensity and spatial metric learning for robust gait recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5705–5715.
[15] K. Sundararajan and D. L. Woodard, "Deep learning for biometrics: A survey," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–34, 2018.
[16] F. Battistone and A. Petrosino, "TGLSTM: A time based graph deep learning approach to gait recognition," *Pattern Recognit. Lett.*, vol. 126, pp. 132–138, Sep. 2019.
[17] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GEINet: View-invariant gait recognition using a convolutional neural network," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.
[18] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017.
[19] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2708–2719, Sep. 2019.
[20] S. Yu *et al.*, "Invariant feature extraction for gait recognition using only one uniform model," *Neurocomputing*, vol. 239, pp. 81–93, 2017.
[21] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh, "GaitGAN: Invariant gait feature extraction using generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 30–37.
[22] F. M. Castro, M. J. Marín-Jiménez, N. G. Mata, and R. Muñoz-Salinas, "Fisher motion descriptor for multiview gait recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 1, Jan. 2017, Art. no. 1756002.
[23] T. Liu, X. Ye, and B. Sun, "Clothing and carrying invariant gait-based gender recognition," *Proc. SPIE*, vol. 10836, Oct. 2018, Art. no. 108360X.
[24] R. Anusha and C. Jaidhar, "Human gait recognition based on histogram of oriented gradients and Haralick texture descriptor," *Multimedia Tools Appl.*, vol. 79, pp. 8213–8234, Jan. 2020.
[25] A. Kalsoom, M. Maqsood, M. A. Ghazanfar, and F. Aadil, "A dimensionality reduction-based efficient software fault prediction using Fisher linear discriminant analysis (FLDA)," *J. Supercomput.*, vol. 74, no. 9, pp. 4568–4602, 2018.
[26] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006, pp. 441–444.
[27] Y. Makihara, H. Mannami, A. Tsuji, M. A. Hossain, K. Sugiura, A. Mori, and Y. Yagi, "The OU-ISIR gait database comprising the treadmill dataset," *IPSJ Trans. Comput. Vis. Appl.*, vol. 4, pp. 53–62, 2012.
[28] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim, and K. Muhammad, "A local and global event sentiment based efficient stock exchange forecasting using deep learning," *Int. J. Inf. Manage.*, vol. 50, pp. 432–451, Feb. 2020.
[29] M. Yasir, M. Y. Durrani, S. Afzal, M. Maqsood, F. Aadil, I. Mehmood, and S. Rho, "An intelligent Event-Sentiment-Based daily foreign exchange rate forecasting system," *Appl. Sci.*, vol. 9, no. 15, p. 2980, Jul. 2019.
[30] F. Jabeen, M. Maqsood, M. A. Ghazanfar, F. Aadil, S. Khan, M. F. Khan, and I. Mehmood, "An IoT based efficient hybrid recommender system for cardiovascular disease," *Peer–Peer Netw. Appl.*, vol. 12, no. 5, pp. 1263–1276, Sep. 2019.
[31] S. Afzal, M. Maqsood, F. Nazir, U. Khan, and F. Aadil, "A data augmentation-based framework to handle class imbalance problem for Alzheimer's stage detection," *IEEE Access*, vol. 7, pp. 115528–115539, 2019.
[32] A. R. Hawas *et al.*, "Gait identification by convolutional neural networks and optical flow," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 25873–25888, 2019.
[33] R. Liao, S. Yu, W. An, and Y. Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107069.
[34] M. Alotaibi and A. Mahmood, "Improved gait recognition based on specialized deep convolutional neural network," *Comput. Vis. Image Understand.*, vol. 164, pp. 103–110, Nov. 2017.
[35] X. Wu, T. Yang, and Z. Xia, "Gait recognition based on densenet transfer learning," *Int. J. Sci., Environ.*, vol. 9, no. 1, pp. 1–14, 2020.
[36] M. Rokanujjaman *et al.*, "Effective part-based gait identification using frequency-domain gait entropy features," *Multimedia Tools Appl.*, vol. 74, no. 9, pp. 3099–3120, 2015.
[37] S. K. Gupta, G. M. Sultaniya, and P. Chattopadhyay, "An efficient descriptor for gait recognition using spatio-temporal cues," in *Emerging Technology in Modelling and Graphics*. Singapore: Springer, 2020, pp. 85–97.
[38] L. Yao *et al.*, "Robust gait recognition using hybrid descriptors based on skeleton gait energy image," *Pattern Recognit. Lett.*, 2019, doi: 10.1016/j.patrec.2019.05.012.
[39] J. Su, Y. Zhao, and X. Li, "Deep metric learning based on center-ranked loss for gait recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4077–4081.
[40] G. Huang *et al.*, "Flexible gait recognition based on flow regulation of local features between key frames," *IEEE Access*, vol. 8, pp. 75381–75392, 2020.

MARYAM BUKHARI is currently pursuing the bachelor's degree with COMSATS University Islamabad–Attock, Pakistan. Her research interests include machine learning and image processing.

KHALID BASHIR BAJWA received the B.Eng. degree in computer engineering from the National University of Engineering and Technology, Pakistan, in 2002, and the M.S. and Ph.D. degrees in computer science from the Queen Mary University of London, U.K., in 2007 and 2011, respectively. He has been an Active Researcher and an Associate Professor with the Faculty of Computer and Information Systems, Islamic University of Madinah. His research interests include computer vision, machine learning, and deep learning.

SAIRA GILLANI received the Ph.D. degree in information sciences from the Corvinus University of Budapest, Hungary. She joined the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2016. She has served as an Assistant Professor for Saudi Electronic University, Jeddah, Saudi Arabia. She is currently serving as a Senior Assistant Professor for Bahria University–Lahore, Pakistan. Previously, she worked as a Research Scholar with Corvinno, Technology Transfer Center of Information Technology and Services in Budapest, Hungary, and also worked as a Research Associate with the Center of Research in Networks and Telecom (CoReNet), CUST, Pakistan. Her research interests include data sciences, text mining, data mining, machine learning, vehicular networks, mobile edge computing, and the Internet of Things.

MUAZZAM MAQSOOD received the Ph.D. degree from UET Taxila, in 2017. He is currently serving as an Assistant Professor for COMSATS University Islamabad–Attock, Pakistan. His research interests include machine learning, recommender systems, and image processing.

MEHR YAHYA DURRANI graduated in information technology. He received the master's degree in computer science. He is currently pursuing the Ph.D. degree in pattern recognition. He is also serving as an Assistant Professor for COMSATS University Islamabad–Attock, Pakistan.

IRFAN MEHMOOD (Member, IEEE) is currently serving as a Senior Lecturer for the University of Bradford, U.K. His sustained contribution at various research and industry-collaborative projects gives him an extra edge to meet the current challenges faced in the field of multimedia analytics. Specifically, he has made significant contribution in the areas of video summarization, medical image analysis, visual surveillance, information mining, deep learning in industrial applications, and data encryption.

HASSAN UGAIL (Member, IEEE) received the B.Sc. degree (Hons.) in mathematics from King's College London, and the Ph.D. degree in geometric design from the School of Mathematics, University of Leeds. He is currently the Director of the Center for Visual Computing, Faculty of Engineering & Informatics, University of Bradford, U.K. His research interests include computer-based geometric and functional design, imaging, and machine learning.

SEUNGMIN RHO received the B.Sc. degree in computer science from Ajou University, South Korea, in 2001, and the M.Sc. and Ph.D. degrees in information and communication technology from the Graduate School of Information and Communication, Ajou University, in 2003 and 2008, respectively. Before, he joined the Computer Sciences Department, Ajou University, he spent two years in the industry. He visited the Multimedia Systems and Networking Laboratory, The University of Texas at Dallas, from 2003 to 2004. From 2008 to 2009, he was a Postdoctoral Research Fellow with the Computer Music Lab, School of Computer Science, Carnegie Mellon University. From 2009 to 2011, he was a Research Professor with the School of Electrical Engineering, Korea University. In 2012, he was an Assistant Professor with the Division of Information and Communication, Baekseok University. From 2013 to 2018, he was an Assistant Professor with the Department of Media Software, Sungkyul University. He is currently a Faculty of the Department of Software, Sejong University, South Korea. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management, and computational intelligence. He has published more than 180 articles in refereed journals and conference proceedings in these areas. He has been involved in more than 20 conferences and workshops as various chairs and more than 30 conferences/workshops as a program committee member.