Recognition of Previously Identified Individuals using Affine Moment Invariants as Feature Descriptors in Gait Sequences

Internship report submitted in partial fulfilment of the requirements for the degree of B.Tech. + M.Tech. in Electronics and Communication Engineering

by

Sandesh V Bharadwaj
(Roll No: ESD15I005)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING, KANCHEEPURAM

December 2019

Certificate

This is to certify that the Internship report titled "**Recognition of Previously Identified Individuals using Affine Moment Invariants as Feature Descriptors in Gait Sequences**" by **Sandesh V Bharadwaj** (ESD15I005) to the **Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram** for the partial fulfilment of "**B. Tech Electronics and Communication Engineering with Specialization in Design and Manufacturing + M. Tech Signal Processing and Communication System Design**", is a bonafide record of the research work done by him under my supervision from **May to October 2019**.

Smt. Soma Mitra

Senior Director
Advanced Signal Processing Group
Centre for Development of Advanced Computing (CDAC)
Kolkata

Abstract

Gait recognition is a biometric technology that identifies individuals in a video sequence by analysing their style of walking or limb movement. However, this identification is generally sensitive to appearance changes and conventional feature descriptors such as Gait Energy Image (GEI) lose some of the dynamic information in the gait sequence. Active Energy Image (AEI) focuses more on dynamic motion changes than GEI and is more suited to deal with appearance changes.

In this paper, we propose a new approach, which allows identifying people by analysing the dynamic motion variations and identifying people in spite of appearance changes, without using a database of predicted appearance changes. In the proposed method, the active energy image is calculated by averaging the difference frames of the silhouette sequence and divided into multiple areas. Affine moment invariants are extracted as gait features for each area. Next, matching weights for all areas are calculated based on the similarity between extracted features and those in the database. Finally, the subject is identified by the weighted combination of similarities in all areas. The CASIA-B Gait Database is used as the principal dataset for the experimental analysis.

Acknowledgements

The research internship at **Centre for Development of Advanced Computing (C-DAC), Kolkata** was a great chance for learning and professional development.

I take this opportunity to express my sincere thanks to **Smt. Soma Mitra** for providing me an opportunity to do the internship work at C-DAC, Kolkata.

I express my sincere thanks and gratitude to **Mr. Kunal Chanda**, who, as my guide and mentor, took a keen interest in my work and guided me all along till the completion of the internship.

I would also like to thank **Mr. Souvik Banik**, **Mr. Washef Ahmed** and **Mr. Sayantan Bhattacharya**, who guided me and supported me during the course of the internship.

This research internship is a big milestone in my career development. The intense and thorough research has exposed me to different aspects of the subject matter and cultivated an interest to pursue it further.

I would also like to thank **IIITDM Kancheepuram** for permitting me to pursue this internship.

Contents

Certificate i

Abstract ii

Acknowledgements iii

Contents iv

List of Figures vi

Abbreviations vii

1 Introduction 1
- 1.1 Gait Recognition . . . 1
- 1.2 Background . . . 2
  - 1.2.1 Model-based approach . . . 2
  - 1.2.2 Motion-based approach . . . 2
- 1.3 Feature Descriptors . . . 3
- 1.4 Literature Survey . . . 5

2 Proposed Methodology 32
- 2.1 Active Energy Image . . . 33
- 2.2 Affine Moment Invariants . . . 35
  - 2.2.1 Image moments and moment invariants . . . 35
- 2.3 Estimation of matching weights . . . 36
  - 2.3.1 Principal Component Analysis . . . 37

3 Experiment and Results 39
- 3.1 Implementation . . . 39
- 3.2 Experiment and Result . . . 40

4 Conclusions and Future Work 42
- 4.1 Conclusion . . . 42
- 4.2 Future Work . . . 43

Bibliography **44**

List of Figures

1.1 Literature Review of Current Gait-Based Person Identification Techniques . 31

2.1 Gait silhouette sequence of individual 33
2.2 Difference Images between consecutive significant frames of gait sequence . 33
2.3 Active Energy Image 34
2.4 Segmented AEI (6 segments) 34

3.1 Best Case Results (K=23, M=5) 41

Abbreviations

AEI	**A**ctive **E**nergy **I**mage
GEI	**G**ait **E**nergy **I**mage
AMI	**A**ffine **M**oment **I**nvariant
PCA	**P**rincipal **C**omponent **A**nalysis
CCR	**C**orrect **C**lassification **R**ate

Chapter 1

Introduction

In the modern world, reliable recognition of individuals has become a fundamental requirement in various real-time applications such as forensics, international travel and surveillance. The motivations for using biometrics are diverse and often overlap. They include improving the convenience and efficiency of routine access transactions, reducing fraud, and enhancing public safety and national security.

1.1 Gait Recognition

Gait recognition is a biometric recognition technique that recognizes individuals based on their walk cycle i.e. gait, and has been a topic of continued interest for person identification due to the following reasons:

- First, gait recognition can be performed with low-resolution videos with relatively simple instrumentation.
- Second, gait recognition can work well remotely and perform unobtrusive identification, especially under conditions of low visibility.
- Third, gait biometric overcomes most of the limitations that other biometric identifiers suffer from such as face, fingerprint and iris recognition which have certain hardware requirements that add to the cost of the system.

- Finally, gait features are typically difficult to impersonate or change, making them somewhat robust to appearance changes.

1.2 Background

In the last two decades, significant efforts have been made to develop robust algorithms that can enable gait-based person recognition on real-time data. Modern gait recognition methods can be classified into two major groups, namely model-based and motion-based methods.

1.2.1 Model-based approach

In model-based methods, the human body structure or motion is described using a mathematical model and the image features are extracted by measuring the structural components of models or by the motion trajectories of the body parts. These models are often simplified based on justifiable assumptions such as the system only pathologically normal gait.

Such systems normally consists of gait capture, a model(s), a feature extraction scheme, a gait signature and a classifier. The model can be 2-dimensional or 3-dimensional structural (or shape) model and motion model that lays the foundation for the extraction and tracking of a moving person. The main advantages of this approach is that it can reliably handle occlusion (especially self-occlusion), noise, scale and rotation well. However, system effects such as viewpoint invariance and effects of physiological, psychological and environmental changes are major limitations in implementing a model-based recognition system.

1.2.2 Motion-based approach

Motion-based methods consider the human gait cycle as a sequence of images and extract binary silhouettes from these images. Motion-based approaches are insensitive to the quality of images and have the added advantage of low computational cost compared to model-based approaches.

A baseline algorithm proposed by Sarkar et al. [11] uses silhouettes as features themselves, scaling and aligning them before use. Bobick and Davis [12] proposed the motion-energy image (MEI) and motion-history image (MHI) to convert the temporal silhouette sequence to a signal format. Han and Bhanu [13] used the idea of MEI to propose the Gait Energy Image (GEI) for individual recognition. GEI converts the spatio-temporal information of one walking cycle into a single 2D gait template, avoiding matching features in temporal sequences. GEI is comparatively robust to noise, but loses dynamic variations between successive frames.

Zhang, Zhao and Xiong [14] proposed an active energy image (AEI) method for gait recognition, in which the active regions of a gait sequence are extracted by calculating the difference between two adjacent silhouette images in a gait sequence. AEI focuses more on dynamic regions than GEI, and can alleviate the effect caused by low quality silhouettes. Current research on gait representation include Gait Entropy Image (GEnI), frequency-domain gait entropy (EnDFT), gait energy volume (GEV) etc., which focus more on dynamic areas and reducing view-dependence of traditional appearance-based techniques.

1.3 Feature Descriptors

A feature descriptor is a metric or a quantifiable value used to describe an image at a high level perspective. Features related to color,texture,shapes, color blobs, edges and other interest points are contained in an image. Feature descriptors need to be unique but repeatable as well, meaning that the same physical interest points in the image must be detectable under different viewing conditions. Descriptors need to be of proper dimensions as well; large descriptors will have higher computational costs, but small descriptors may discard useful information.

Feature descriptors are of the following types:

1. **Color Descriptors**:

- *Histograms*: It is the distribution of number of pixels of an image. The number of elements in a histogram relates to the number of bits in each pixel of an image.
- *Color Coherent Vector (CCV)* [22] : Each histogram bin is partitioned into two types; coherent type contains pixel value belonging to a large informally colored region.
- *Color Moments and Moment invariants* [23]: Color moments are shift-invariant measures that centralize color distribution in an image. Moment invariants are derived from these color moments, where only central moments are invariants themselves
- *Color Scale Invariant Feature Transform (SIFT) Descriptors* [24]: The local shape of a region is described by edge orientation histograms. SIFT descriptors are shift-invariant and normalized, due to which gradient magnitude changes have no effect. However, these descriptors are not invariant to light color changes, as the R, G and B channels are combined to form the intensity channel.

2. **Texture Descriptors**:

- *Gray level co-occurrence matrix (GLCM)*[25]: Used in motion estimation of images to extract second order statistical texture features.
- *Haralick Texture Feature*[26]: Used for image classification, these 13 features capture information about the patterns emerging in patterns of texture. These features are calculated by using co-occurrence matrix.

3. **Visual Descriptors**:

- *Visual color descriptors*
- *Visual texture descriptors* [28]
- *Visual shape descriptors* [27]

4. **Frequency Domain Descriptors**

1.4 Literature Survey

Citation	Date of Publication	Title of Paper	Defn of key terms and concepts	Objectives from Abstract and Conclusion (point-wise)	Research Methodology (include steps of Algorithm point-wise)	Summary of Research Results
[1]	April 23, 2018	**Person Identification from Partial Gait Cycle Using Fully Convolutional Neural Network**	• **Gait Recognition** - to recognize people by the way they walk • **FCN** - Fully Convolutional Neural Network • **GEI** – Gait Energy Image, obtained by averaging binary silhouette over one gait cycle • **ReLU** – Rectified Linear Unit, a piecewise activation function used in CNNs. • **IC-GEI** – Incomplete GEI • **RC-GEI** – Reconstructed Complete GEI • **TC-GEI** – True Complete GEI • **ITCNet** – Incomplete-to-	• To identify individuals from gait features when *few (or single) frame(s)* is available. • To build a fully convolutional neural network for GEI reconstruction from *incomplete gait cycle*; this is done by training several auto-encoders independently and then combing them into a model. • Fully reconstruct a true GEI from incomple	1. Incomplete GEI computed by averaging silhouette images. $GEI(x,y)=\frac{1}{N}\sum_{i=1}^{N} f_i(x,y)$ Where f(x, y) denotes binary value of pixel in position (x, y) at time t N is no. of frames in gait cycle. 2. Training for Incremental GEI reconstruction approach using 9 FCNs that each single FCN enhances the quality of input GEI. • Each FCN is trained on different types of GEI, but has the same architecture. • Structure of FCN – *encoder* (convolutional) part and *decoder* (deconvolutional) part. ○ ***Encoder*** – three convolutional layers. Each layer is followed by a **ReLU** (activation function), and a **pooling layer** (down-sampling unit). ○ Batch normalization and dropout is then used to prevent overfitting and increase rate of convergence. ○ ***Decoder*** – up-sampling layer, convolutional layer, ReLU layer, batch normalization and dropout. ○ *Up-sampling layer* is used to remove convolutional artifacts like checkerboard patterns, which occur if	1. Reconstruction: a. ***Increasing no. of frames*** of **IC-GEIs** generates **RC-GEIs** closer to **TC-GEIs**. However, upward trend becomes slower as 18/20f-GEIs are very close to true GEIs. b. **OULP** – Min. Accuracy – *89.35%* Max. Accuracy – *96.15%* c. **Casia-B** – Min. Accuracy – *92.27%* Max. Accuracy – *94.48%* 2. Recognition: a. Recognition between OULP and Caisa -B was compared using Euclidean distance as similarity metric between GEI samples. b. ITCNet improved identification performances for IC-GEIs, especially of those with fewer frames. c. As no. of frames increase, rank-1 and rank-5 identification gets closer to that of TC-GEIs. (CMC curve

			complete GEI Network • **ROC** – Receiver Operating Characteristic Curve, used for measuring performance of classification model, plotting true positive rate vs false positive rate. • **CMC** – Cumulative Matching Characteristic curve, metric used to measure performance of identification and recognition algorithms based on precision for each rank.	te GEI, despite starting frame and number of available frames.	transposed convolutional layer is **directly used** for the encoder output. ○ ***Sigmoid function*** is placed at output of decoder to get a *gray-scale output* image. • Input image – *64x64* pixels, • Convolution layer parameters – ○ Kernel size – *(4,4)* ○ Strides – *(1,1)* ○ Output Dimensions – *128, 64, 32* • Pooling Layer: ○ Kernel size and pooling strides – *(2,2)* each 3. ITCNet converter: • The trained auto-encoders from Step 2 are stacked together. • Input – any type of incomplete GEI • Output – complete GEI • Auto-encoder in layer i maps ***m***f-GEI to ***n***f-GEI, where ***n*** – ***m*** = ***T***/10, ***T*** = gait cycle length. • Weight initialization using *Gaussian distribution* with std. deviation σ_c $\sigma_c = \sqrt{\frac{2}{f_w * f_h * f_t * f_d}}$ Where f_w – spatial width f_h – spatial height f_t – time extent of filter f_d – no. of filters in convolutional layer • Mean Square Error (L) – Loss value between RC-GEI and complete GEI $L = \sum_{i=1}^{N} \|x - y\|^2$ Where x – RC-GEI y – original complete GEI • MSE minimized by using stochastic gradient descent, using Adam optimizer.	comparison) d. For *verification*, similarity between two GEIs is computed. If this score is less than or equal to threshold, this is regarded as positive pair (same subject). 3. Conclusion – ITCNet successfully reconstructed TC-GEI from IC-GEI regardless of starting frame and no. of available frames. Model improves recognition rate, particularly when only 0.1 part of gait cycle is available.

					○ Optimizer parameters – $B_1 = 0.8$, $\beta_2 = 0.99$, $\varepsilon = 10 \times 10^{-8}$ • Learning rate- optimal solution through quantitative testing together with weight decay and momentum parameter. ○ Initial value – 10^{-3}, decreases every 5 epochs by factor of 10 • Reconstruction accuracy – portion of errors ***<0.08*** between pixels in TC-GEI and RC-GEI • 50 epochs, batches of size 80 for training. • Dropout probability – 0.5 during training 4. Datasets used – a. *OULP* – OU-ISIR Large Population Dataset, section A used for gait recognition, gait data selected at *85°*. 2254 subjects for training set, 1000 subjects in validation and test sets. b. *Casia-B* – normal walking section under *90°* used, with 124 subjects. Each subject has 1 gallery set and 5 probe sets. 5. Evaluation Metrics – a. ***Reconstruction***: i. MSE value is computed. ii. SSIM – Structural Similarity Index,which is a decimal value between -1 and 1. If *identical*, output is 1. b. ***Recognition*** – i. For verification, **receiver operating characteristic** (ROC) curve and **equal error rates** (EER) ii. For identification, ***Rank-1/Rank-5*** metrics and ***cumulative matching characteristic*** (CMC) curve.	

[2]	Sept. 2018	**Appearance and Gait-Based Progressive Person Re-identification for Surveillance Systems**	• **Progressive search** – techniques that can be used • **Gait feature** • **LSTM** - Long short-term memory, special kind of RNN, capable of learning long-term dependencies. • **Siamese network** – neural network that contain two or more identical sub-networks. • **GoogLeNet (Inception)** – A 22-layer deep CNN model developed by Google, using a CNN inspired by LeNet and implemented an inception module, which used batch normalization, image distortions and several	• To implement a progressive person re-identification (**PROPRID**) approach to simultaneously improve timeliness and accuracy of identifying the target person • To utilize the appearance of the target person as a course filter to *reduce computational complexity* of the fine search	1. Person image sequences are taken as input, considered as progressive processes 2. **Appearance-based person coarse filtering** – Hair, clothes, skin and other visual information used to filter out dissimilar people in large-scale surveillance videos. ○ Low-level, mid-level and high-level appearance features are used as coarse filter ○ For low-level and mid-level features – CN, HOG, and LOMO descriptors ➢ CN is benchmark for person Re-ID on Market-1501 dataset, quantized by BOW model for fast color feature matching. ➢ HOG is used to characterize poses and appearances (shape feature) ➢ LOMO is used for stable texture representation against changes in the viewpoint. ○ For high level features, **GoogLeNet** is used as feature extractor. ➢ GoogLeNet used ImageNet as training data, cannot apply directly to person re-id. ➢ GoogLeNet model is fine-tuned using PRID-2011 and iLIDS-VID datasets. ○ All four types of features are integrated by distance-level fusion. 3. **Human silhouette extraction** done by **Mask R-CNN** to generate GEI on PRID-2011 and iLIDS-VID datasets. ○ Mask R-CNN is trained (fine-tuned) using **Casia-B** dataset. It cannot be used	1. Evaluations of Appearance-Based Person Re-Identification: ○ 5 approaches evaluated on PRID-2011 and iLIDS-VID: ➢ *CN* ➢ *HOG* ➢ *LOMO* ➢ *GoogLeNet:* 1024-D feature vector from last pooling layer obtained to represent semantic features. ➢ *FMLF*: Multi-level features combined into appearance features. *Fusion weights obtained on validation subset of training data*. ○ GoogLeNet Features are more discriminative and robust than CN, HOG and LOMO descriptors. ○ Multi-level fusion features (FMLF) achieve **highest accuracy**. 2. Evaluation of Progressive Person Re-Identification: ○ Two methods i.e. **FMLF** and **PROPRID** were compared on PRID-2011 and iLIDS-VID datasets with the following fusion weights: ➢ *FMLF*: PRID-2011 – CN = 0.1 HOG = 0.2 LOMO = 0.4

			very small convolutions to reduce no. of parameters. • **CN** – Color Name Descriptor, assigning names to colors for identification • **BOW** – Bag-of-words model, a vector of occurrence counts of local image features • **HOG** – Histogram of oriented gradients • **LOMO** – Local Maximal Occurrence, analyses the horizontal occurrence of local features and maximizes the occurrence for stable representation in viewpoint changes. • **Mask R-CNN** – Mask regional CNN, consisting		directly as it is trained on COCO dataset. 4. **Gait Cycle Feature extraction**: ○ Gait sequences are processed by ResNet-50 model to gain spatial features of each image. ○ Spatial features fed into LSTM layers to obtain temporal feature sets. ○ Feature sets summarized by mean-pooling to generate sequence-level features. ○ **Gait recognition** performed by computing cosine distance between two gait sequences. **Siamese LSTM** (SLSTM) Network is used for learning subtle periodic features. ○ Consists of two LSTM sub-networks, parallel in structure and same parameters. ○ SLSTM takes pairs of gait sequences as inputs, maps each gait sequence to a feature vector. ○ For all gait sequences, two sequences belonging to same identity selected randomly as positive pairs, whereas two sequences belonging to different identities selected randomly as negative pairs. ○ **Contrastive loss** layer applied to connect the two branches of SLSTM network. $D(v_i, v_j) = \|v_i - v_j\|_2$ Where v_i and v_j are sequence-level feature vectors The contrastive loss function is: $L(v_i, v_j, Y) =$ ¿ $Y\frac{1}{2}D^2 + (1-Y)\frac{1}{2}(max(m - D, 0))^2$	GoogLeNet = 0.3 iLIDS-VID – CN = 0.1 HOG = 0.2 LOMO = 0.1 GoogLeNet = 0.6 ➢ PROPRID: PROPRID fuses scores of FMLF and human gait models. Before late fusion, similarity vectors are normalized to (0,1) range. PRID-2011: FMLF – 0.65 Gait – 0.35 iLIDS-VID: FMLF – 0.6 Gait – 0.4 ○ PROPRID achieves *81.00%* and *62.33%* for Rank-1 recognition on both datasets, which exceeds second-best methods by 4.00% and 0.33% respectively. 3. Comparison with state-of-the-art methods: ○ *ASTPN* – Spatial and Temporal Attention Pooling Network, enables feature extractor to be aware of current input video sequence. ○ *RCN* – Recurrent CNN with temporal pooling. ○ *RFA* – Recurrent Feature Aggregation network, based on LSTM.

			of a region proposal netowrk and a binary mask classifier. • **ResNet-50** – 50-layer Residual Network, uses skip connection to propagate informatio n over layers. • **Contrasti ve Loss** – Distance-based loss function, used to find similarity between image pairs. • **Backprop agation** – algorithms used to train neural networks by using gradient descent. • **FMLF**: Fusion of Multi-Level Features, features are combined into comprehen sive appearanc e features. Weights obtained		where **m** is the min. distance margin of different identities' gait sequences. Total loss is sum of all gait sequence pair losses. ○ In training, both LSTM sub-networks are optimized simultaneously with weight sharing. ○ Pairwise gait sequences with similar/dissimilar labels fed into the two LSTM sub-networks *separately*. ○ Outputs are combined by contrastive loss layer to output contrastive loss. ○ Backpropagation with contrastive loss is used to train model. ○ During testing, Siamese architecture and loss function is **discarded**. ○ Fine-tuned ResNet-50 model and LSTM sub-netowrk jointly used as feature extractor. ○ Final decision made by *comparing cosine distances*. ○ Comparison is done using CMC curves.	PROPRID achieves the best performance in terms of rank-1 recognition for both datasets.

			on validation subset of training data. • **PRID-2011** – Person Re-ID Dataset, consists of images extracted from multiple person trajectories recorded from two different, static surveillance cameras. Images contain viewpoint change and differences in illusion, background and camera characteristics. • **iLIDS-VID** – iLIDS Video re-Identification dataset, 600 image sequences of 300 different pedestrians observed across 2 disjoint cameras			
[3]	April 2014	**Gait-assisted Person Re-identification in**	• **Person Re-ID** • **GEI** – Gait Energy Image • **Frame**	• Investigate if gait extracted from real-world	1. Silhouette sequence selection is used to ensure reliable gait extraction a. Non-zero pixels > 40% of image size for reliable gait extraction.	1. SAIVT SoftBio Dataset - a. Combined Color and Gait Features model for Re-ID: i. 150 subjects in

		Wide Area Surveillance	**Difference Energy Images (FDEI)** – Gait feature used to deal with incomplete silhouettes while retaining shape and motion changes, by constructing multiple thresholded GEIs from single gait cycle, and summing of these GEIs with differences between consecutive silhouettes. Positive values are retained, and negative values discarded. • **Z normalization** – standardization, by converting all indicators to a common scale with average of zero and standard deviation of one i.e	video sequences can be successfully leveraged as an additional feature along with appearance features for person re-ID. • Identify robust gait features that can be extracted from noisy and incomplete silhouettes and retain discriminative capability for Re-ID. • Implement a sparsified representation-based gait recognition method, where probe gait features	b. Aspect ratio of silhouettes over sequence of frames is used to estimate gait period. Assumption - silhouettes extracted and pre-processed. Pre-processing – size normalization and horizontal alignment of silhouettes. 2. Gait Period estimation: Gait period estimation is performed using lower half of silhouette region a. Aspect ratio of silhouette over frame sequence represented as 1D temporal signal, which is z-normalized and smoothed using moving average filter. b. Peaks in signal magnified by computing autocorrelation sequence; first derivative of autocorrelation signal used to detect zero-crossings. c. Zero-crossing of positive and negative peaks used to compute distance between prominent peaks d. Average of distances between consecutive peaks => Gait period in number of frames. 3. Gait Features – a. GEI – computed by averaging the silhouettes in spatial domain over gait cycle $$GEI(i,j)=\frac{1}{N}\sum_{t=1}^{N} S_t(i,j)$$ Where S_t is silhouette at frame t, (i,j) are spatial image co-ordinates N is estimated gait cycle period. b. FDEI - deals with incomplete silhouettes while retaining changes in shape and motion. Steps:	8 camera views; only 122 subjects in atleast 2 camera views. ii. Constraint: same subjects' gallery and probe camera views are different. iii. Simple background subtraction followed by low level image processing to extract silhouette sequences for gait feature extraction. iv. 5 randomly selected frames from a sequence used to extract color features v. All frames used to extract gait features. vi. As weight assigned to gait features increases, significant boost to performance. vii. Any weight combination of color and sparsified representation gait similarity-based Re-ID outperforms color and nearest-neighbour gait similarity. viii. Difference in performance between GEI and FDEI is

			$x_i^{'}=\frac{x_i-\mu}{\sigma}$ • **SAIVT SoftBio Dataset** • **MCID dataset –** multi-camera Re-ID dataset. • **Sparsified Representation** • **Bhattacharyya Distance** – Measures the similarity of two probability distributions For probability distributions *p* and *q* over the same domain X, $D_B(p,q)=-\ln$ Where $BC(p,q)=\sum_{x\in}$ is the **Bhattacharyya Coefficient** for discrete probability distributions .	are represented as linear combination of gallery gait features, using ○Gait Energy Images ○Frame Difference Energy Images	i.Construct GEIs from sub-cycles – $I_c(i,j)=\frac{1}{N_c}\sum_{t=1}^{N_c} S_t(i,j)$ Where N_c is period of sub-cycle ii.Thresholding GEIs to remove noise, creating dominant GEIs (DGEIs) for each sub-cycle – $DGEI_c(i,j)=¿$ iii.Frame difference computed by subtracting consecutive silhouettes and including only positive portion of difference in feature – $DS_t(i,j)=\begin{cases} 0,if\ S_t(i,j)\geq S_{t-1}(i \\ S_{t-1}(i,j)-S_t(i,j),oth \end{cases}$ iv.FDEI generated by summation of +ve frame difference and corresponding sub-cycle DGEI – $FDEI(i,j)=DS_t(i,j)+DGEI_c(i,$ 4. Feature Matching for Re-ID – a. Color features: Distribution of colors characterized using weighted HSV algorithm. i. Silhouette divided into three parts - head, torso and legs, by detection of one vertical axis of symmetry and 2 horizontal axes of symmetry. ii. Histogram for each body part weighted by distance from axis of symmetry. iii. Each histogram concatenated channel-wise to generate single color feature descriptor. b. Combined similarity measure: Similarity	negligible. b. Only Gait features for Re-ID: i. Re-ID only using gait features; only 23 subjects with usable gait features available across views. ii. Gait features outperform color features significantly iii. FDEI yield better Rank-1 matching accuracy than GEI, yet comparable. 2. MCID dataset – 9-camera network in and around a building; 40 subjects, 19 in multiple cameras. w_{gait} = 0.2 and w_{color}= 0.8 for best Re-ID accuracy a. Combined Color and Gait features model for Re-ID: i. For each camera pair, only subset of probe set i.e closed probe set (intersection b/w probe and gallery set IDs) used to establish Re-ID. ii. Gait features of MCID extracted only from lower body silhouettes (truncated GEIs). iii. Performance consistent w.r.t. SAIVT SoftBio

					measure between gallery subject G and probe subject P is weighted sum of color feature-based similarity and gait features based similarity i.e $dist(G,P)=w_{*d_{(G,P}}*w_{gait}.d_{gait}|G,P|$ Where d_{color} is color feature-based similarity d_{gait} is gait feaure-based similarity, w_{color} and w_{gait} are weights given to color and gait similarity, such that: w_{gait} = 1 - w_{color} i. Color similarity is calculated using Bhattacharyya distance. Color similarity for given probe-gallery pair is maximum similarity among all probe-gallery frame pairs. ii. Gait similarity is computed as reconstruction error between probe gait features and gallery gait model, utilizing sparsified representation. iii. Given a dictionary matrix built using labeled training features of several subjects, a test feature of i-th subject is linear combination of training images on same subject from dictionary. $V = [v_1,.....,v_n]$ is a dictionary matrix, where each column is obtained by vectorizing gait features belonging to all gallery subjects. Multiple columns in V may belong to same gallery subject. iv. Separate dictionary is built for each gallery subject, ensures that if probe subject is as close	dataset results. b. Only Gait features model for Re-ID: i. FDEI Re-ID within 4% of color accuracy. ii. Combining either GEI or FDEI boots Re-ID accuracy. iii. Pure GEI or FDEI performs poorly on MCID relative to SAIVT SoftBio dataset, due to truncated gait features. **Conclusion**: Sparsified representation based cross view gait recognition eliminates view angle estimation or gait feature reconstruction for matching.

					as possible to gallery subject in identity and view angle, probe image lies in linear span defined by subset of gait features constituting V. $I_p = V.\alpha$ Goal – to find sparest α that generated GEI_P in V i.e. solving the l_1-minimization $\hat{\alpha} = arg\,min \|\alpha\|_1 s.t \geq I_p = V.\alpha$ Solved using linear programming, leveraging augmented Lagrange multipliers. v. $d_{gait}(G,P) = \|I_P - V.\hat{\alpha}\|_2$ This is an estimate of how well $\hat{\alpha}$ reproduces GEI_P vi. If good quality silhouette sequences available and gait features extractable, w_{gait} is set to non-zero value. In absence of usable gait features, Re-ID established using only color features. Acceptable quality – No. of +ve pixels > 40% of total pixels in image.	
[4]	November 2012	**Person Re-identification using View-dependent Score-level Fusion of Gait and Color Features**	• **View-dependent score-level fusion** – adaptive weight control of gait and color features, where greater weight is given to gait features in similar-view	• Person re-identification across multiple non-overlapping cameras, using a spatio-temporal histogram as a gradient-based shape and	1. Color and gait features, observation view extracted from image sequence; color and gait feature matching returns individual distances. 2. Feature Extraction – a. Gait Feature: i. STHOG features used as gait feature containing both shape and motion information. ii. Pedestrian window size-normalized and divided into multiple spatio-temporal cells. $G = [G_x, G_y, G_t]^T$ is first order derivative for each of x-, y-, and t-axes from two successive frames.	1. Constructed own dataset; a. Walking image sequences of 27 subjects (14 for training, 13 for testing); 7 non-overlapping views ranging from front view to rear-oblique view. b. Near side view used as query view c. Size normalization window -> 100x140 pixels d. Spatio-temporal

			matching and to the color features in different-view matching. • **Assumptions** – ○ Pedestrian detection and tracking for each camera preprocessed. ○ Observation view for each pedestrian given. • **STHOG** – Spatio-temporal histogram of oriented gradient. • **Tanimoto distance** – Ratio of the intersecting set to the union set as a measure of similarity • **Baseline algorithm** • **LLR –** Linear Logistic Regression classifier • **CMC** – Cumulative matching characteristics.	motion gait feature to discriminate persons with similar color clothing in conjunction with background edge attenuation • Implement view-dependent score-level fusion framework to adaptively control weights of gait and color features.	$\phi = \tan^{-1}\left(\frac{G_y}{G_x}\right)$ where φ is spatial gradient. $\theta = \tan^{-1}\left(\frac{G_t}{\sqrt{G_x^2+G_y^2}}\right)$ where θ is temporal gradient. iii. Orientation of spatial and temporal gradients voted separately according to spatial gradient magnitude into individual 9-bin histograms for each cell, then normalized w.r.t. L_1 norm. iv. Both histograms combined into single 18-bin histogram, and histograms from all cells combined into single histogram. v. Effect of background variations must be suppressed; gradient-based background attenuation rather than background subtraction. vi. Attenuates edge at certain position if background edge also exists at same position and edge patterns of background/input images are similar. $G_s = \frac{\sqrt{G_x^2+G_y^2}}{\left(1+K G_x^{B^2} e^{\frac{-z_x^2}{\sigma^2}}\right)\left(1+K G_y^{B^2} e^{\frac{-z_y^2}{\sigma^2}}\right)}$ K and σ are hyper parameters to control background attenuation. G_x^B and G_y^B are horizontal and vertical gradients in background image. z_x and z_y denote dissimilarity of horizontal and vertical	cell sizes -> 100 pixels x 20 pixels x 3 frames, given 7 cells per window. e. K = 5.0; σ=10.0 f. Length of subsequence = 15 frames Repeated experiments for 20 trials, random choice of training and test subjects. Avg. performances of 20 trials evaluated using CMC. 2. Results – a. Performance using STHOG feature better than color feature for similar observation views; lower than color features for different observation views b. Adaptive fusion controls weight of STHOG and color features; achieves best performance

					edge patterns in background and input images. $z_x = max\{\|I^B_{x,y} - I_{x,y}\|, \|I^B_{x+1,y} - I_{x}$ $z_y = max\{\|I^B_{x,y} - I_{x,y}\|, \|I^B_{x,y+1} - I_{x}$ Where $I^B_{x,y}$ and $I_{x,y}$ are pixel intensities at (x,y) in background and input images. b. Color Feature: i. Color histogram used with modified HSV system i.e. Histogram for each cell with 10 bins; 7 hue bins and 3 value bins (white, gray, and black). ii. Region-based feature; hence, background region masked out by background subtraction during voting. iii. Color histograms normalized for individual cells and concatenated into single histogram. 3. Matching – a. Features are synchronized when matching using baseline algorithm as follows: i. Query sequence segmented into subsequences with certain length (gait period) ii. Each segment phase-synchronized with other sequence by shifting frame to minimize total distance b/w them iii. Baseline algorithm utilizes Tanimoto distance; replaced with $L_{0.5}$ norm to improve robustness wrt outliers. iv. Distance of individual sub-sequences computed and minimum	

					value chosen as final distance. 4. View-dependent Score-level Fusion – a. Score-level fusion function maps 2-D distance vectors derived from STHOG and color features into single distance. b. Fusion function trained separately for each discrete observation view difference, due to reliability of STHOG features being highly dependent on observation view difference. c. LLR of likelihood ratio b/w +ve and -ve samples is used for simplicity and high generalization capability. d. Relative distance of each negative sample w.r.t corresponding positive sample is input into fusion framework. e. After training, input 2-D distance vector converted into single fused distance. f. Minimum fused distance between person in one camera image to another camera image used for Re-ID.	
[5]	June 2013	**Gait-Based Person Identification Robust to Changes in Appearance**	• **Affine moment invariants** – moment-based descriptors, invariant under general affine transform. Centralized moment of order(p+q) of object O: $\mu_{pq}=\sum_{(x,y)\in O}\sum$	• Implement a robust method of person identification in spite of changes in appearance – without using database of predicte	1. Average GEI is calculated over one gait cycle and the human body area is divided into multiple areas (5). a. Silhouette is extracted by background subtraction b. Human body area is scaled to uniform height and set to 128 pixels $\acute{I}(x,y)=\frac{1}{T}\sum_{t=1}^{T} I(x,y,t)$ Where T is no. of frames in one gait cycle I(x,y,t) represents intensity of pixel (x,y) at time t. c. High intensity values – body parts that move little during gait cycle i.e. head,	1. First Experiment (robust to noise and deficit in silhouette images) a. CASIA-B: Each subject has 6 walking sequences i. First four sequences are used for training datasets. ii. CCR calculated by dividing the sequences of each subject into two sets.

			Where x_g and y_g define center of object. $A_1=\frac{1}{\mu_{00}^4}(\mu_{20}$ Where A_1 is the first affine moment. • **CCR** – Correct classification rate	d appearance changes – by dividing the human body image into multiple areas and performing feature extraction.	torso; d. Low intensity values – body parts that move constantly i.e. lower parts of legs, arms; e. Gait cycle estimated by calculating first affine moment invariant at each frame. 2. Estimation of matching weight: a. Affine moment invariants in database & of subject are whitened at each area b. Distance between features of subject and those of all datasets in database determined ($d^k_{n,s}$) – $d^k_{n,s}=¿$ Where ${}^wA^k_{SUB}$ and ${}^wA^k_{DB(n,s)}$ show whitened affine moment invariants of subject and person in database. c. Whitening is done by – i. Applying principal component analysis to calculated affine moment invariants ii. Projecting them to new features space iii. Normalizing projected affine features based on corresponding eigenvalue $1\leq n\leq N$ Where N is number of people in database $1\leq s\leq S$ Where S is no. of sequences of each person $1\leq k\leq K$ Where K is no. of divided areas d. $d^k_{n,s}$ calculated between features of subject and those of each sequence in database at each area. 3. Estimate matching weights based on similarity between features of subject and database.	iii. Testing by 2-fold cross-validation method (124x3 sequences for training, rest for testing) iv. CCR calculated by dividing no. of test datasets correctly classified by that of all test datasets. Proposed method shows highest performance of 97.7% (K=17,M=45). CCR increases with K and M b. CASIA-C: Each subject has 4 sequences i. Sequences divided into two sets. ii. Testing through 4-fold cross validation (153x3 sequences used for training, rest for testing) iii. Highest performance of 94% (K=7,M=65). Although silhouette images of worse quality than CASIA-B, person identification with high performance. 2. Person Identification robust to appearance changes: a. CASIA-B-BG and

					High matching weights – less appearance changes Low matching weights – more appearance changes a. At each area k, select sequences from database if $d^k_{n,s} < d^k_{min}$. $\acute{d}^k_n = min_n \acute{d}^k_n$ $\acute{d}^k_n = \frac{1}{S}\sum_{s=1}^{S} d^k_{n,s}$ b. Consider that similarities of non-selected sequences are low, so distance of these sequences are redefined as d_{max}. $d^k_{n,s} = d_{max}$ This process allows setting low similarities to areas of each sequence in database. c. Repeat a and b for all areas. d. Sum of distances for all areas calculated by $D_{n,s} = \sum_{k=1}^{K} d^k_{n,s}$ and k-nearest neighbor method used to identify subject. (k=1 in experiment)	CASIA-B-CL dataset used. b. K=17 and M=45 for CCR calculations, similar to first experiment. c. CASIA-B-BG: i. Handbag (42 sequences) ii. Shoulder bag (171 sequences) iii. Backpack (30 sequences) iv. Others (3 sequences) CCR = 91.9% CCR with method without matching weights = 20.2% d. CASIA-B-CL: i. Thin coat with hood (30 sequences) ii. Coat (24 sequences) iii. Coat with hood (16 sequences) iv. Jacket (70 sequences) v. Down jacket (62 sequences) vi. Down jacket with hood (28 sequences) vii. Down coat with hood (16 sequences) CCR = 79% CCR without matching weights method = 22.4%
[6]	September 2013	**Person re-identification using height-based gait in**	• **KL-Divergence (relative entropy)** – measure of difference between	• To implement person re-identification in color-	Training dataset of N people a. Height dynamics extraction – i. Multiple people tracking framework (position and appearance-based) to generate trajectories of detected people, using	1. TUM-GAID Dataset – a. Public dataset, containing multiple trials of subjects, captured with fronto-parallel

		colour depth camera	two probability distributions, given by D_{KL} • **DGEI** – Depth gradient GEI	depth camera using height temporal information of people, by using a feature-selection scheme for each person based on KL-divergence to identify a discriminative subset of frequency bins for height-based gait feature. • To formulate a probabilistic matching framework incorporating the selected frequency bins, using the discriminative ability of feature vector obtained	3D color depth cameras. ii. Each 3D bounded box contains only foreground 3D color-depth blobs. iii. Top-down camera view: 1. Since origin is defined in ground plane, person's height estimated from depth value corresponding to highest point in detected blob. 2. To increase robustness, mean of *t* highest points calculated. iv. Front-parallel camera view: 1. Height derived from foreground/depth silhouette height. 2. Height/scale variations are adjusted using foreground depth info. Given apparent person height *h*, average foreground depth *d* and pre-defined normalizing depth value *s*, normalized person height *h'* given as *h' = (h/s)*d.* *b.* Feature vector from frequency response (Frequency domain transformation of height dynamics) – *i.* k-point DFT used to generate frequency response and obtain k-dimensional feature vector $y_m^n = \left(dft\left(h_m^n, k \right) \right)$ ∨ Where $\mathbf{y}^n_m$ represent absolute real part of k-Fourier coefficients (frequency bins), $\mathbf{h}^n_m$ is estimated height dynamics for person *n* and trajectory *m*. c. Feature Selection – Technique of identifying subset of feature vectors to enhance overall	color depth camera. b. Trials partitioned into two clusters based on clothing; clothing-1, clothing-2. c. During training, multiple trials of clothing-1 for each subject. d. Test dataset partitioned into two subsets, based on clothing for each subject; i. First test partition – trials of clothing-1. ii. Second partition – trials of clothing-1 and clothing-2. 2. Dataset-1 and Dataset-2 – a. Multiple walking subjects using top-down color depth cameras b. Dataset-1: 2 sequences @15Hz using 5 ceiling mounted color depth cameras with 5 and 10 subjects, walking in indoor environment with varying illumination. c. Dataset-2: 3 subjects@15Hz with uniform illumination. k = 256-DFT 10 features selected. Results: • For first test partition, baseline

				by integration of height temporal information with a color & height-based appearance model.	identification accuracy We identify subset of frequency bins unique to each person i. For each person *n*, compute empirical Gaussian mean and standard deviation over set of feature vectors M i.e. $\left\{y_m^n\right\}_{m=1}^{M_n}$ and obtain k unimodal Gaussian distributions $\left\{\widetilde{N}_n\left(\mu^k,\sigma^k\right)\right\}_{k=1}^{K}$ corresponding to k frequency bins. ii. KL-divergence for *n* and frequency bin *k*: $v^n(k)=\sum_{\substack{p=1\\p\neq n}}^{N} D_{kl}\left(\widetilde{N}_n(k)\right) \vee \widetilde{N}_p(k$ iii. Large KL divergence values – minimum overlap Small KL divergence values – high degree of overlap iv. KL divergence value tends to zero when two gaussian distributions overlap completely. v. Selected frequency bin indices $\Lambda^n=\left\{\lambda^n(j)\right\}_{j=1}^{J}$ vi. To avoid overfitting, test trajectories not present in training dataset during feature selection d. Probabilistic Identification – Maximum likelihood classification used to identify test person. i. Extract test person's height dynamics and obtain frequency response. ii. Identify label in training dataset using likelihood function. Frequency response $q=\left\{q(k)\right\}_{k=1}^{K}$ Label l	algorithm (RGB-HT) performs marginally better than gait-based algorithm. • For second test partition, gait-based algorithm performs better than baseline algorithm. • Robust algorithm is limited to batch-mode applications due to supervised feature selection.

					$l=argma\,x_n\prod_{j=1}^{J} p\left(q\left(\lambda^n(j)\right);\mu^n\left(\lambda^n(\right.$ Where $q\left(\lambda^n(j)\right)$ represents frequency bin j in k-dim test feature vector, $\mu^n\left(\lambda^n(j)\right) \wedge \sigma^n\left(\lambda^n(j)\right)$ are empirical mean and standard deviation. e. Baseline Identification Algorithm – i. RGB-height histogram: 4D histogram with discrete bins corresponding to person height and red, green and blue channels. ii. 3D RGB histogram created at each non-overlapping person height bin. iii. Nearest-neighbour classification using min. difference of pairwise assignment distance to mean feature vector ($\dot{G}_n$) for person n in the training dataset. $l=argmax_n\left(-d\left(q,\dot{G}_n\right)\right)$ f. Combined Feature vector-based algorithm – i. Combination of color, person-height and gait as feature. ii. RGB-height histogram and height-dynamics gait features combined i.e. respective maximization-based matching measures are combined.	
[7]	June 2016	**Gait-Based Person Identification Considering Clothing Variatio**	• **GEI** – Gait Energy Image • **GEnI** – Gait Entropy Image • **DFT** – Discrete	• To implement the proposed method of person-reidentification	1. Silhouettes extracted from each frame of video sequence using background subtraction 2. Silhouettes normalized into 120x88 pixels and registered to form spatio-temporal GSV. 3. Gait period detection over complete gait cycle.	***OU-ISIR treadmill dataset-B*** was only used for training/testing. 1. **Performance Evaluation of GEI, GEnI, DFT and EnDFT Feature**

		n	Fourier Transform • **EnDFT** – Frequency-domain Gait Entropy • **Sequence combination** • **Dynamic part selection** • **GSV** – Gait Silhouette Volume	using a combination of gait sequence and dynamically part weighting using threshold value and weighted integration. • To test the proposed method on OU-ISIR Treadmill dataset B, which includes 68 subjects with 32 combinations of clothing types. • To calculate GEI, GEnI, DFT and EnDFT over a complete gait cycle and utilize dynamic weighting on the gait features.	4. **Gait Representation** – a. **GEI**: $GEI(x,y)=\frac{1}{N}\sum_{n=1}^{N}B(x,y,n)$ b. **GEnI**: Avg. self-information obtained from k number of outputs generated from source that generates k symbols: $-k\sum_{j=1}^{j}p(a_j)\log p(a_j)$ Where a_j are symbols, $p(a_j)$ are source symbols probability. $H(Z)=-k\sum_{j=1}^{j}p(a_j)\log p(a_j)$ Entropy of this variable over each gait period: $H(x,y)=-k\sum_{j=1}^{1}p_j(x,y)\log p_j($ Where $p_1(x,y)$ is value from GEI and $p_0(x,y) = 1 – p_1(x,y)$ c. **DFT**: G(x,y,k) [DFT] and amplitude A(x,y,k) for temporal axis: $G(x,y,k)=\sum_{n=0}^{N-1}B(x,y,n)e^{-jw_0k}$ $A(x,y,k)=\frac{1}{N}\vee G(x,y,k)\vee$ Where N is no. of frames in gait cycle, w_0 is base angular frequency, k is frequency component. DFT has 3 components: (0-2 frequency component) (higher frequency components removed as noise) • 1st component equivalent to GEI • Middle component shows asymmetry of left and right motion • Last component represents symmetry. d. **EnDFT**: EnDFT is computed from	**sequence combination**: a. CCR calculated for rank-1 – EnDFT has best performance at 70.13% recognition rate. 2. **Comparison with existing methods**: Proposed sequence combination +EnDFT has the highest recognition rate relative to existing methods such as **Baseline+GEI** (60.75%), **Part-based (without weighting)** (65.99%).

					DFT image using the Gait Entropy equation, where $p_1(x,y)$ is the intensity value of DFT. EnDFT gives more weights into dynamic areas; less (near to zero) for static areas; uses all three components of DFT 5. **Dynamic Part Weighting and Sequence Combination** – a. Divide human body structure into small fragments; i.e each small fragment is single row b. Evaluation of each row from bottom of gait image, merging with next immediate upper row and finds recognition rate. c. Process continued till top most row is evaluated. d. Median value b/w local maxima and local minima calculated to divide the body image into multiple areas. e. Median value is calculated for next local maxima and so on to bottom of image. f. Estimate matching weight for each area based on similarity b/w extracted probe features and gallery features for standard clothes g. Probe identified by weighted integration of similarities in all areas h. Sequence combination used to calculate final weight. i. K-nearest neighbor classifier used to classify probe.	
[8]	Febru ary 2019	**Gait-based person re-identific ation**	• **LBP** – local binary pattern • **HOG** – Histogram of oriented	• To impleme nt a new method for gait-based	**Signature Extraction** – 1. Gait Cycle Estimation 2. GEI generation 3. Semantic classification: Attributes such as carried objects (backpack,	CASIA-B database used. Results compared between LBP, HOG and P-LBP: 1. *Descriptor Choice*

		under covariate factors	gradients • **Gait features**	person re-identification relying on dynamic selection of human parts by computing a new person descriptor from relevant human parts, selected depending on presence of semantic information.	shoulder bag, and handbag) and clothes are semantic. a. Offline stage: i. Data preparation of training database related to semantic attribute: Construct 2D table from database; each row represents bounding box's image; each column represents feature. Last column contains sematic attribute class. ii. Predictive model for each attribute class constructed using a. Support vector machines (SVM) b. Tree bagger-based decision tree (FT) c. Neural Network (NN) iii. Online Stage: 1. Extracting features to represent global information of each image. 2. Labels generated depending on sematic info. of image. 4. Feature extraction: Dynamic selection of parts from divided GEI image. (7 Parts) Set of relevant parts V: $V=\{c_1.G_{H1}(x,y),\ldots,c_7.G_{H7}($ $\begin{cases} c_i=0, if\ semantic\ attribute \exists \\ c_i=1, otherwise \end{cases}$ where i is index of part $G_{Hi}(x,y)$. a. Extraction of salient and suitable features to capture gait characteristics using partial-LBP (local binary pattern). b. Takes pixels of image by thresholding 3x3 neighborhood of each pixel with center value, and mapping result as binary number (0/1).	*experiments*: 204 images used. a. P-LBP has advantages in terms of rank 1; 23.92% better than HOG and 47.05% better than LBP. 2. *Semantic Attributes classification:* a. 124 person images, 10 sequences per person. b. 744 carrying nothing, 248 with bags, 248 wearing coats. c. Classification done using SVM, NN and FT(tree bagger). d. Neural network gives best accuracy (>89%) for detecting semantic attributes. e. For normal semantic attributes (nothing), SVM shows better performance. 3. *Dynamic selection of human parts*: a. Comparison between P-LBP 7parts, invariant parts, and dynamic selection; P-LBP dynamic selection gives best results, with rank-1 performance of 52.352%, rank-3 of 69.8% and rank-5 of 78.23%.

					$P-LBP(x_c,y_c)=\sum_{n=0}^{7} s(i_n-i$ Where i_c corresponds to value of center pixel, i_n to value of 8 surrounding pixels and s(x) defined as $s(x)=\begin{cases}1, \wedge x\geq 0\\0, \wedge x<0\end{cases}$ c. P-LBP immediately records bit sequences to maintain more local information. d. Binary sequence is further used as gait signature for each relevant part. 5. Signature Matching: a. Measures distance between probe person signature and each gallery image signature b. Common parts not affected by semantic attribute for gallery and probe image are used. c. Gallery set arranged according to similarity to generate ranked list. d. Top rank person's identity assigned to probe image. e. Euclidean distance used as metric.	
[9]	May 2015	**Enhancing person re-identification by integrating gait biometric**	• **PCA** – Principal component analysis • **Gabor Features** • **Score level fusion**	• To implement a method to enhance person-reidentification by integrating gait biometric • Using hierarchical feature extraction and descripto	**1. Hierarchical feature extraction:** a. Gait feature: Original gait feature is 2048-D vector for each sequence; dimensionality reduction done using PCA to avoid overfitting. b. Appearance features: i. HSV space color histogram: Each dimension of HSV (transformed from RGB and equalized to reinforce illumination robustness) is divided into 128 bins to count number of pixels in corresponding bins	CASIA Gait Database B used, contains 11 views of 124 individuals with 3 conditions: Bag, clothes, normal. View angles: 0,18,36,54,72,90,108,126,144,162 and 180 degrees. 1. Comparison with baselines: Feature-level + MLR and score-level + MLR boh perform much better compared to other

				r matching with learned metric matrices.	ii. Gabor features: 2D Gabor filter is Gaussian kernel function, modulated by complex sinusoidal plane wave. Used to generate texture feature in face and fingerprint recognition. $\psi_{u,v}(z)=\frac{\|k_{u,v}\|^2}{\sigma^2}e^{\frac{-\|k_{u,v}\|^2\|z\|^2}{2\sigma^2}}\left[e^{i}\right.$ Where u is the orientation, v is scale of Gabor filters and z = (x,y) $k_{u,v}=k_v e^{i\varphi}$ Gabor filter of 5 different scale and eight orientations, with: $\sigma=\pi, k_{max}=\frac{\pi}{2}, f=\sqrt{2}$ c. Metric learning to rank: i. General metric learning algorithm, based on structural SVM & view metric learning. ii. Used to build view-angle independent mapping. iii. Data from different views of same person clustered in metric space, so that similarities between people measured as distances in metric space, regardless of view angles. **2. Fusion:** a. Score-level fusion: i. View-independent global function, formulated as $D_{FIN}(S_1,S_2)=D_{APP}(S_1,S_2)+\theta D_{GEI}$ Where θ is the weighting factor. Value of θ range from 0.001 to 10; best result at $\theta=1$ b. Feature-level fusion: i. Fuse two type features before calculating distance. $\vec{F}=[\vec{A},\vec{G}]$ Where $\vec{A}\wedge\vec{G}$ is	methods. 2. Closed/Open set person-reidentification: Feature-level +MLR performs better than score-level+MLr as GEI has weaker discrimination than color and textures in short period Re-ID. However, GEI is appropriate supplement to appearance features. Integrating gait feature alleviates the negative effects of different appearance of people and make algorithm more robust.

					appearance feature vector and GEI feature vector.	
[10]	December 2016	**Person Re-identification in frontal gait sequences via Histogram of Optic flow Energy Image**	• **Optic flow** • **Histogram of Flow Energy Image** • **GEI**	• Implement a novel method of re-identifying people in frontal video sequences, based spatio-temporal representation of gait based on optic flow features, called HOFEI.	1. Histogram of Flow: a. HOG encoding scheme on human detection bounding boxes, by using 'ground truth' annotation. b. Polar cells used to better represent spatial location of limbs and head w.r.t time. c. Optic flow image divided into cells using polar sampling strategy; compute histogram of flow orientation weighed by magnitude. $HOF^t = [HOF_1^t \ldots HOF_i^t \ldots H($ Where nR is no. of angular regions (cells) and nB is no. of bins defining each cell. HOF^t is of dimension 64 with nR=8 and nB=8. d. HOF^t computed for each frame through video sequence S 2. Gait Cycle Estimation: a. To estimate gait period, subset of histogram bins $HOF_2^t \wedge HOF_3^t$ are used, which represent lower limb motions. Amplitude has good SNR ratio for detection. b. The periodic sinusoidal curve generated by plotting HOF peaks of single leg against frame (as function of time). Frames between two consecutive peaks represents gait cycle. 3. Histogram of Flow Energy Image: a. HOF^t representations averaged over full gait cycle to obtain HOF energy image $HOFEI = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} HOF^t$ $t_1 \wedge t_2$ are the beginning	Experiments conducted in two scenarios. Re-ID in controlled scenario vs Re-ID in uncontrolled (busy office) scenario. For first scenario, CASIA Dataset B used: 1. contains multiple videos of subjects including normal and apparel change conditions. 2. 105 subjects considered, 3 gait cycles extracted for each subject. 3. Stefans implementation, using the Lucas-Kanade method was applied. For second scenario, HDA Person Dataset is used: 1. 13 indoor cameras, recording simultaneously for 30 min during busy noon hour inside university building. 2. Only camera19 was used, containing frontal gait sequences. 3. Resolution of 640x480, frame rate of 5fps. 4. 12 people, min. of 3 gait cycles extracted per person for training and testing. Due to limitations of large video sequences and

					and end frame indices of gait cycle	varying appearance conditions per person, Experiment 1 and 2 excluded in HDA Dataset. Experiments: 1. Recognition in regular conditions: a. First four sequences used for training, last two used for probe set. b. CMC Curve shows CCR of 74.29%. c. Compared to silhouette based approaches (GEI, FDEI, GHI, GMI), CCR in between others. d. GEI and FDEI have better CCR due to use of segmented binary silhouettes and better classification method. 2. Re-identification under change in appearance: a. Apparel change CCR for bag is 66.67%. b. CCR for coat is 59.05%. c. Lower CCR of coat is due to global change in flow features. 3. Variable Distance to camera: a. CASIA: i. For 1st case study, 6 $\mathbf{D_{far}}$ probe and 12 training set ($\mathbf{D_{middle}}$ and $\mathbf{D_{near}}$) per person. ii. For 2nd case

						study, $\mathbf{D_{middle}}$ is considered as probe and $\mathbf{D_{near}}$ and $\mathbf{D_{far}}$ are training sets. iii. For 3rd case study, $\mathbf{D_{near}}$ is probe set, $\mathbf{D_{middle}}$ and $\mathbf{D_{far}}$ are training sets. iv. $\mathbf{D_{middle}}$ case outperforms other two cases with **33.81%** CCR; as trained on extreme ranges, classifier performs interpolation when predicting values for middle range. b. HDA Dataset: i. 24 training descriptors, 12 test probes. ii. $\mathbf{D_{middle}}$ has highest CCR of **75%**, with **50%** and **58.33%** for $\mathbf{D_{far}}$ and $\mathbf{D_{near}}$ cases.

FIGURE 1.1: Literature Review of Current Gait-Based Person Identification Techniques

Chapter 2

Proposed Methodology

In this project, we proposed a motion-based approach to person identification, using *affine moment invariants* (AMIs) as feature descriptors. *Active Energy Image* (AEI) are generated from the subject silhouette sequence. The AEI is divided into multiple areas and features are extracted for each area. Matching weights are estimated for each area based on similarity between extracted features and those in the database sequences, using Euclidean distance as a metric. Finally, the subject is identified using the weighted sum of similarities of all areas. A nearest-neighbor classifier is implemented for identification of the subject. Experiments are performed on the CASIA-B Gait Database, using the normal walking sequences.

The approach can be divided into the following steps:

1. Extract active energy image (AEI) from the gait sequence and divide the image into multiple areas.

2. Extract affine moment invariants (AMI) at each area to use as gait features. The database consists of a set of affine moment invariants of multiple people who wear standard clothing with no accessories.

3. AEI of subject person is also divided in the same way as that of the dataset and gait features (AMIs) are extracted.

4. Matching weights are estimated at each area based on similarity between features of the subject and those in the database.

5. The subject is identified by the weighted combination of the similarities of all areas. Nearest Neighbor classifier is used to classify the test subject.

2.1 Active Energy Image

In gait recognition, two types of information - static and dynamic - are extracted from the gait sequence. According to [13], GEI is efficient in extracting static and dynamic information, both implicitly and explicitly. However, since GEI represents gait as a single image, there is a loss of dynamic information such as the fore-and-aft frame relations. Active Energy Image (AEI) feature representation can be used to solve the above problems.

FIGURE 2.1: Gait silhouette sequence of individual

Given a pre-processed binary gait silhouette[14] $i = i_0, i_1, ..., i_N - 1$, where f_j represents the j^{th} silhouette (Figure 2.1), N is the total number of frames in the sequence, the difference image between frames is calculated as follows (Figure 2.2):

$$I_j = \begin{cases} i_j(p,q), & j = 0 \\ ||i_j(p,q) - i_{j-1}(p,q)||, & j > 0 \end{cases} \tag{2.1}$$

FIGURE 2.2: Difference Images between consecutive significant frames of gait sequence

FIGURE 2.3: Active Energy Image

The AEI is defined as (Figure 2.3):

$$A(p,q) = \frac{1}{N}\sum_{j=0}^{N-1} I_j(p,q) \tag{2.2}$$

The active energy image representation is divided into 'K' segments as shown in Figure 2.4:

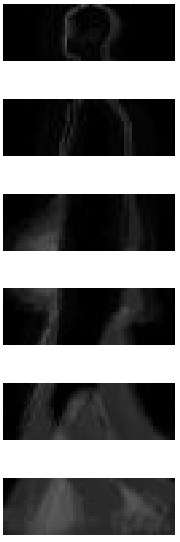

FIGURE 2.4: Segmented AEI (6 segments)

2.2 Affine Moment Invariants

2.2.1 Image moments and moment invariants

Image moments are certain particular weighted average (moment) of the image pixels' intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. Some properties of images which are found via image moments include area (or total intensity), its centroid, and information about its orientation.

Image moments are well-known in their applications for image analysis, since they are used to derive **moment invariants**, which are invariants to transformations such as translation and scaling. The well-known ***Hu moment invariants*** were shown to be invariant to translation, scale and rotation. However, Flusser[29] showed that the traditional set of Hu moment invariants is neither independent nor complete.

Affine moment invariants, proposed by Flusser and Suk [17] are moment-based descriptors, which are invariant under a general affine transformation. The invariants are generally derived by means of the classical theory of algebraic invariants [18], graph theory [19,20] or the method of normalization [21]. The most common method is the use of graph theory.

The moments describe shape properties of an object as it appears. For an image, the centralized moment of order *(p+q)* of an object O is given by

$$\mu_{pq} = \sum \sum_{(x,y)\in O} (x - x_g)^p (y - y_g)^q A(x, y) \tag{2.3}$$

Here, x_g and y_g define the center of the object. More specifically, they are calculated from the geometric moments m_{pq}, given by $x_g = \dfrac{m_1 0}{m_0 0}$ and $y_g = \dfrac{m01}{m00}$.

The affine transformation can be expressed as

$$u = a_0 + a_1 x + a_2 y \tag{2.4}$$

$$v = b_0 + b_1 x + b_2 y \tag{2.5}$$

[20] In our method, the number of affine moment invariants ($\mathbf{A} = (A_1, A_2, ..., A_M)^T$) M is 10. We show 5 such invariants:

$$\begin{aligned}
A_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2) \\
A_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 \\
&\quad -3\mu_{21}^2\mu_{12}^2) \\
A_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
&\quad +\mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \\
A_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} \\
&\quad +9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} \\
&\quad +6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\
&\quad -8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad +12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2) \\
A_5 &= \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2)
\end{aligned}$$

2.3 Estimation of matching weights

[5] Matching weights are calculated by first whitening the AMI in the database and of the subject at each area. The distance $d_{n,s}^k$ between the features of the subject and those of all datasets in the database is computed as follows:

$$d_{n,s}^k = ||^w A_{SUB}^k - ^w A_{DB_{n,s}}^k|| \tag{2.6}$$

where $^w A_{SUB}^k$ and $^w A_{DB_{n,s}}^k$ are the whitened AMI of the subject and those of a person in the database.

The whitening of the affine moment invariants is done by:

1. Applying *(*principal component analysis (PCA)) to the calculated affine invariants and projecting them to a new a feature space

2. Normalizing the projected features based on the corresponding eigenvalue.

*(*n,s and k) are $1 \leq n \leq N$ (N is number of persons in database), $1 \leq s \leq S$ (S is number of sequences of each person), and $1 \leq k \leq K$ (K is number of divided areas). $d_{n,s}^{k}$ is calculated in the Euclidean norm.

To estimate matching weights, we use the similarity between subject features and database features; high matching weights are set to areas with less appearance changes, and low matching weights set to those with more appearance changes.

Steps to estimate matching weights:

1. At each area k, select sequences from the database where $d_{n,s}^{k} < \bar{d}_{min}$. These selected sequences are considered to have high similarity with the subject. $\bar{d}_{min}$ is defined as:

$$\bar{d}_{min} = min_n \bar{d}_n^k \tag{2.7}$$

$$\bar{d}_n^k = \frac{1}{S} \sum_{s=1}^{S} d_{n,s}^{k} \tag{2.8}$$

   We consider that in each area, if at least one sequence of a person in the database is selected, then the matching scores of all sequences of that person are also high.

2. The non-selected sequences are all considered to be low similarities, so the distances of these sequences are redefined as d_{max} *(*$d_{max} = max_{n,s,k} d_{n,s}^{k}$*)*

3. The above procedure is applied for all areas and the sum of distances for all areas is calculated by $D_{n,s} = \sum_{k=1}^{K} d_{n,s}^{k}$. Subject is identified by nearest-neighbor method.

2.3.1 Principal Component Analysis

In machine learning classification problems, there are often too many features on the basis of which the final classification is done. The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, most of these features are correlated, and hence redundant. This leads to a need for dimensionality reduction, which reduces the complexity of a model and avoids overfitting.

Principal Component Analysis (PCA) is one of the most popular algorithms used for dimensionality reduction. Proposed by Karl Pearson, PCA is an unsupervised linear transformation technique used in identifying patterns in data based on the correlation in features. It aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one.

The algorithm can be summarized in the following steps:

1. Standardize the d-dimensional dataset: This is usually done by standardizing the data such that the d-dimensional dataset will have a mean of 0 and a standard deviation of 1.
2. Construct the covariance matrix.
3. Decompose the covariance matrix into its eigenvectors and eigenvalues.
4. Sort the eigenvalues by decreasing order to rank the corresponding eigenvectors.
5. Select k eigenvectors which correspond to the k largest eigenvalues, where k is the dimensionality of the new feature subspace ($k \leq d$).
6. Construct a projection matrix W from the "top" k eigenvectors.
7. Transform the d-dimensional input dataset X using the projection matrix W to obtain the new k-dimensional feature subspace.

Chapter 3

Experiment and Results

3.1 Implementation

The implementation was done in Python 3.6.8, on a Windows 10 laptop, with an Intel i7 8th Generation processor and 8GB of RAM. The following modules with the mentioned versions were used:

- Numpy - 1.17.2
- Scikit-learn - 0.21.3
- SciPy - 1.3.1
- OpenCV - 4.1.1.26
- Pillow - 6.2.0
- Scikit-image - 0.15.0
- ImageIO - 2.6.1
- PyWavelets - 1.0.3

Two custom modules were also written to provide the following functionality for the main program:

1. *functions_module.py*: Contains functions to generate the active energy image (AEI), divide the image into areas and compute the accuracy.(Nearest Neighbor Implementation)
2. *ami.py*: Contains functions to extract affine moment invariants from the AEI images.

3.2 Experiment and Result

The proposed method was applied to the CASIA-B Gait Dataset. The CASIA-B Dataset is a large multi-view gait database, created in January 2005. There are 124 subjects, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered.

In the experiment, I used the lateral view (90 degree) standard walking sequences to analyze the prediction rate of the algorithm. The CASIA-B Dataset consists of six standard walking sequences for each person. I calculated CCRs by dividing the six sequences of each subject into two sets; 124 x 5 sequences were used for training and the rest were used for testing.

The total number of affine moment invariants M were varied from 1 to 10 and the parameter K (number of divided areas) from 5 to 30. In case of $K = 23$ and $M = 5$, the proposed method shows the highest accuracy of **91.13**% (Figure 3.1).

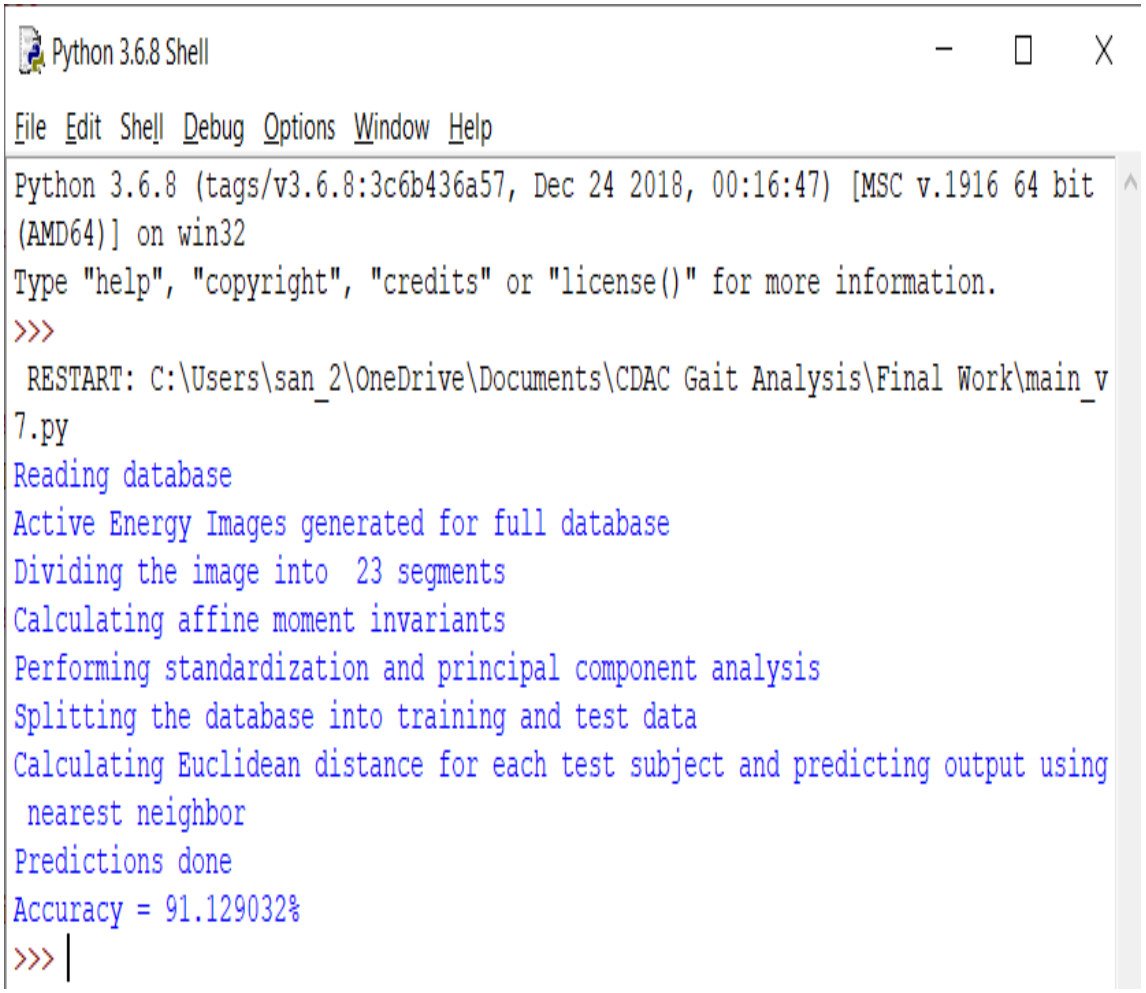

FIGURE 3.1: Best Case Results (K=23, M=5)

Chapter 4

Conclusions and Future Work

4.1 Conclusion

A person re-identification method was proposed that utilizes active energy images,focusing on the dynamic information in the gait sequence, and is robust to appearance changes. In this method, the active energy image of the gait sequence was calculated and the human body area was divided into multiple areas. Affine moment invariants were extracted at each area as gait features and a matching weight was estimated for each area based on the similarity between features of the subject and those in the database. Then, the subject was identified by weighted sum of similarities in all areas using the nearest neighbor classifier.

In this research, we focused on developing a method that utilizes gait features using the dynamic information embedded in the walking speed and style. There are other potential factors that may affect the performance of the algorithm, such as different viewing angle, walking speed, etc. The specific immediate improvement to be done is to increase the current recognition rate without compromising on the dynamic features.

4.2 Future Work

Future work on the algorithm will address the following:

1. Improvement of the recognition rate by
    - Extracting features specific to the human body segments such as head, torso, arms and legs, and assigning matching weights based on these body segments.
    - Increasing the number of feature descriptors (AMIs) to get a better measure of variance in dynamic information of the image.
2. Testing and fine-tuning the algorithm on other viewing angles.
3. Testing the algorithm on other gait sequences that include additional clothing and accessories on the human body area.
4. Implementing a neural network to improve estimation of matching weights for similarity of subject and database features.
5. Testing of the algorithm on real-time wide area surveillance to identify and track subjects over multiple non-overlapping cameras.

Bibliography

[1] M. Babaee, L. Li, and G. Rigoll, "Person identification from partial gait cycle using fully convolutional neural network," *Neurocomputing*, 04 2018.

[2] S. Li, M. Zhang, W. Liu, H. Ma, and Z. Meng, "Appearance and gait-based progressive person re-identification for surveillance systems," 09 2018, pp. 1–7.

[3] A. Gala and S. Shah, "Gait-assisted person re-identification in wide area surveillance," 04 2015, pp. 633–649.

[4] R. Kawai, Y. Makihara, C. Hua, H. Iwama, and Y. Yagi, "Person re-identification using view-dependent score-level fusion of gait and color features," 01 2012, pp. 2694–2697.

[5] Y. Iwashita, K. Uchino, and R. Kurazume, "Gait-based person identification robust to changes in appearance," *Sensors (Basel, Switzerland)*, vol. 13, pp. 7884–901, 06 2013.

[6] V. John, G. Englebienne, and B. Krose, "Person re-identification using height-based gait in colour depth camera," 09 2013.

[7] I. Ahmed and M. Rokanujjaman, *International Journal of Innovations & Advancement in Computer Science (IJIACS)*, vol. 5, pp. 28–42, 06 2016.

[8] E. Fendri, I. Chtourou, and M. Hammami, "Gait-based person re-identification under covariate factors," *Pattern Analysis and Applications*, 02 2019.

[9] Z. Liu, Z. Zhang, and Q. Wu, "Enhancing person re-identification by integrating gait biometric," *Neurocomputing*, vol. 168, 05 2015.

[10] A. Nambiar, J. Nascimento, A. Bernardino, and J. Santos-Victor, "Person re-identification in frontal gait sequences via histogram of optic flow energy image," vol. 10016, 10 2016, pp. 250–262.

[11] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo Vega, P. Grother, and K. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, pp. 162–77, 03 2005.

[12] A. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters," vol. 1, 02 2001, pp. I–423.

[13] J. Han and B. Bhanu, ""individual recognition using gait energy image"," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, pp. 316–22, 03 2006.

[14] E. Zhang, Y. Zhao, and W. Xiong, "Active energy image plus 2dlpp for gait recognition," *Signal Processing*, vol. 90, pp. 2295–2302, 07 2010.

[15] K. Bashir, T. Xiang, and S. Gong, "Gait recognition using gait entropy image," vol. 2009, 01 2010, pp. 1 – 6.

[16] M. Rokanujjaman, M. Islam, M. A. Hossain, M. Islam, Y. Makihara, and Y. Yagi, "Effective part-based gait identification using frequency-domain gait entropy features," *Multimedia Tools and Applications*, vol. 74, 11 2013.

[17] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, pp. 167–174, 1993.

[18] S. Marxsen, D. Hilbert, R. Laubenbacher, B. Sturmfels, H. David, and R. Laubenbacher, *Theory of Algebraic Invariants*, ser. Cambridge Mathematical Library. Cambridge University Press, 1993. [Online]. Available: https://books.google.co.in/books?id=xCneO62aHiIC

[19] T. Suk and J. Flusser, "Graph method for generating affine moment invariants," vol. 2, 09 2004, pp. 192 – 195 Vol.2.

[20] T. Suk, "Tables of affine moment invariants generated by the graph method," 01 2005.

[21] ——, "Affine normalization of symmetric objects," 10 2005, pp. 100–107.

[22] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *ACM Multimedia*, 1996.

[23] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009.

[24] A. Abdel-Hakim and A. Farag, "Csift: A sift descriptor with color invariant characteristics," vol. 2, 02 2006, pp. 1978 – 1983.

[25] B. Pathak and D. Barooah, "Texture analysis based on the gray-level co-occurrence matrix considering possible orientations," 2013.

[26] E. Miyamoto and T. E. Merryman, "Fast calculation of haralick texture features," 2005.

[27] S. Zhang, Y. Huang, Y. Yu, H. Li, and D. Metaxas, "Automatic image annotation using group sparsity," 06 2010, pp. 3312–3319.

[28] B. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, pp. 703 – 715, 07 2001.

[29] J. Flusser, "On the independence of rotation moment invariants," *Pattern Recognition*, vol. 33, pp. 1405–1410, 2000.