World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:15, No:12, 2021

Gait Biometric for Person Re-Identification

Lavanya Srinivasan

***Abstract*—**Biometric identification is to identify unique features in a person like fingerprints, iris, ear, and voice recognition that need the subject's permission and physical contact. Gait biometric is used to identify the unique gait of the person by extracting moving features. The main advantage of gait biometric to identify the gait of a person at a distance, without any physical contact. In this work, the gait biometric is used for person re-identification. The person walking naturally compared with the same person walking with bag, coat and case recorded using long wave infrared, short wave infrared, medium wave infrared and visible cameras. The videos are recorded in rural and in urban environments. The pre-processing technique includes human identified using You Only Look Once, background subtraction, silhouettes extraction and synthesis Gait Entropy Image by averaging the silhouettes. The moving features are extracted from the Gait Entropy Energy Image. The extracted features are dimensionality reduced by the Principal Component Analysis and recognized using different classifiers. The comparative results with the different classifier show that Linear Discriminant Analysis outperform other classifiers with 95.8% for visible in the rural dataset and 94.8% for longwave infrared in the urban dataset.

***Keywords*—**Biometric, gait, silhouettes, You Only Look Once.

I. Introduction

THE most challenging problem is detecting humans in a video owing to variations in background, illumination, clothing, pose, body shape and appearance. Infrared cameras and changing background make it even harder.

The object appearance and shape are characterized by the distribution of local intensity gradients or edge directions. The gradients and edge directions are implemented by dividing the image window into cells, for each cell accumulating a local l-D histogram of gradient directions or edge orientations over the pixels of the cell. Contrast-normalize can be done by accumulating a measure of energy over blocks and using results to normalize all the cells in the block. The normalized descriptor blocks are referred as Histogram of Oriented Gradient (HOG) descriptors. Dalal et al. [1] describe that tiling the detection window with a dense grid of HOG descriptors and using the combined feature vector in a conventional SVM based window classifier gives human detection chain. Dalal et al. [2] build a detector combine gradient based appearance descriptors with differential optical flow-based motion descriptors in a linear SVM framework to detect human in a challenging environment.

Current object detection datasets are limited compared to datasets for other tasks like classification and tagging. The most common detection datasets contain thousands to hundreds of thousands of images with dozens to hundreds of tags [3]-[5].

Lavanya Srinivasan is with Electronics and Computer Science Department, University of Southampton, United Kingdom (corresponding author, e-mail: L.Srinivasan@soton.ac.uk).

Classification datasets have millions of images with tens or hundreds of thousands of categories [6], [5]. You Only Look Once (YOLO) [7], a real-time object detector, can detect over 9000 different object categories. Mask R-CNN [8] framework for object instance segmentation is simple and flexible. The framework includes instance segmentation, bounding box object detection and person key point detection. Framework detects objects in an image and generates a high-quality segmentation mask.

A static camera observing a region of interest is a common case for monitoring in a surveillance system. Detecting objects of the region of interest is an essential step in analyzing the scene. A statistical model of a scene exhibits some regular behavior. In background subtraction, pedestrians are detected in the scene when the full body exactly fitted in the model.

A Gaussian mixture model (GMM) was proposed for the background subtraction in [9] and efficient update equations are given in [10]. In [11], the GMM is extended with a hysteresis threshold. In GMM, the kernel method is much simpler, the processing time is less, and the segmentation is better than the traditional methods [12], [13]. The GMM gives a compact representation and a better model for simple static scenes.

ViBe is another method for background subtraction as proposed in [14]. This method requires a minimum memory compared to the other background subtraction technique, it compares the current pixel value with the neighborhood value to determine whether that pixel belongs to the background and remodel by substitute values from the background. Finally, the part of the background pixel value is propagated to the neighboring pixel of the background. In this work, gait recognition is done by extracting moving features using Gait entropy images, the features are dimensionality reduced using Principal Component Analysis and gait recognized using classifiers.

II. Methodology

The Gait video data are collected from two different locations representing urban and rural environments. The data were collected from volunteers of different ethnicity, religion, and a range of body forms from slim to fat. The participants, both men and women volunteers, wearing different clothing, shoes, coats, and bags, are considered for this analysis. Four different cameras, Long wavelength infrared (LWIR), Medium wavelength infrared (MWIR), Short-wavelength infrared (SWIR) and visible cameras are used for recording by walking along straight lines perpendicular to the camera view axis in the urban and rural environments. The rural data consist of 24

subjects, and the urban data consist of 31 subjects.

A. Preprocessing

There are three pre-processing steps: Human detection, Background subtraction and Silhouette's extraction.

1. Human Detection

The human-based detection uses HOG, YOLO and Mask Region Based CNN (RCNN). The YOLO based object detection outperforms other methods.

a. HOG

HOG is for object detection. The following steps are required to calculate HOG for an object:

1. Image normalization to reduce the influence of illumination effects.
2. Computing the gradient image in x and y to add further resistance to illumination variations.
3. Computing gradient histograms provides resistant to small changes in pose or appearance.
4. Normalizing across blocks provides better invariance to illumination, shadowing, and edge contrast.
5. Flattening into a feature vector.

b. You Only Look Once

A single convolutional neural network predicts bounding boxes, class labels and probabilities directly from full images in one evaluation. The main advantage of YOLO is it extremely fast and makes predictions that are comparatively better than traditional methods for object detection. YOLO makes less than half the number of background errors and false positive and negative compared to other methods. In YOLO, the detected box is bounded towards the object approximately as the same size as the object. The limitation of YOLO imposes strong spatial constraint and struggles to generalize aspect ratios or configurations to objects.

c. Mask R-CNN

Mask R-CNN is for semantic segmentation and extends Faster R-CNN for the bounding box recognition. Mask R-CNN detects objects and generates a segmentation mask for each instance.

The results of HOG, YOLO and Mask R-CNN are shown in Fig. 1. The bounding box of HOG is larger than the object, and false positive and false negative are comparatively higher than in YOLO. Mask R-CNN segmentation mask gives the rectangular effect. With a compact bounding box around the object, YOLO outperforms with a smaller number of false positive and false negative.

1.2. Background Subtraction

The background subtraction was performed to check the quality of the image using GMM and ViBe methods. The results of both the methods are shown in Fig. 2. The figure shows that the ViBe results are comparatively better with less artefacts and clutter than GMM.

1.3. Silhouettes Extraction

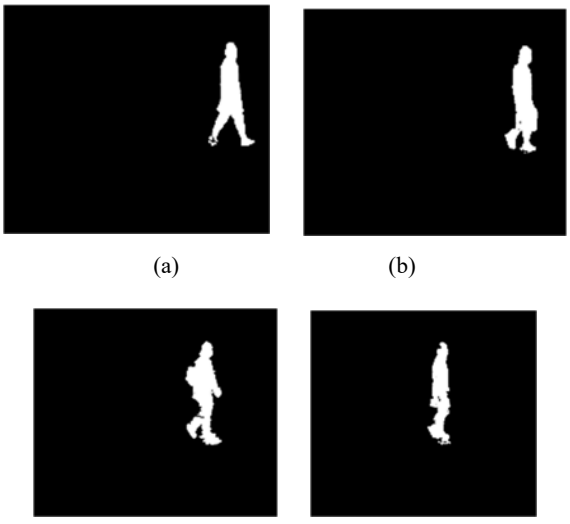
Each subject is divided into four groups normal, coat, bag, and suitcase. The silhouettes for normal data consist of 12 sequences, six sequences of walking from left to right and six sequences of walking from right to left. The coat, bag, and suitcase data consist of four sequences, two sequences of walking from left to right and two sequences of walking from right to left. In this work, left to right walking sequences are considered for gait analysis. The silhouette data are divided into training and testing. The training data consist of four sequences of normal silhouette, the testing data consist of two sequences of normal and two sequences coat, bag and suitcase left to right walking silhouettes. The extracted silhouettes are shown in Fig. 3.

(a) (b) (c)

Fig. 1 Human detection using (a) HOG (b) YOLO (c) Mask RCNN

(a) (b)

Fig. 2 Background Subtraction (a) GMM (b) ViBe



(a) (b)

(c) (d)

Fig. 3 Silhouette Extraction (a) normal, (b) carrying suitcase, (c) carrying bag, (d) wearing coat

B. Gait Recognition

The gait of the person is recognized using Gait Entropy Image. Gait Entropy Image based on computing entropy, encodes in a single image the randomness of pixel value in the silhouette images over a complete gait cycle. The dynamic parts of the image give high gait entropy value and the static parts remain low value. The Gait Entropy Image captures dynamic information and remains robust to covariate changes that effect the static information of human body.

1. Gait Entropy Based Method

In [15], gait entropy image (GEnI) is computed from normalized silhouettes. The silhouettes are extracted using the proposed method and height of the silhouettes are normalized followed by Centre alignment. Given a gait cycle of size-normalized and center-aligned silhouettes, a GEnI is computed by calculating Shannon entropy for each pixel in the silhouette images.

Entropy over a completed gait cycle is calculated as:

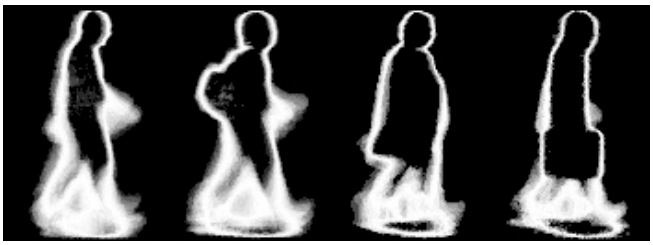
$$H(x,y) = -\sum_{k=1}^{K} p_k(x,y) log_2 p_k(x,y) \quad (1)$$

where x,y are pixel coordinates $p_k(x,y)$ is the probability that the pixel takes on the kth value. GEnI$G(x,y)$ can be obtained by scaling and discretising $H(x,y)$ so the value ranges from 0 to 255.

$$G(x,y) = \frac{(H(x,y) - H_{min}) * 255}{H_{max} - H_{min}} \quad (2)$$

where $H_{min} = \min(H(x,y))$ and $H_{max} = \max(H(x,y))$

Fig. 4 shows some examples of GEnI from our gait dataset. It clearly shows that the dynamic area of the human body, including legs and arms which undergo motions in relation to other body parts, are represented by higher intensity values.



(a) (b) (c) (d)

Fig. 4 GEnI (a) Normal (b) Bag (c) Coat (d) Briefcase.

2. Principal Component Analysis

Principal Component Analysis (PCA) [16] reduces data by geometrically projecting them from higher dimension to lower dimensional features. PCA by projecting simplifies the complexity in high-dimensional data while retaining trends and patterns. The gait sequences are represented as GEnI, gait recognition can be performed by matching testing dataset to the training dataset that has the minimal distance to the testing GEnI. PCA projects the original features to the subspace of the lower dimensionality so best data representation and class separability can be achieved simultaneously. The reduced dimension features are used for gait recognition by using classifiers.

3. Classifiers

In this work, classifiers such as K-Nearest Neighbor, Random Forest, Naive Bayes, Linear Discriminant, Support Vector Machine and Linear regression are analyzed for recognition.

a. K-Nearest Neighbor

The K-Nearest Neighbor (K-NN) classifier [17] is based on the class of their nearest neighbors considering more than one neighbor. Classification is based directly on the training examples and the Memory-Based Classification needs to be in the memory at run-time during the training process.

b. Random Forest (RF)

Random forest approach, a machine learning technique, was first proposed by Breiman [18] by combining classification and regression tree [19] and bagging [20]. Briefly, in a random forest, prediction is obtained by averaging the results of classification and regression trees that are grown on bootstrap samples. Thus, when growing a tree, training data are divided into a bootstrap sample data and out-of-bag (OOB) data, and cross validation is possible in random forest by using the OOB data.

c. Naive Bayes (NB)

The Naive Bayes classifier [21] greatly simplifies learning by assuming that features are independent.

$$P(X/C) = \prod_{i=1}^{n} P(X_i/C) \quad (3)$$

where $X = (X_1, \ldots X_n)$ is a feature vector and C is a class.

d. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is used for dimensionality reduction and classification. LDA [22] tries to maximize the ratio of the between-group variance and the within-group variance. When the ratio is maximum, the scatter between the group is small and the groups separated from one another the most.

The ratio between and within class variance is given as:

$$S = \frac{w^T s_b w}{w^T s_w w} \quad (4)$$

where, s_b and s_w are between and within group variance, S is the Scatter Matrix and w is the eigen vector.

e. Support Vector Machine (SVM)

SVM [23] are a set of supervised learning methods used for classification and regression problems. SVM is effective in high dimensional spaces. Hyperplane separates the two classes to generalize to new data and make accurate classification predictions.

f. Logistic Regression (LR)

LR [24] models the probabilities for classification problems with two possible outcomes. The LR model uses the sigmoid

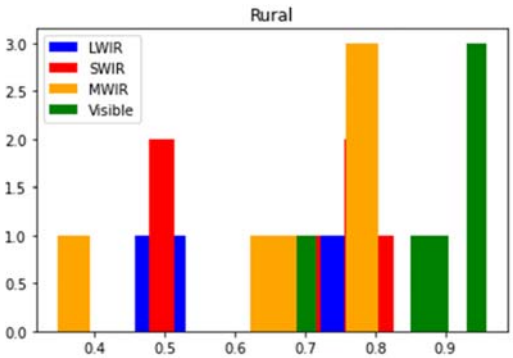
function to squeeze the output of a linear equation between 0 and 1. In particular, an input producing an outcome greater than the threshold is considered to belong to class 1. The output is less than the threshold, the corresponding input is classified as belonging to the 0 class.

TABLE I
THE EXPERIMENT RESULTS OF GENI

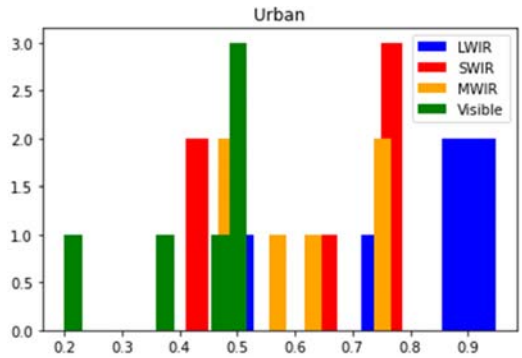
Classifier	Subject	LWIR		SWIR		MWIR		Visible	
		Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
K-NN	Normal vs. Normal	0.931	0.729	0.786	0.783	0.776	0.783	0.500	0.896
	Normal vs. Bag	0.288	0.354	0.768	0.565	0.303	0.457	0.655	0.022
	Normal vs. Case	0.317	0.109	0.554	0.045	0.379	0.261	0.283	0.045
	Normal vs. Coat	0.300	0.354	0.554	0.348	0.379	0.30	0.224	0.000
RF	Normal vs. Normal	0.741	0.667	0.643	0.717	0.776	0.696	0.483	0.854
	Normal vs. Bag	0.203	0.208	0.464	0.413	0.089	0.239	0.466	0.065
	Normal vs. Case	0.200	0.087	0.411	0.114	0.293	0.283	0.133	0.023
	Normal vs. Coat	0.217	0.271	0.268	0.239	0.155	0.196	0.241	0.023
NB	Normal vs. Normal	0.482	0.458	0.446	0.478	0.569	0.348	0.383	0.688
	Normal vs. Bag	0.033	0.041	0.196	0.087	0.180	0.130	0.134	0.022
	Normal vs. Case	0.100	0.065	0.143	0.045	0.100	0.174	0.000	0.023
	Normal vs. Coat	0.050	0.125	0.125	0.109	0.155	0.109	0.155	0.045
LDA	Normal vs. Normal	0.879	0.521	0.411	0.500	0.466	0.652	0.200	0.938
	Normal vs. Bag	0.254	0.083	0.214	0.130	0.214	0.283	0.207	0.065
	Normal vs. Case	0.167	0.109	0.107	0.045	0.155	0.196	0.067	0.045
	Normal vs. Coat	0.133	0.104	0.161	0.152	0.069	0.196	0.138	0.000
SVM	Normal vs. Normal	0.862	0.750	0.768	0.826	0.483	0.761	0.500	0.938
	Normal vs. Bag	0.169	0.354	0.768	0.543	0.375	0.435	0.655	0.065
	Normal vs. Case	0.300	0.130	0.553	0.068	0.240	0.304	0.283	0.045
	Normal vs Coat	0.267	0.271	0.428	0.326	0.310	0.326	0.293	0.023
LR	Normal vs. Normal	0.948	0.813	0.786	0.783	0.621	0.804	0.517	0.958
	Normal vs. Bag	0.271	0.333	0.750	0.522	0.303	0.478	0.689	0.065
	Normal vs. Case	0.367	0.174	0.500	0.682	0.276	0.239	0.167	0.068
	Normal vs Coat	0.3	0.354	0.5	0.261	0.275	0.391	0.276	0.023

III. RESULT AND DISCUSSIONS

The experimental results of the GEnI are shown in Table I. The classification accuracy for normal data is comparatively higher compared to bag, coat, and briefcase video sequences. The histogram images of rural and urban are shown in Fig. 5. In the rural dataset, LR for visible data recognizes with the highest accuracy of 95.8%. In the urban dataset, LR for LWIR data recognizes with the highest accuracy of 94.8%. Linear regression outperforms in recognition compared to the other classifiers.



(a)



(b)

Fig. 5 Histogram analysis of Normal subject: (a) Rural, (b) Urban

IV. CONCLUSION

Video surveillance plays a major important role and acts as a part of everyone's life for security reasons. In public places, identifying a person in different cameras is a challenge when those individual changes their appearance. This paper proposes that the gait biometric of a person can be identified at a distance. This biometric measure can identify a person even with changes in their appearance based on their gait (walking). In this work, person re-identification is analyzed using gait moving feature extraction. The features extracted from the GEnI are

dimensionality reduced using PCA and recognized using different classifiers. In the future, the work could be extended using soft biometric features with traditional biometric features.

REFERENCES

[1] N. Dalal, B. Triggs, "Histogram of Oriented Gradients for Human Detection", IEEE international conference on computer vision and pattern recognition (CVPR), pp. 886-893, 2005.
[2] N. Dala1, B. Triggs, C. Schmid, "Human Detection Using Oriented Histograms of flow and appearance", European Conference on Computer Vision (ECCV), pp. 428-441, May 2006.
[3] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge", International journal of computer vision, pp.88(2) :303– 338, 2010.
[4] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Com- ´ mon objects in context", In European Conference on Computer Vision, pp. 740–755, Springer, 2014.
[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei, "Imagenet: A large-scale hierarchical image database", In Computer Vision and Pattern Recognition, IEEE Conference on CVPR, pp.248–255, 2009.
[6] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. "Yfcc100m: The new data in multimedia research", Communications of the ACM, pp.59(2):64–73, 2016.
[7] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition, 2016.
[8] Kaiming He, Georgia Gkioxari, Piotr Dollar and Ross Girshick, "Mask R-CNN", Computer Vision and Pattern Recognition, 2018.
[9] Friedman, N., Russell, S., "Image segmentation in video sequences: a probabilistic approach", In: Proc. 13th Conf. on Uncertainty in Artificial Intelligence, 1997.
[10] Stauffer, C., Grimson, W., "Adaptive background mixture models for real-time tracking", In: Proc. of the Conf. on Computer Vision and Pattern Recognition. pp. 246–252, 1999.
[11] Power, P.W., Schoonees, J.A., "Understanding background mixture models for foreground segmentation", In: Proc. of the Image and Vision Computing New Zealand, 2002.
[12] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel. "Towards robust automatic traffic scene analysis in real-time", In Proc. of the International Conference on Pattern Recognition, Israel, November 1994.
[13] Christof Ridder, Olaf Munkelt, and Harald Kirchner "Adaptive Background Estimation and Foreground Detection using Kalman-Filtering," Proceedings of International Conference on recent Advances in Mechatronics, ICRAM'95, UNESCO Chair on Mechatronics, pp. 193-199, 1995.
[14] O. Barnich, M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences", IEEE Transactions on Image Processing, 20(6), pp. 1709-1724, June 2011.
[15] K. Bashir, T. Xiang, S. Gong, "Gait recognition without subject cooperation", Pattern Recognition Letters, 31(13), pp. 2052-2060, October 2010.
[16] JOLLIFFE, I.T., "Principal Component Analysis", second edition, New York: Springer-Verlag New York, 2002.
[17] Padraig Cunningham, Sarah Jane Delany, "k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)", Machine Learning, 2020.
[18] Leo Breiman, "Random Forests", Machine Learning, volume 45, pp.5-32,2001.
[19] Leo Breiman, Jerome Friedman, Charles J. Stone, R.A. Olshen, "Classification and Regression Trees", Taylor & Francis, Mathematics, pp. 368, 1984.
[20] Leo Breiman, "Bagging Predictors", Machine Learning, volume 24, pp. 123–140, 1996.
[21] I. Rish, "An empirical study of the naive Bayes classifier", 2001.
[22] Ian H. Witten, EibeFrank, Mark A.Hall, Christopher J.Pal, "Data transformations", Data Mining (Fourth Edition), Practical Machine Learning Tools and Techniques, pp. 285-334, 2017.
[23] S. Gunn, "Support Vector Machines for Classification and Regression", Mathematics, 1998.
[24] N. H. Bingham, John M. Fry, "Regression Linear Models in Statistics", Springer Undergraduate Mathematics Series, 2010.