Improved Gait Recognition Accuracy Based on DFT-GEI

Lavanya Srinivasan (drlavphd@gmail.com)
University of Southampton

Research

Keywords: Biometric, Gait, Silhouettes, Infrared, Reidentification.

Posted Date: June 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-655061/v1

Improved Gait Recognition Accuracy Based on DFT- GEI

Lavanya Srinivasan
University of Southampton, United Kingdom
drlavphd@gmail.com

Abstract- Person reidentification is a challenging task in computer vision. Identify a person from different cameras due to changes in appearance based on cofactors. Cofactors such as changing clothes, a suitcase, backpack, etc. The gait biometric is used to identify a person vary with different cofactors at different backgrounds. The person's gait can be identified at a distance, based on a walking pattern, without any physical contact. In this work, the videos are recorded using Long Wave Infrared and Visible cameras at different locations such as urban and rural environments. The pre-processing includes the recorded videos are converted into frames, person identification using deep learning techniques, background subtraction, artifacts removal, silhouettes extraction, calculating gait cycle, and synthesis frequency domain gait energy image by averaging the silhouettes. The moving features are extracted from the frequency domain gait energy image and gait energy image are dimensionally reduced by principal component analysis, recognized using classifier K nearest neighbour and results are compared. Experiments are conducted on urban and rural datasets recorded using Long Wave Infrared and Visible cameras.

Keywords—*Biometric, Gait, Silhouettes, Infrared, Reidentification.*

I. INTRODUCTION

Video surveillance is to track people across a network of cameras to detect abnormal behavior. Recognizing a person by gaits is more popular due to identify at a distance from the low-resolution videos or images, without the cooperation of individuals, without any physical contact with instruments such as faces and fingerprints requires physical contact with instruments. Gait can be recognized by moving features from a distance while other features such as faces, ears, and fingerprints are hidden. Gait features are typically difficult to be caricatured. The most challenging problems in gait recognition in a video, owing to variations in background, changes in daylight illuminations, different cofactors, body shapes, person's pose and appearances.

Gait identification plays a major role in video-based wide-area surveillance [1,2] in finding terrorists in airports, stations, car parking, banks, crowd gathering places, law enforcement to identify criminals, detecting health disorders such as identifying the early stage of Parkinson's and in the sports training to provide optimal training strategies.

Thermal imaging is a benefit to armed forces such as the army, navy and air force. The border surveillance and law enforcement work in all weather conditions and day-night, they use thermal detectors to capture the infrared cameras. The infrared cameras capture the radiation emitted from objects, which are above absolute zero temperature. Thermal imaging is mainly used to locate moving objects, recognize, target and differentiate from the own to enemy forces. Thermal imaging, due to its various advantages, has many applications in the military and defence [3].

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 1-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. [4] 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. [5] build a detector combine gradient based appearance descriptors with differential optical flow-based motion descriptors in a linear Support Vector Machine (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 [6]-[8]. Classification datasets have millions of images with tens or hundreds of thousands of categories [9], [8]. You Only Look Once (YOLO) [10], a real-time object detector, can detect over 9000 different object categories. Regions with convolutional neural networks (Mask R-CNN) [11] 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 [12] and efficient update equations are given in [13]. In [14], 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 [15], [16]. The GMM gives a compact representation and a better model for simple static scenes.

Visual Background Extractor (ViBe) is another method for background subtraction as proposed in [17]. 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 and frequency domain gait entropy, the features are dimensionality reduced using Principal Component Analysis (PCA) [18] and gait recognized using K-Nearest Neighbor (K-NN) [19-20].

II. METHODOLOGY

The Gait video data collected from two different locations in the 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 were both men and women volunteers were wearing different clothing, coats, case, and backpack are considered for this analysis. Walking along straight lines perpendicular to the camera view axis in the urban and rural environments are recorded using Longwave infrared (LWIR) and Visible cameras. The rural data consists of 24 subjects and the urban data consists of 31 subjects. Two walking data sequences, Right to Left and Left to Right. In this work, we considered Left to Right walking data sequences.

A. Preprocessing

The frames are extracted from videos. NVIDIA GeForce RTX 2060, Processor Intel (R) Core (TM) i7-10750H CPU @ 2.60GHz, 2592 MHz, 6 Core(s), 12 Logical Processor(s), is used to conduct experiments.

2.1 Human detection

The algorithms used for human-based detection are HOG, YOLO and Mask-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.

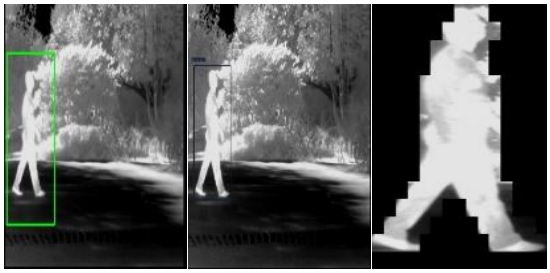
b. You Only Look Once

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 positives and negatives compared to other methods. In YOLO, the detected box is bounded towards the object approximately the same size as the object. The limitation of YOLO imposes strong spatial constraints 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. False positives and false negatives are relatively higher than YOLO. Instance segmentation of Mask R-CNN results in a rectangular effect. YOLO outperforms other methods with a smaller number of false positives and false negatives with a compact bounding box around the object.

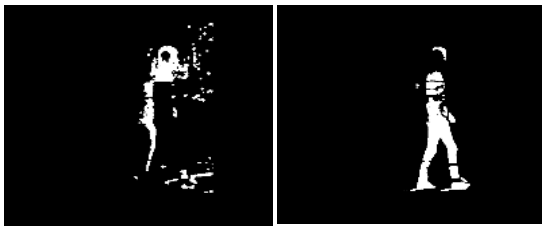


(a) (b) (c)

Fig. 1 Human detection using (a) HOG (b) YOLO (c) Mask R-CNN

2.2. Background Subtraction

The background subtraction was 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 fewer artifacts and clutter than GMM.



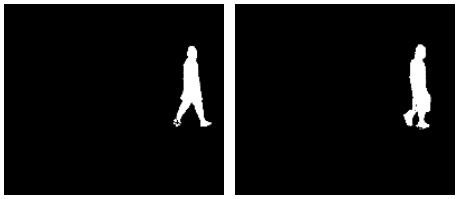
(a) (b)

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

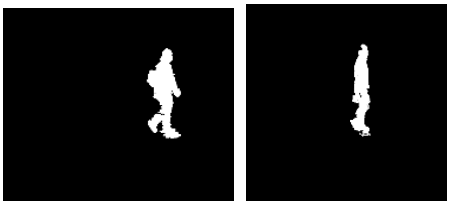
2.3 Silhouettes Extraction

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 sets.

The training data set consists of four sequences of normal silhouettes. The testing data sets consist of two sequences of normal, coat, bag, and suitcase. The silhouettes are shown in Fig. 3.



(a) (b)

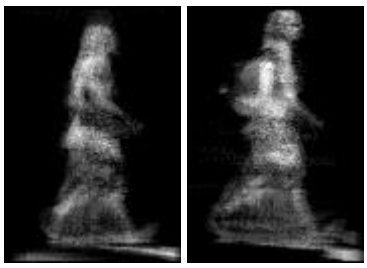


(c) (d)

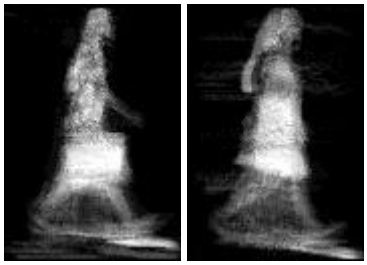
Fig. 3 Silhouette Extraction (a) normal, (b) carrying suitcase, (c) carrying bag, (d) wearing coat

2.4 Gait Energy Image

The Spatio-temporal silhouettes are averaged over the Gait cycle to calculate Gait Energy Image (GEI). The GEI are shown in Fig. 4.



(a) (b)



(c) (d)

Fig. 4 Gait Energy Image (a) normal, (b) carrying bag, (c) carrying suitcase, (d) wearing coat

Gait Energy Image is defined as,

$$GEI(x,y) = \frac{1}{N}\sum_{n=1}^{N} B(x,y,n) \quad (1)$$

where $B(x,y,n)$ pre-processed binary gait silhouette, N is the number of frames in a gait cycle, n is the frame number and x and y values are the 2D image coordinates.

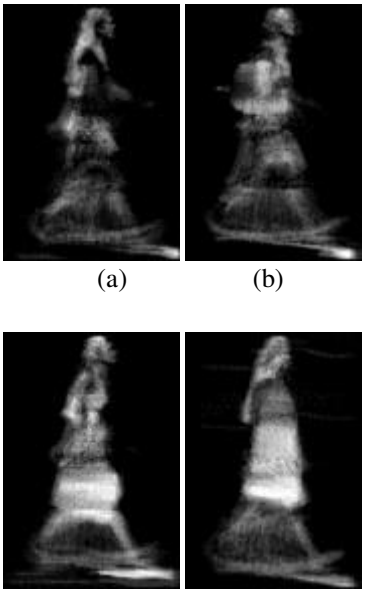
2.5 Discrete Fourier Transform

The amplitude spectra of Gait Silhouette Volume (GSV) are calculated by Discrete Fourier Transform (DFT) analysis based on the gait period.

$$G(x,y,k) = \sum_{n=0}^{N-1} B(x,y,n) exp^{-j\omega_0 kn} \quad (2)$$

$$A(x,y,k) = \frac{1}{N}|G(x,y,k)| \quad (3)$$

where $A(x, y, k)$ is amplitude for temporal axis, N is the number of frames in a gait cycle, ω_0 is a base angular frequency for a gait cycle and k is the frequency component. The DFT analysis of gait period is shown in Fig. 5.



(a) (b)

(c) (d)

Fig. 5 DFT analysis (a) normal, (b) carrying bag, (c) carrying suitcase, (d) wearing coat

2.6 Principal Component Analysis

Principal Component Analysis 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 GEI and DFT-GEI, gait recognition can be performed by matching testing dataset to the training dataset that has the minimal distance to the testing GEI and DFT-GEI. The 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.

III. CLASSIFIER

In this work, classifier K-Nearest Neighbour are analysed for recognition.

3.1 K-Nearest Neighbour

The K-Nearest Neighbour classifier 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.

IV. RESULTS AND DISCUSSIONS

The experimental results of the DFT-GEI are shown in Table I. The classification accuracy for normal data is comparatively higher compared to bag, coat, and briefcase video sequences. In the urban dataset, K-NN for visible data recognizes with the highest accuracy of 98% and LWIR data recognizes with the highest accuracy of 91%. The result shows urban data outperforms compared to rural data. The experimental results of the GEI are shown in Table II. In the rural dataset, K-NN for visible data recognizes with the highest accuracy of 89%. In the urban dataset, K-NN for LWIR data recognizes with the highest accuracy of 86%.

TABLE I Experimental results of DFT- GEI

Classifier	Subject	LWIR		Visible	
		Urban	Rural	Urban	Rural
KNN	Normal vs Normal	0.91	0.75	0.98	0.94
	Normal vs Bag	0.33	0.33	0.34	0.85
	Normal vs Case	0.43	0.35	0.28	0.5
	Normal vs Coat	0.16	0.48	0.12	0.41

TABLE II Experimental results of GEI

Classifier	Subject	LWIR		Visible	
		Urban	Rural	Urban	Rural
KNN	Normal vs Normal	0.86	0.73	0.50	0.89
	Normal vs Bag	0.29	0.35	0.26	0.02
	Normal vs Case	0.32	0.11	0.28	0.04
	Normal vs Coat	0.20	0.35	0.22	0.00

V. 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. In this work, person re-identification is analyzed using gait moving feature extraction. The features extracted from the GEI and DFT-GEI are dimensionality reduced using PCA and recognized using K-NN classifier. The DFT-GEI recognition

rate is higher compared to GEI. In the future, the work could be extended using soft biometric features with traditional and part based biometric features.

Author's Contribution

Lavanya wrote the manuscript, revised, and edited.

Author's information

Lavanya is a research fellow at the University of Southampton, United Kingdom. She has 6 years of experience in various computer vision projects. Her research areas of interests include biometrics, image processing and medical image processing.

Funding

Not applicable.

Availability of data and materials

Materials used in the manuscript may be requested from the corresponding author.

Competing interests

The author declare that she has no competing interests and that there is no conflict of interest regarding the publication of this manuscript.

Author details

Lavanya Srinivasan, Research Fellow, School of Electronics and Computer science, University of Southampton, Southampton, United Kingdom, SO17 1BJ.

REFERENCES

[1] H Lu, K. Plataniotis and A. Venetsanopoulos , "Uncorrelated multilinear discriminant analysis with regularization for gait recognition", In: Proceedings of biometrics symposium. Baltimore, pp 1–6,2007

[2] E Rogers, "Gait recognition", In: Bell Canada chair in multimedia IPSI: Identity, Privacy and Security Initiative. Department of Electrical and Computer Engineering, University of Toronto.

[3] A. Aparna, G. Ripul and H K Sardana, 'Thermal Imaging and Its Application In Defence Systems', AIP Conf Proc, Vol. 1391, 2011.

[4] N. Dalal and B. Triggs, "Histogram of Oriented Gradients for Human Detection", IEEE international conference on computer vision and pattern recognition (CVPR), pp. 886-893, 2005.

[5] N. Dala1, B. Triggs and C. Schmid, "Human Detection Using Oriented Histograms of flow and appearance", European Conference on Computer Vision (ECCV), pp. 428-441, May 2006.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Joseph Redmon and Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition, 2016.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollar and Ross Girshick, "Mask R-CNN", Computer Vision and Pattern Recognition, 2018.

[12] N. Friedman and S. Russell, "Image segmentation in video sequences: a probabilistic approach", In: Proc. 13th Conf. on Uncertainty in Artificial Intelligence, 1997.

[13] C. Stauffer, W. Grimson, "Adaptive background mixture models for real-time tracking", In: Proc. of the Conf. on Computer Vision and Pattern Recognition. pp. 246–252, 1999.

[14] P.W. Power, J.A. Schoonees, "Understanding background mixture models for foreground segmentation", In: Proc. of the Image and Vision Computing New Zealand, 2002.

[15] 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.

[16] 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.

[17] O. Barnich and M. Van Droogenbroeck, “ViBe: A universal background subtraction algorithm for video sequences”, IEEE Transactions on Image Processing, 20(6), pp. 1709-1724, June 2011.

[18] I.T. JOLLIFFE, “Principal Component Analysis”, second edition, New York: Springer-Verlag New York, 2002.

[19] Padraig Cunningham and Sarah Jane Delany, “k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)”, Machine Learning, 2020.

[20] Marcin Derlatka, “Modified kNN Algorithm for Improved Recognition Accuracy of Biometrics System Based on Gait”, Computer Information Systems and Industrial Management, pp. 59-66, 2013.