Analysis of Human Gait for Person Identification and Human Action Recognition

Deepak N.A.
Research Scholar
National Aerospace Laboratories (NAL)
Bangalore, India

Sinha U.N.
Former Distinguished Scientist
National Aerospace Laboratories-CMMACS
Bangalore, India

ABSTRACT

The gait sequence analysis finds applications in several fields such as, person identification, human action recognition and event classification. The extraction of gait sequences is the initial step in gait sequence analysis. The Latent Dirichlet Allocation (LDA) a type of 'Topic Model' is used to analyse the gait sequences for person identification and human action recognition. This is achieved by proposing a novel method that transforms the gait sequences into words representation. Finally, the generated words are then analysed using LDA for person identification and human action recognition. The proposed person identification algorithm is tested using CASIA gait dataset A and human action recognition algorithm is tested using Weizmann action recognition dataset, resulting in correct classification rate of 85% using CASIA Dataset A and 85% using Weizmann action recognition dataset.

General Terms

Gait Domain, Text Domain

Keywords

Binary Silhouettes, Gait Analysis, Images-Topics-Words, Transformation, Latent Dirichlet Allocation

1. INTRODUCTION

The application of gait analysis in multiple domains requires analysing the extracted gait features. These features are extracted from input binary silhouettes. The topic models developed for text classification is one of the potential approaches to use in gait domain framework, such as person identification and human action recognition. Some of the existing methods to identify person and recognize human actions uses different methods or techniques, to name few, the boundary centroid distance in [1] is used as a gait feature for identification. The gait features like, step length, stride length, cadence, cycle length, is used along with body height, weight and gender are used for recognizing individuals [2, 3, 4, 5]. The spatio temporal local features in [6] are integrated with support vector machine, for human action recognition. The probabilistic latent semantic indexing in [7] uses unsupervised learning method to recognize human actions. Gait biometric cues in [8] and human body movements in [9] are also used to recognize individuals. The study of text analysis, uses 'Topic Models' in text domain framework, here an attempt is made to use topic model in gait domain framework. The major crux in use of topic models in gait domain framework is the lack of textual concepts. This is resolved by proposing a novel transformation method that transforms the input documents into word representation. The Fig 1, shows the block diagram of the proposed algorithm, that specifies the steps used to arrive at the results.

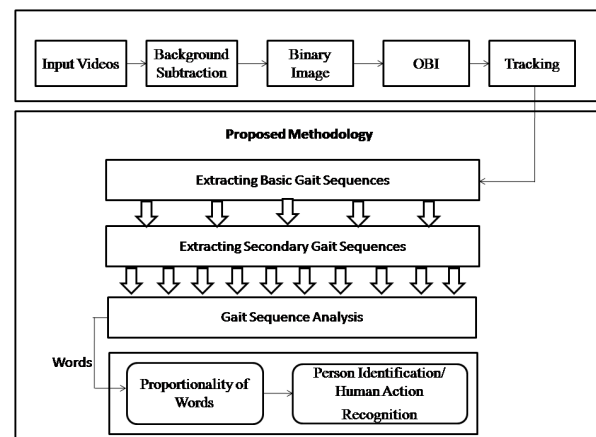The rest of the paper is structured as follows: Section 2 gives the overview of methods used for person identification and human action recognition, Section 3 specifies the steps of the proposed algorithm and Section 4 discusses the results, followed by the conclusion and references.

Fig. 1. Block Diagram of the Proposed Methodology

2. RELATED WORKS

This section deals with the overview of the existing methods and algorithms that are currently being used or investigated for recognizing the human actions. An arbitrary view transformation model in [10] is used to identify individuals. The body shape information and their movements in [11] are extracted as a gait features for identification. The multiple gait features in [8, 12, 13, 14] are extracted, smoothed and normalized using principal component analysis. Finally, a supervised pattern classification technique is employed for person identification. The template matching based on supervised learning method [15] uses the features, extracted from the silhouettes for real-time human action recognition. The human action recognition in [16] uses dynamic texture descriptors. These descriptors are used to describe human motions in a spatio temporal manner for recognizing actions. The sparse representations for human action recognition in [17] uses the models of human actions and its corresponding sparse representation found using spatio-temporal descriptors extracted from the input documents. Latent dirichlet allocation in [18] specify the topic model, that has been used for identifying topics and concepts in textual data, which are then analysed for identification and recognition. In the proposed algorithm, latent dirichlet allocation (LDA) a 'Topic Model' is used to identify person and recognize human actions. This is achieved by proposing a novel transformation method that transforms the gait sequences into document-word representation. Finally LDA uses these words to identify person and recognize human actions.

3. PERSON IDENTIFICATION AND HUMAN ACTION RECOGNITION - PROPOSED METHODOLOGY

The 'Topic Model' like LDA is applicable in gait domain framework, only if, the concepts of image-words are mapped appropriately. The words are generated from the input documents, (where document refers to an input video hereafter). The words are the basic building blocks, which are used in the proposed algorithm for person identification and human action recognition. The steps to generate words involves i) Extraction of primary gait sequences from the input documents ii) Deriving sub-primary gait sequences from basic gait sequences and finally, v)Analysing the sub-primary gait sequences to generate words.

3.1 Extraction of Primary Gait Sequences (Δ) and Sub-Primary Gait Sequences (Υ)

The extraction of primary gait sequences (Δ) is an initial and basic step in the proposed algorithm. This is achieved by enclosing the pixels that corresponds to a person within the bounding box. The Enclosed pixels are then tracked over the binary frames (B_I) to extract primary gait sequences such as: 1) Horizontal-Movement (Δ_{HM}), 2) Vertical-Movement (Δ_{VM}), 3) Pixel-Count (Δ_{PC}), 4) Left-Area-Counter (Δ_{LAC}) and 5) Right-Area-Counter (Δ_{RAC}). The gait sequences horizontal-movement (Δ_{HM}) and vertical-movement (Δ_{VM}) represents the measure of the distance moved by an individual during an action. The (Δ_{HM}) is measured horizontally and (Δ_{VM}), is measured vertical. The gait sequence pixel-count (Δ_{PC}) represents, the total number of white coloured pixels within an binary image frame. The gait sequences (Δ_{LAC}) and (Δ_{RAC}), are extracted as the measure of change in gait-area of the bounding box (rectangle), labelled as (A,B,C,D) and (B,E,F,C) respectively, as given in Fig 2.

The performance of the proposed LDA algorithm is tested with the extracted primary gait sequences. It is found that, the performance reduces when executed along with the primary gait sequences. This is because the LDA will not be so effective, if the words generated in its vocabulary are small, as this would influence the topic space and its relation to words generation. In order to increase the performance of proposed LDA algorithm, the sub-primary gait sequences (Υ) are extracted from the primary gait sequences using above-below standard deviation sequencing technique. The sub-primary gait sequences increases the number of gait sequences extracted, which in-turn increases the vocabulary space of LDA related to gaits, in-turn increasing the correct classification rate.

3.1.1 Above-Below Standard Deviation. In this technique, the standard deviation ($\delta_{(\Delta_i)}$) is found, by considering each elements of the selected primary gait sequence ($\Delta_i(E_j)$). The standard deviation ($\delta_{(\Delta_i)}$) value is then subtracted from each elements of the same sequence, the sign of the element is also considered in calculating above-below standard deviation sequence as shown in Equation (1).

$$\prod(i) = sign[|(\Delta_i(E_j)) - \delta_{(\Delta_i)}| \geq 0]$$ (1)

3.2 Generation of Words

The words for LDA are generated by transforming the extracted sub-primary gait sequences (Υ) into word representation. The steps for generating the words are given below. These steps are executed iteratively by considering other elements of the same sequence to generate other sets of words. This process is repeated with each sub-primary gait sequence (Υ_i) to populate the word count. Finally, the generated words will be analysed by LDA to generate two matrices as output, i) Word probability matrix or
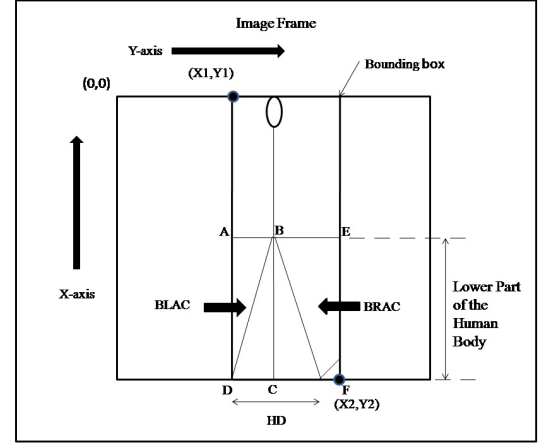

Fig. 2. Region of Interest

word-topic distribution matrix (β) and ii) Topic probability matrix or topic-document distribution matrix (α).

(1) The word (w_1) will be incremented by one, if the first two elements of the selected sequence are equal.

(2) Similarly, the word (w_2) will be incremented, if the other two elements of the selected sequence are equal.

(3) The word(w_3) will be incremented, if summation value of first two elements and second two elements of the selected sub-primary sequence are equal.

(4) The word (w_4) will be incremented, if the difference of first two elements is zero or the difference of second two elements is zero or if the sign of the difference of the first two elements and second two elements are equal.

3.3 Calculating the Words Proportionality

The word probability matrix or word-topic distribution matrix (β) of LDA, is used in the proposed algorithm to find the words proportionality. Each element of the matrix (β) is analysed as a product of two normalized factors as shown in the Equation (2).

$$\boxplus(i,j) = \frac{\nabla(i)}{\sum_{k=1}^{N} \nabla(k)} * \frac{\beta(i,j)}{\sum_{i=1}^{M} \lambda(i)} .$$ (2)

Where ($\boxplus$) represents the proportionality of words, which is found as the product of two normalized factors.

3.4 Person Identification and Human Action Recognition

In this section, the person is identified and human actions are recognized from the input video documents. The input videos also known as document are segmented initially into image frames. The binary images are constructed and the object/person needs to be identified is selected and tracked over the segmented frames to extract the primary gait sequences. As the performance of LDA reduces with the extracted primary gait sequences, the sub-primary gait sequences gains importance in classification. The sub-primary gait sequences increase the classification rate by increasing the number of words. Finally, the generated words are analysed by LDA to generate two matrices as output, that is (α) and (β).

3.4.1 Person Identification. In order to identify person, the proposed algorithm divides the input documents into two groups,

namely test and training. The proposed algorithm is then executed independently on both these groups to generate matrix (β) and (α) as the output of LDA from the training set of documents. Whereas the words and its occurrences are found from the test set. In order to identify the person, the proposed algorithm selects, some of the words from the test set, based on the mean value ($\mu_{test)}$), where the mean (μ_{test}) is found using the word occurrences of the test document set. For each word, the corresponding topic is found using the matrix (β). Then, for each topic, the corresponding document is found, using the matrix (α). The document selected, would be the maximum value found with respect row (k^{th}) row of the matrix (α), where (k) is the selected topic index. Finally, the person is identified by selecting a single document as a result of the proposed algorithm, which is the maximum summation value of the selected documents in the matrix (α).

3.4.2 Human Action Recognition. In order recognize human actions the proposed algorithm uses quality threshold clustering. The quality threshold clustering algorithm groups the input documents into clusters based on the human action performed. The quality threshold clustering is an alternative method of portioning the data. This requires more computation power in order to cluster the individuals. The working of quality threshold clustering is given below:

(1) Initially, choose the maximum diameter for the cluster

(2) Build a candidate cluster for each point, by including the closest point, the next closest and so on until the diameter of the cluster surpasses the threshold.

(3) Save the candidate cluster with most points, as the first true cluster, and remove all the points in cluster from further considerations.

(4) Recurs with reduce set of points

4. RESULTS AND DISCUSSION

The proposed person identification algorithm has been illustrated using CASIA gait Dataset A [13], and human action recognition algorithm is tested using Weizmann action recognition dataset [20]. These two are the largest datasets available as a benchmark for person identification and human action recognition. The brief overview of these datasets is provided below.

4.1 CASIA Gait Dataset A

The CASIA gait dataset contains 4 subsets i) Dataset A, ii) Dataset B, iii) Dataset C and iv) Dataset D. The gait dataset A contains people walking along a path at free cadence in three different views (directions) with respect to the image plane. This dataset includes 20 persons, performing walking actions captured at a rate of 25 frames per second. The proposed algorithm identifies person by considering only the videos, where the person is allowed to walk perpendicular to the stationary camera that is lateral view. Out of four sequences per person, three sequences are taken as training and the remaining sequence is considered as test. In total 60 sequences are considered as training, whereas the remaining 20 sequences are taken as test. This partitioning is repeated for 4 different times, so that each video of a person is tested. (For ex. in the first experiment, Fig 3(a), the first three sequences, per person is taken as training, while the remaining single sequence is considered as test. In the successive experiments, Fig (3(b), 3(c), and 3(d)), shows the other combinations of input sequences taken as test and training. At any instance of time only three sequences per person is taken as training and the single sequence is considered as test.

4.2 Weizmann Action Recognition Dataset

The Weizmann action recognition dataset contains 9 persons performing 10 types of actions. Each action sequence contains of
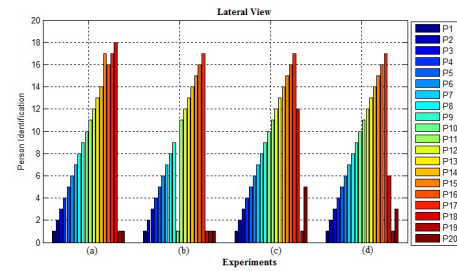

Fig. 3. Person Identification - Lateral View

50-90 frames normalized to a resolution of 80X70. The number of human actions considered for recognition in the proposed algorithm is limited to five and these actions are run, walk, jump, hands-wave and action side. The performance of the proposed human action recognition algorithm is tested using quality threshold clustering. The confusion matrix in recognizing human actions is shown in the Table 1. In quality threshold clustering the maximum diameter of the cluster is chosen first, the candidate cluster is then built from each point, by including the closest point, the next closest and so on. Finally, this process is re-cursed with reduced set of points for generating clusters based on the human actions performed.

Table 1: Confusion Matrix using Weizmann Dataset

	Walk	Run	Jump	Wave	Side
Walk	0.9				
Run		0.9			
Jump			0.9		
Hand-Wave	0.3			0.5	0.1
Side	0.2	0.1	0.1		0.5

5. CONCLUSION

The gait feature extraction and gait sequence analysis becomes an important step in the proposed methodology. The 'Topic Model', developed only for text mining applications, is one of the potential approaches to study gait sequence analysis. The application of topic model like LDA in gait domain framework is achieved by proposing a novel transformation method that transforms the extracted secondary gait sequences into words representation. Finally, document-topic-word distribution is considered to recognize individuals. The performance of the proposed LDA algorithm has been illustrated using CASIA Dataset A and Weizmann action recognition dataset, which resulted in correct classification of 85% using dataset A and 85% using Weizmann action recognition dataset.

6. REFERENCES

[1] Dong Ming, Cong Zhang, Yanru Bai, Baikun Wan, Yong Hu and KDK Luk, "Gait Recognition Based on Multiple Views Fusion of Wavelet Descriptor and Human Skeleton Model", *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurements Systems*, pp. 246-249, 2009.

[2] Jinyan Chen, Jiansheng Liu, "Average Gait Differential Image Based Human Recognition", *The Scientific World Journal*, vol. 2014, pp. 1-8, 2014.

[3] Jinyan Chen, "Gait Correlation Analysis Based Human Identification", *Scientific World Journal*, vol. 16, pp. 275-283, 2014.

[4] Muhammad Rasyid Aqmar, Koichi Shinoda and Sadaoki Furui, "Efficient Model Training for HMM-based Person Identification by Gait", *IEEE Signal and Information Processing*, pp. 1-4, 2012.

[5] Chiraz BenAbdelkader, Ross Cutler and Larry Davis, "View-Invariant Estimation of Height and Stride for Gait Recognition", *Biometric Authentication, European Conference on Computer Vision: International. Workshop*, vol. 4, pp. 155-167, 2002.

[6] I. Laptev and P. Prez, "Retrieving actions in Movies", *IEEE International Conference on Conference Vision*, 2007.

[7] H. Wang, J.C. Niebles and L.Fei-Fei, "Unsupervised learning of human action categories using spatial temporal words", *International Journals on Computers*, vol. 79(3), pp. 299-318, 2008.

[8] Hu Ng, Hau-Lee Tong, Wooi-Haw Tan and Timothy Tzen-Vun Yap, "Human Identification Based on Extracted Gait Features", *International Journal on New Computer Architectures and Their Applications*, vol. 1, pp. 358-370, 2011.

[9] Emdad Hossain and Girija Chetty, "Person Identification in Surveillance Video Using Gait Biometric cues", *International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 1877-1881, 2012.

[10] Daigo Muramatsu, Akira Shiraishi, Yasushi Makihara, Md. Zasim Uddin and Yasushi Yagi, "Gait-Based Person Recognition using Arbitrary View Transformation Model", *IEEE Transactions on Image Processing*, vol. 2, pp. 140-154, 2015.

[11] Chunsheng HUA, Yasushi Makihara and Yasushi Yagi, "Pedestrian Detection by using Spatio Temporal Histogram of Oriented Gradients", *IEICE Transaction. Information & System*, vol. 96, pp. 1376-1386, 2013.

[12] Jiwen Lu and Erhu Zhang, "Gait Recognition for Human Identification based on ICA and Fuzzy SVM Through Multiple Views Fusion", *IEEE Pattern Recognition Letters*, vol. 28, pp. 2401-2411, 2007.

[13] Liang Wang, Tieniu Tan, Huazhong Ning and Weiming Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1505-1518, 2003.

[14] Murat Ekinci, " A New Attempt to Silhouette-Based Gait Recognition for Human Identification", *Computer Science Journals*, vol. 4013, pp. 443-454, 2006.

[15] Y. Dedeolu, "Human action recognition using Gaussian mixture model based background segmentation", vol. 1(1), pp. 1-9, 2009.

[16] G. Zhao,V. Kellokumpu and M. Pietikinen, "Human activity recognition using a dynamic texture based method", *IEEE Transactions on Pattern Analysis and Machine Intelligence Journal*, vol. 32(9), pp. 1705-1720, 2010.

[17] T. Guha and Rabak. K Ward, "Learning sparse representations for human action recognition", *IEEE Transaction, Pattern Analysis Machine Intelligence*, vol. 34(8), pp. 1576-1588, 2012.

[18] Andrew Y. Nag, D.M. Blei and Michael. I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, pp. 993-1022, 2003.

[19] Changhong Chen, Jimin Liang and Xiuchang Zhu, "Gait Recognition Based on Improved Dynamic Bayesian Networks", *IEEE Pattern Recognition Letters*, vol. 44, pp. 988-995, 2011.

[20] www. wisdom. weizmann.ac.il/ vision/ SpaceTimeActions.html.