ABSTRACT

Gait sequence analysis from the input binary silhouettes, has various applications, such as person identification, human action recognition, event recognition and classification. The gait feature extraction is a key step in gait analysis. The 'Topic Model', used for text classification, is one of the potential semantic approaches to study gait sequence analysis. The proposed algorithm uses Latent Dirichlet Allocation (LDA), a 'Topic Model', to analyse the gait sequence for person identification. This has been achieved by proposing a novel transformation method that transforms the gait sequence into word representation suitable for topic models like LDA. The latent dirichlet allocation algorithm, then calculates the word-topic and topic-image distributions, using the words generated by transforming the gait sequences using a transformation method. Finally, the image-topic-word distributions are used to identify person. The performance of the proposed latent dirichlet allocation algorithm, has been illustrated using CASIA dataset A, dataset B, and TUM-IITKGP gait dataset, resulting in an average classification rate of 82.2% using dataset B, and 85%, 85%, and 85% for lateral, oblique and frontal view respectively with respect to dataset A and 90% using TUM-IITKGP gait dataset.

Keywords

Gait Analysis, Gait Sequence, Latent Dirichlet Allocation, Person Identification, Primary Gait Sequence.

1. INTRODUCTION

The gait sequence analysis has applications in multiple domains such as, surveillance, person identification, human action recognition, event detection and classification etc. The extraction of gait features is the key concept in gait sequence analysis. The 'Topic Model', used for text classification, is one of the potential semantic approaches to study gait sequence analysis. The proposed algorithm uses Latent Dirichlet Allocation (LDA), a 'Topic Model', to analyse the gait sequences for person identification. There are several gait recognition algorithms, that are used to recognize individuals by analysing the extracted gait sequences, to name few, the silhouette boundary centroid distance \( d \) is used for person identification. The boundary centroid distance along with the other gait features, are analysed to recognize individuals. The gait features, such as step length, stride length, cadence, cycle length, height, weight and gender are used in \([2,3,4,5]\) for person identification. The spatio-temporal silhouette and frequency-domain features \([6]\) are used for recognizing individuals. Gait biometric cues \([7]\) and human body movements \([8]\) are also used for person identification. The linear time normalization (LTN) \([9]\) is used for recognizing individuals. The LTN transforms the human silhouettes into low-dimensional feature vectors. These vectors are then analysed for recognizing individuals.

The potential semantic approaches, to study text analysis, uses Topic Models in text domain framework. Here, an attempt is made to use Latent Dirichlet Allocation (LDA), a 'Topic Model' in gait domain framework for recognizing individuals and recognizing human actions. The major bottleneck in using topic models in gait domain framework is the lack of natural text concepts such as word, sentences in image or video domain. In this paper, a framework is proposed, that transforms the gait sequences into word representation that are suitable for topic model like LDA. The performance of the proposed LDA algorithm is tested using CASIA gait dataset A, dataset B, dataset C and TUM-IITKGP gait dataset. The Fig 1 shows the block diagram of the proposed algorithm, where the gait sequences are extracted from the moving silhouettes of binary images. These sequences are then analysed to generate words. Finally, LDA uses these words to recognize individuals.

The rest of the paper is structured as follows: Section 2 gives the overview of methods used in person identification by analysing the extracted gait sequences, Section 3 introduces LDA algorithm, and specifies the steps involved in the proposed method for person identification and Section 4 discusses the results, followed by the conclusion and references.

2. RELATED WORKS

In this section, the current state of the art techniques applied in the field of person identification is provided. This will give the background information of all the techniques used in the proposed algorithm. The person identification \([21]\) uses arbitrary view transformation model, to match a pair of gait traits selected from an arbitrary view for identification. The human body is divided into multiple segments \([23]\) during an action, the movements extracted from these segments, which are then analysed in frequency-domain
framework for person identification. In a similar approach, the gait features [10] are extracted by dividing the human body into multiple areas. The weights are assigned for the areas based on the similarities found between the extracted features and those stored in the database. Finally, person identification is achieved as a weighted intergeneration of similarities found in all the areas. The human body shape and their motion are extracted as gait features in [25] for person identification. Person identification in [25] is based on score level fusion of face, gait and height biometric features. These features are extracted from single walking image sequence. The weights are assigned for each feature, which is then integrated for recognition. The dynamic features, derived using inverse dynamics on human body [26] is used for recognizing individuals. The multiple gait features in [11][12][17] are used for recognizing individuals. The multiple gait features in [11][12][17] are used for person identification, where the extracted multiple gait features are smoothed and normalized using principal component analysis and finally, a supervised pattern classification technique is employed for identification. In another similar method [22] the multiple gait features extracted using radon transformation is used to find the gait energy images, which is then used to recognize individuals.

3. LATENT DIRICHLET ALLOCATION

Latent Dirichlet Allocation (LDA) aims to identify the thematic information in large archives of images. The key concepts in LDA are topics and gait-word distribution. LDA is a statistical model, which applies statistical distribution on the proportion of different topics in an image. The statistical distribution is applied on the topics to estimate the parameters using the training database of images. LDA has wide application, due its applicability to various non-text domains in the context with probabilistic models of machine learning. The LDA model consists of several parameters such as:

1. $\beta_{h,w}$ represents the probability distribution of gait-word ($w$) related to $h^{th}$ topic, where $\beta$ is the gait-word probability matrix.
2. $\theta_{d,p}$ is the proportionality of $p$ topic in image $d$.
3. $Z_{d,n}$ is the topic assignment for $n^{th}$ gait-word in image $d$.
4. $W_{d,q}$ is $q^{th}$ observed gait-word in image $d$.

The hidden variables are topics and observed variables are gait-word.

\[
p(\beta_{1:k}, \theta_{1:D}, Z_{1:D}, W_{1:D}) = \prod_{i=1}^{k} \phi(\beta_{i}) \prod_{d=1}^{D} \prod_{n=1}^{N} p(Z_{d,n}|\theta_{d}) p(W_{d,q}|\beta_{1:k}, Z_{d,n})
\]

$p(Z_{d,n}|\theta_{d})$ represents the probability of $\theta_{d}$ on $Z_{d,n}$, where $\theta_{d}$ represents image-topic distribution, $Z_{d,n}$ is topic assignment and $p(W_{d,q}|\beta_{1:k}, Z_{d,n})$ represents the image-topic distribution on observed gait-word $W_{d,q}$, where $\beta_{1:k}$ represents the gait-word distribution per topic, $Z_{d,n}$ is the topic assignment for the gait-word in an image.

4. GAIT SEQUENCE ANALYSIS FOR PERSON IDENTIFICATION - PROPOSED METHOD

The application of LDA in gait analysis is facilitated, only if the concept of image and words are mapped appropriately. This is achieved by proposing a novel framework for defining the words. The steps for generating words involves i) Extraction of primary gait sequences from the binary silhouettes ii) Deriving sub-primary gait sequences from the primary gait sequences and iii) Generating words based on sub-primary gait sequences. Initially, the input video streams are converted into gray scale image frames, the background motion is then subtracted and the binary images $B_{l}$ are constructed using the threshold value as shown in the Equation (1), (2) and (3) respectively.

\[
D_{k}(i, j) = |I_{k}(i, j) - I_{k+1}(i, j)|
\]

\[
\mu = \frac{1}{M + N} \sum_{i=1}^{M} \sum_{j=1}^{N} D_{k}(i, j)
\]

\[
\delta = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (D_{k}(i, j) - \mu)^2}{N(M - 1)}}
\]

\[
T = \mu + 2 * \delta
\]

$I_{k}$ is the $k^{th}$ image frame and $I_{k+1}$ is the $k + 1^{th}$ image frame, $D_{k}$ is the difference image, found using the consecutive images. The variables $i$ and $j$ are used to denote the coordinates (pixel coordinates in case of image and row-column index in case of matrix) of the given image, $(M)$ and $(N)$ determines the size of an image frame. The threshold $(T)$ is found using $(\mu)$ and $(\delta)$ using Equation (5).

\[
B_{l}(i, j) = \begin{cases} 
255 & \text{if } D_{k}(i, j) \geq T \\
0 & \text{Otherwise}
\end{cases}
\]

4.1 Extraction of Primary Gait Sequences

The extraction of primary gait sequences ($\Delta$) is an initial step of the proposed algorithm. This is achieved by enclosing the pixels, that corresponds to a person within the bounding box and tracked over the frames of binary images ($B_{l}$) to extract the primary gait sequences like Horizontal - Movement ($\Delta_{HM}$), Vertical-Movement ($\Delta_{VM}$), Pixel-Count (PC), Left-Area-Counter ($\Delta_{LAC}$) and Right-Area-Counter ($\Delta_{RAC}$). The gait sequences horizontal-movement ($\Delta_{HM}$) and vertical-movement ($\Delta_{VM}$) represents the measure of the distance moved by the person during an human action. The ($\Delta_{HM}$) is measured horizontally and
shows the related equations in generating up-down sequence. The Equations (7)
ues in a sequence (\(\Delta\)) is then subtracted from each element of the same sequence. The
quence, the mean value is found from a sequence (\(\mu\)) is used by the proposed algorithm to calculate the
words proportionality. This is the product of two normalized factors. Each element of the matrix (\(\beta\)) is used to calculate the
portrayal of words generated in its vocabulary are small, as this would influence the topic space and its relation to words
gen the performance of proposed LDA, the sub-primary gait sequences are extracted from the primary gait seq-
sequences, which increase the vocabulary space for LDA related to gaits.

4.2 Extraction of Sub-Primary Gait Sequences
It is noted that the performance of proposed LDA algorithm, reduces with the extracted primary gait sequences, hence the
primary gait sequences gains importance in recognizing individuals. The sub-primary gait sequences extracted from the primary gait sequences increases the words for LDA related to gait sequences. The proposed methodology, extracts the sub-primary gait sequences (\(\Upsilon\)) from the primary gait sequences (\(\Delta\)) using two sequencing techniques i) Above-Below Mean sequence and ii) Up-Down sequence.

4.2.1 Above-Below Mean Sequence. In above-below mean sequence, the mean value is found from a sequence (\(j, \Delta\)). This value is then subtracted from each element of the same sequence. The sign of the element is also considered in calculating above-below mean sequence as shown in Equation (6).

\[
AMBMeq = \text{sign} \left| \Delta(i(E_j)) - \mu_{\Delta} \right| \geq 0
\]  

4.2.2 Up-Down Sequence. Similarly, the up-down sequencing technique is used to generate the sub-primary gait sequences from the primary gait sequences. Here the number of up and down values in a sequence (\(\Delta\)) are counted by subtracting each element with the successive elements of the same sequence. The Equations (7) shows the related equations in generating up-down sequence.

\[
RF_{\text{seq}} = |\Delta(i(E_j)) - \Delta(i(E_{j+1}))| \geq 0
\]  

4.3 Generation of Words
The words are generated by transforming the extracted sub-primary gait sequences (\(\Upsilon\)) into word representation. The steps for generating words, are given below, these steps are executed iteratively within a sequence to generate words. This is repeated with each sub-primary gait sequence (\(\Upsilon\)) to generate words.

In each iteration of generating words:

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 by one, 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.

4.4 Proportionality of Words
The word probability matrix or word-topic distribution matrix (\(\beta\)), is used by the proposed algorithm to calculate the words proportionality. Each element of the matrix (\(\beta\)) is used to calculate the proportionality of words. This is the product of two normalized factors as shown in the Equation (8).

\[
\mathcal{U}(i,j) = \frac{\nabla(i) \cdot \beta(i,j)}{\sum_{k=1}^{M} \nabla(k) \cdot \sum_{i=1}^{N} \lambda(i)}
\]  

The (\(\mathcal{U}\)) refers to proportionality of words, which is the product of two normalized factors.

4.5 Person Identification
In the proposed approach, the input documents, are broadly divided into two groups, namely test and training. The proposed algorithm is executed independently on both these groups. The i) Matrix (\(\beta\)) and ii) Matrix (\(\alpha\)) are generated from the training set of documents, whereas the words are found from the test set. In order to identify the person, the proposed algorithm selects, some of the words from the test, based on the standard deviation value (\(\delta_{\text{test}}\)) given in Equation (9), where \(j = 1..n\) in Equation (9) represents the words and \(w_{(j)}\) is the \(j^{th}\) word occurrence in an input test document. The standard deviation (\(\delta_{\text{test}}\)) given in Equation (10) is found from the word occurrence \((w_{(j)}, w_{(j+1)}, w_{(j+2)}, \ldots, w_{(n)})\) of the test document.

\[
v_{\text{test}(j)} = \left| \sum_{j=1}^{n} (w(j) - \delta_{\text{test}}) \geq 0 \right|
\]

\[
\delta_{\text{test}} = \left| \sum_{j=1}^{n} (\delta(w_{(j)}, w_{(j+1)}, w_{(j+2)}, \ldots, w_{(n)})) \right|
\]

Then for each word in (\(v_{\text{test}(j)}\) Equation (9), the corresponding topic is found, using the word-topic distribution matrix (\(\beta\)), of the training set. The topic selected, is the maximum value with respect
The correct classification rate (CCR) of the documents shown in the Figures 3, 4 and 5 respectively. The CASIA dataset A contains 4 subsets: i) Dataset A (standard gait dataset), ii) Dataset B (multi-view gait dataset), iii) Dataset C (infra-red gait dataset) and iv) Dataset D (gait and its corresponding footprint dataset). The CASIA gait dataset A contains people walking along a path at free cadence in three different views (directions) with respect to the image plane. Namely, lateral (0°), obliquely (45°) and frontal (90°) views. This dataset includes 20 persons, four sequence for each viewing angle per person. Therefore the database includes in total 240 video sequences of 24-bit full color, captured at a rate of 25 frames per second with the resolution 352X240. The proposed algorithm is tested by considering each view separately. Three sequences per person is taken as training, and the remaining single sequence is considered as test. In total 60 sequences are considered as training, whereas remaining 20 sequences are taken as test. This partitioning is repeated for 4 different times, so that at each instance of time only three sequences per person is taken as training, and a remaining single sequence per person is considered as test. The correct classification rate of the proposed algorithm, along the different views (lateral, obliquely and frontal views) are shown in the Figures 3, 4 and 5 respectively.

The correct classification rate (CCR) of the documents (p(1), p(2), p(3), p(4), p(5), p(6), p(7), p(8) and p(9)) in Fig 3(a), 3(b), 3(c) and 3(d) with respect to lateral view is 100%. The CCR drops to 75% for the document (p(10)), that is, in the Fig 3(b) the document (p(10)), is wrongly identified as document (p(1)). Similarly, for the document p(15) the CCR is 75%, that is, in the

<table>
<thead>
<tr>
<th>Method</th>
<th>0°</th>
<th>45°</th>
<th>90°</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC-NN [12]</td>
<td>65.0%</td>
<td>63.73%</td>
<td>77.5%</td>
</tr>
<tr>
<td>NED-NN [12]</td>
<td>65.0%</td>
<td>66.25%</td>
<td>85%</td>
</tr>
<tr>
<td>NED-ENN [12]</td>
<td>75.0%</td>
<td>81.25%</td>
<td>93.75%</td>
</tr>
<tr>
<td>FD+NN-SVM+GFSVM [11]</td>
<td>90.0%</td>
<td>85.5%</td>
<td>82.5%</td>
</tr>
<tr>
<td>WD+NN-SVM+GFSVM [11]</td>
<td>90.0%</td>
<td>87.5%</td>
<td>83.3%</td>
</tr>
<tr>
<td>PSM+NN-SVM+GFSVM [11]</td>
<td>89.5%</td>
<td>88.5%</td>
<td>84.2%</td>
</tr>
<tr>
<td>PCA-NED [19]</td>
<td>90.0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Proposed LDA (First Scenario)</td>
<td>85.0%</td>
<td>85.0%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Proposed LDA (Second Scenario)</td>
<td>80.0%</td>
<td>85.0%</td>
<td>85.0%</td>
</tr>
<tr>
<td>Proposed LDA (Third Scenario)</td>
<td>85.0%</td>
<td>85.0%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Proposed LDA (Fourth Scenario)</td>
<td>75.0%</td>
<td>85.0%</td>
<td>70.0%</td>
</tr>
</tbody>
</table>

Table 1. Reported classification rate (CCR) in identifying person - CASIA Dataset A

The correct classification rate in recognizing individuals using CASIA dataset A.

Fig 3(a), the document (p(15)) is wrongly identified as document (p(17)). The CCR for the document (p(18)) is 25% that is, in the Fig 3(b), 3(c) and 3(d), the document (p(18)), is wrongly identified as documents (p(1), p(12) and p(6)) respectively. Similarly for documents (p(19) and p(20)) the CCR drops heavily. The same assumption is considered to derive the correct classification rate form the oblique and frontal views using Fig 4 and 5 respectively.

The CASIA dataset B, is an indoor multi-view gait dataset of 124 persons, performing 10 sequence of walking actions captured from 11 different views. This dataset includes 2 sequences of walk with carrying bag, 2 sequences of walk with wearing coat and 6 sequences of normal walk. Several experiments were carried out on this dataset, by selecting 4 sequences of normal walk (nm = 01, nm = 02, nm = 03, nm = 04) as training and 2 sequences (nm = 05, nm = 06) as test. The Table 2 shows the reported correct classification rate in recognizing individuals using wavelet (w) and frieze (f) features.
Table 3. Reported Results in Identifying Individuals using TUM-IITKGP Gait Dataset - Colour Histogram (BL 1), Cropped Gait Energy Image (BL 2a), and Cropped Gait Energy Image (BL 2b).

<table>
<thead>
<tr>
<th>Configuration</th>
<th>BL 1</th>
<th>BL 2a</th>
<th>BL 2b</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular (Conf.1)</td>
<td>97.9%</td>
<td>68.6%</td>
<td>77.1%</td>
<td>90%</td>
</tr>
<tr>
<td>Hand-in-pocket (Conf.2)</td>
<td>93.3%</td>
<td>67.1%</td>
<td>75.7%</td>
<td>85%</td>
</tr>
<tr>
<td>Backpack (Conf.3)</td>
<td>75.0%</td>
<td>11.4%</td>
<td>77.1%</td>
<td>80%</td>
</tr>
<tr>
<td>Gown (Conf.4)</td>
<td>20.0%</td>
<td>8.6%</td>
<td>32.9%</td>
<td>40%</td>
</tr>
</tbody>
</table>

5.3 TUM-IITKGP Gait Database

The TUM-IITKGP gait database offers different kinds of gait variations namely, regular, hand-in-pocket, backpack, gown, dynamic occlusion and static occlusion. This database consists of 840 sequences of 35 individuals. The camera is set-up in a narrow hallway, reflecting a realistic set-up in a potential real world surveillance application. The camera is positioned at a medium height of 1.85 meters and is oriented perpendicular to the hallway direction. The people are allowed to walk from right to left and from left to right. Each person is captured in six different configurations like regular (Conf.1), hand-in-pocket (Conf.2), backpack (Conf.3), gown (Conf.4), dynamic occlusion (Conf.5), static occlusion (Conf.6). Furthermore, each configuration is repeated for two times in a right-to-left motion and two times in a left-to-right motion, resulting in a total of 840 sequences. Each person was primarily recorded in a regular walking configuration, followed by three degenerated configurations including hands in pocket, backpack and gown. These configurations can be used to evaluate the recognition methods in the presence of different kinds of gait variation. The proposed algorithm is tested using Conf.1, Conf.2, Conf.3 and Conf.4. The other two configurations Conf.5 and Conf.6 are not considered for identification by the proposed algorithm. The first (initial) walk sequence either left-right or form right-left (140 sequences) of 35 individuals are taken as test, whereas the remaining 700 sequences of 35 individuals are considered as training set. The Table 3 shows the reported results in recognizing individuals using TUM-IITKGP dataset.

6. CONCLUSION

Gait sequence analysis, has several applications in the field of surveillance, person identification, human action recognition, event recognition and classification. The extraction of gait features is an important step in gait sequence analysis. The topic model, developed for text mining applications, is one of the potential approaches to study gait sequence analysis. In the proposed approach, an attempt is made to use latent dirichlet allocation (LDA) a ‘Topic Model’ in gait domain framework. This is achieved by proposing a novel transformation method that transforms the extracted sub-primary gait sequences into words representation. The transformation involves i) Extracting sub-primary gait sequences from the primary gait sequences and ii) Generating words from the extracted sub-primary gait sequences. The words are then used by LDA to calculate word-topic and topic-image distributions. Finally, image-topic-word distributions are considered to recognize individuals. The performance of the proposed LDA algorithm has been illustrated using CASIA dataset A, dataset B and TUM-IITKGP gait database, resulting in correct classification rate of 85%, 85% and 85% for lateral, oblique and frontal view using CASIA dataset A, and an average classification rate 82.2% using CASIA dataset B and 90% using TUM-IITKGP gait dataset.

7. REFERENCES


5


