Visual Biometric based Animal Identification using Topic Models

Deepak N.A. 1, Shobha N.S. 2

1Associate Professor, Department of Computer Science and Engineering, R V Institute of Technology and Management, Bengaluru, Karnataka, India; Email: deepakna.rvitm@rvei.edu.in.

2Assistant Professor, Department of IEM, R V College of Engineering, Bengaluru, Karnataka, India; Email: shobhakoushik@gmail.com

Abstract: Monitoring animal behaviors, activities and natural their habitat is necessary for conservation and management of animal life. The surveillance camera trapped videos are helpful for automatic animal life management and it acts as an important source that provides images for monitoring the animal behavior. The constant monitoring of animal activities, may result in generation of large volume of data processing. However, processing such a huge amount of data manually, is a complex and time-consuming task. Hence their an immense need to automate this process, and to speed-up the recognition task. In this paper, a visual biometric based automatic identification and recognition of individual animals is proposed. The different motion patterns of individual animals found in KTH dataset are used by the proposed system. The proposed algorithm uses motion parameters of individual animals to generate words, which are then fed to” Topic-Modelling” algorithm for identification. The topic modelling algorithm developed for clustering textual data, has some hidden qualities, that helps in study of motion patterns of animals in video or image domain.

Keywords: Monitoring; Visual Biometric; Motion Features; Clustering; Topic Models

1. INTRODUCTION

Monitoring the animals and identifying their behavior is essential in maintaining a balanced ecosystem. The explosion in growth of human population is the major cause of reduction of natural resources. The de-forestation makes enormous change in wildlife and results in environment changes of earth [3]. This result in change of animal behavior and their habitat. The radio tracking system, Sensor based tracking system [8], GPS based tracking [4] are evolved to track the animal and its behavior

But monitoring the wild animals through captured video streams gains popularity due to advancement in technology, like commercial availability of high-resolution cameras results in ease of development, deployment and maintenance. Night vision cameras, 360° rotational cameras help in monitoring the animal behavior’s round the clock in a cost-effective manner. The camera trapping based monitoring round the clock generates a huge amount of data, where manual processing becomes an almost impossible and a hectic task. A computer-based technique, integrated with machine learning outperforms the automatic classification of animals and their behavior’s without human intervention. The paper is organized as follows: Section 2 gives the literature related to proposed work. Section 3 presents the implementation part of the proposed work and Section 4 shows the results of experiments carried out. Finally, the paper concludes with a brief conclusion and references.

2. RELATED WORKS

An automatic wildlife monitoring system for detection and recognition of individual animal species described in [9] The faster region CNN is used to identify the animal body and the flank region is used to recognize individual animals. The features and flanks extracted are used to train the logistic regression for identification. Deep learning-based wildlife surveillance system recognizes the animal species from the captured images [3]. Initially, the method subtracts the image
background, extracts the necessary features and finally, recognizes the animals using standard classifiers. The state-of-the-art deep learning neural network-based method [5], automatically identifies animals from the camera trapped images. Their method uses various deep learning architectures to classify animals. A sparse coding spatial pyramid matching (ScSPM) technique is used along with cell-structured LBP (cLBP) [1] improves the animal classification accuracy. The Linear support vector machine algorithm is also used to achieve improved classification rate.

3. IMPLEMENTATION
The applicability of Hierarchical Latent Dirichlet Allocation (HLDA) a “Topic Model “on animal recognition is made easier, only if the input images and words-sentences are mapped appropriately. This is achieved by proposing a novel technique, that interprets the words from extracted features. The generation of words involves the following steps:

- Tracking the subject of interest over the frames of binary sequences to extract of basic features.
- Generation of words by analyzing the extracted basic features.

3.1 Video-Segmentation
Initially, the input video streams are segmented into gray scale image frames. The image background is subtracted and the binary image (BI) is constructed using the standard mean-delta-threshold method Eq. (1).

\[ BI(m, n) = \begin{cases} 255 & \text{if } DI(m, n) \geq T \\ 0 & \text{Otherwise} \end{cases} \]  

where \( T \) is the threshold found using equation

\[ T_j = \mu_j - 2 \ast \delta_j. \]

The symbol \((\mu)\) represents the mean and \((\delta)\) is the standard deviation found from the pixels of the difference image (\(D_i\)). The \(D_i\), which is found using the equation

\[ DI(m, n) = |I_j[m, n] - I_{j+1}[m, n]|, I_j \]  

is the current image frame and \(I_{j+1}\) is the successive image frame.

3.2 Extraction – Basic Sequences
The primary goal of sequence extraction, is to project the pixel /data onto a smaller dimensional space. This process mainly relies on linear transformation or any other transformations. This is an important step and is required because the extracted sequence dimensionality might be high; this needs to lower down, which reduces the computational complexity. The dimensionality reduction is achieved by selecting some of the required features. The extraction of basic sequence (\(U\)) is the initial step of the proposed algorithm. The feature-based parameter extraction technique is used to extract the basic sequences. This method selects a part of the region or a section in the image space. The selected region is tracked over the successive frames of binary images to extract basic sequences.

The basic sequences extracted are:

- Leg-Maneuver (\(U_{LM}\)): Measures the distance moved by the animal leg, in the direction perpendicular to the stationary camera.
- Knee-Maneuver (\(U_{KM}\)): Measured in degrees, which represents the angle between thigh and the lower portion of the leg (Calf) during an action.
- Head-Maneuver (\(U_{HM}\)): Represents movements occurred in the head of human body during an action.
- Progression-Line(\(U_{LM}\)): Measures the line-of-motion, in which the object or a person progress during an action.

3.3 Generation of Words
The words for Hierarchical latent Dirichlet allocation (HLDA), are generated by transforming the extracted basic sequences (\(U\)) into words representation. The words form the basic building blocks for animal identification in the proposed methodology. In order to generate words, initially four elements \(\delta_1, \delta_2, \delta_3\) and \(\delta_4\) are selected from a sequence (\(U_k\)), where \(k\), is the index of the selected sequence and \(i\) refers to an element of the selected sequence. In the process of generating words, the four elements \(\delta_1, \delta_2, \delta_3\) and \(\delta_4\) are assigned with the first four values found within the selected sequence (\(U_k\)). In the successive iteration, these elements will be assigned with the new values, found by moving each element towards the right by one position, of the selected sequence(\(U_k\)). This process is repeated with each sequence to generate the words. In each iteration, only four elements are considered from a sequence as shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Generation of Words for HLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: Basic Sequences ((U))</td>
</tr>
<tr>
<td>Output: Words ((w))</td>
</tr>
<tr>
<td>for i=1 to Length((U))</td>
</tr>
<tr>
<td>\text{Initialize } \delta_1 = U_i(Ej)</td>
</tr>
<tr>
<td>\text{Initialize } \delta_2 = U_i(Ej + 1);</td>
</tr>
<tr>
<td>\text{Initialize } \delta_3 = U_i(Ej + 2);</td>
</tr>
<tr>
<td>\text{Initialize } \delta_4 = U_i(Ej + 3);</td>
</tr>
<tr>
<td>Step 1: if ((\delta_1 = \delta_2)), (w_{(i+1)} = w_{(i+1)} + 1; )</td>
</tr>
<tr>
<td>Step 2: if ((\delta_3 = \delta_4)), (w_{(2)} = w_{(2)} + 1; )</td>
</tr>
<tr>
<td>Step 3: if ((\delta_1 + \delta_2) = (\delta_3 + \delta_4)), (w_{(3)} = w_{(3)} + 1; )</td>
</tr>
<tr>
<td>Step 4: if ((\delta_1 &lt;= \delta_2)) and ((\delta_3 &lt;= \delta_4)), (w_{(4)} = w_{(4)} + 1; )</td>
</tr>
<tr>
<td>End for</td>
</tr>
</tbody>
</table>

3.4 Standard Techniques
The method used in this study is derived from Gibbs Sampling and Hierarchical Latent Dirichlet Allocation Algorithm.

3.4.1 Probabilistic Inference - Gibbs Sampling: Gibbs sampling is one of the Monte Carlo Markov Chain (MCMC) technique suitable for the task Table 2. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values [6]. The underlying logic of MCMC sampling is that we can estimate any desired expectation by ergodic averages. That is, to compute any statistic of a
posterior distribution as long as we have $N$ simulated samples from that distribution: Eq. (2)

$$E[f(s)]_p \approx \frac{1}{N} \sum_{i=1}^{N} f(s^{(i)})$$

where $(p)$ is the posterior distribution of interest, $f(s)$ is the desired expectation, and $f(s^{(i)})$, is the $i^{th}$ is the simulated sample $(p)$.

Gibbs sampling is commonly used for statistical inference to determine the best value of a parameter. The standard step for Gibbs sampling over a space of variables $a, b, c$.

1. Draw a conditioned-on $b, c$
2. Draw $b$ conditioned on $a, c$
3. Draw $c$ conditioned on $a, b$

The variables for Gibbs sampling
- $W_{m,n}$: the $n^{th}$ word in the $m^{th}$ document (the only observed variable in the model).
- $C_{m,l}$: the document corresponding to the $l^{th}$ topic in document $(m)$.
- $Z_{m,n}$: the assignment of the $n^{th}$ word in the $m^{th}$ document to one of the ($L$) available topics.

**Table 2: Gibbs Sampling Algorithm**

<table>
<thead>
<tr>
<th>Initialize $x^{(0)} \sim q(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>for iteration 1,2,... do</td>
</tr>
<tr>
<td>$x_1^{(i)} \sim p(X_1 = x_1</td>
</tr>
<tr>
<td>$x_2^{(i)} \sim p(X_2 = x_2</td>
</tr>
<tr>
<td>$x_0^{(i)} \sim p(X_D = X_D</td>
</tr>
<tr>
<td>End for</td>
</tr>
</tbody>
</table>

**3.4.2 Probabilistic Inference - Hierarchical Latent Dirichlet Allocation (HLDA):**
Hierarchical Latent Dirichlet Allocation (HLDA) attempts to find latent (hidden) topics within a collection of documents [2]. This is a non-parametric version of Latent Dirichlet Allocation (LDA), meaning that the model identifies the number of topics from the data itself, in which a tree of topics is built Fig.1. The algorithm generates a path through the tree for each document, starting from the root to a leaf node, and allocates each word in the document to one of the nodes in the path. On subsequent passes of the Gibbs Sampler, the document is de-allocated from its current path, a new path is chosen and words are re-allocated to the nodes along the path. The working of hierarchical latent Dirichlet allocation (HLDA), involves in generating 1) A model of multiple-topic documents. 2) A mixture distribution on topics and 3) Picks a topic according to their distribution and generate words according to the word distribution for the topic as shown in Fig.2.

![Fig. 1. Tree of Topics Topic Identification and Assigning Words - (Figure from Blei, et al 2003 [2])](image)

There are some additional parameters that can be used to tune how the model works.

1. **maxDepth**: Sets how many levels exist in the tree. The default value is 3.
2. **gamma**: Governs how likely new paths are generated in the tree. Smaller values mean fewer new paths will be generated. The default value is 1.0.
3. **eta**: Controls how sparse or rich topics become. Smaller values result in sparser topics, meaning more topics will likely be discovered. The eta parameter is set as an array. Each level in the tree may have a different value of eta. Since the default maxDepth is 3, the eta default must have 3 values, which are 2.0, 1.0, 0.5. This default favors more words in the root node.
4. **m**: Controls the proportion of general to specific words. Large values of $m$ favor more words being pushed to the leaf nodes. The default value is 0.5.
5. **pi**: Used for the stick breaking process. The $pi$ value dictates how strongly $m$ should be upheld. The default value is 1.0.

**Simple Representation of HLDA**
1. Let $C_1$ be the root Document (Fig. 2).
2. For each level $l \in \{2,...,L\}$:
   - Draw a topic from the Document $C_{l-1}$.
   - Set $C_l$ to be the Document referred to by that Topic.
3. Draw an $L$-dimensional topic proportion vector $\Theta$ from $\text{Dir}(\alpha)$.
4. For each word $n \in \{1,...,N\}$:
   - Draw $z \in \{1,...,L\}$ from $\text{mult}(\Theta)$. Draw $w_n$ from the topic associated with Document $C_z$.

Finally, the proposed approach uses the standard HLDA software [2] for animal identification. The output of hierarchical latent Dirichlet allocation is then analyzed using a novel approach to identify a final result.
3.5 HLDA based Animal Recognition

In order to identify animal and their behavior, the input documents are divided into test and training samples. The proposed algorithm is then executed on these two samples independently. First, the output of HLDA is found from the documents of the training set. The words and its occurrences within a document are found from the test set. Animal identification involves in filtering out, some of the generated words found from the test set. This filtering process is carried out based on the standard deviation value Eq. (3).

\[
\rho_{\text{test}}(j) = |w(j) - \delta_{\text{test}}| \geq 0
\]  

(3)

For each word the corresponding topic is selected. The selected topic will be the maximum value found in the row. In the next step, for each selected topic, the corresponding document is found, which is the maximum value found in the selected row. Finally, an animal is identified by selecting a single document from the training set as a result of the proposed algorithm, that is, the maximum value found in the selected row.

4. Experiments and Results

In order to test the performance of the proposed algorithm, we divide the dataset broadly into two sets, 1) Training and 2) Testing as shown Table 3. The algorithm is executed on these two sets separately. Initially, the output of HLDA is found from the images of the training set, whereas the words are calculated from the test set. The experiments were conducted on KTH [7] datasets to evaluate the performance of the proposed algorithm. The Fig.3 and Fig. 4. shows the animal species and binary images of KTH dataset.

The performance of the proposed model reduces on colored images of the dataset. In-total 12 different categories of animals are considered for recognition. To investigate the adaptability of the model in the absence of color images, testing is performed on segmented binary image frames on same 12 different categories chosen. The corresponding results and classification accuracy obtained on original color image frames and binary images are presented in the Fig.5. This implies that the model achieves excellent performance on segmented binary images compared colored images.

<table>
<thead>
<tr>
<th>Animal category</th>
<th>Number of Colored Images</th>
<th>Number of Segmented Binary Images</th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear</td>
<td>1586</td>
<td>1586</td>
<td>586</td>
<td>1000</td>
</tr>
<tr>
<td>Coyote</td>
<td>1500</td>
<td>1500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Giraffe</td>
<td>1260</td>
<td>1260</td>
<td>260</td>
<td>245</td>
</tr>
<tr>
<td>Dear</td>
<td>1500</td>
<td>1500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Gorilla</td>
<td>1245</td>
<td>1245</td>
<td>245</td>
<td>245</td>
</tr>
<tr>
<td>Elephant</td>
<td>1500</td>
<td>1500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Leopard</td>
<td>1500</td>
<td>1500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Lion</td>
<td>1470</td>
<td>1470</td>
<td>470</td>
<td>470</td>
</tr>
<tr>
<td>Kangaroo</td>
<td>1350</td>
<td>1350</td>
<td>350</td>
<td>350</td>
</tr>
<tr>
<td>Panda</td>
<td>1455</td>
<td>1455</td>
<td>455</td>
<td>455</td>
</tr>
<tr>
<td>Zebra</td>
<td>1155</td>
<td>1155</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>Tiger</td>
<td>1500</td>
<td>1500</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

Fig. 5. Comparison of accuracy - Color Images and Segmented Binary Image Frames

Conclusion
In this paper we propose an automatic identification of animals using KTH animal dataset. The KTH animal dataset consists of 1134 images of 12 different classes. The proposed algorithm extracts the following basic sequences by the tracking the object of interest from the segmented binary image frames 1) Leg-Maneuver, which is the distance moved by the animal leg during an action, 2) Knee-Maneuver measured in degrees, which represents the angle between thigh and the lower portion of the leg. 3) Head-Maneuver represents movements of head during an action. 4) Progression-Line measures the line-of-motion, in which the animal progress during an action. Words are then generated by analyzing these features, which are then fed to standard package of HLDA to compute the output of HLDA. The first basic contribution is the extraction of basic sequences from the binary images. The second basic contribution is in the generation of words from the extracted basic sequences and the third basic contribution is in recognizing animal from the output of HLDA, with these three new elements, the results become qualitatively superior which is not surprising.

References

Journal


Conference
[3] Hung Nguyen; Sarah J. Maclagan; Tu Dinh Nguyen; Thin Nguyen; Paul Flemons; Kylie Andrews; Euan G. Ritchie; Dinh Phung; Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring, International Conference on Data Science and Advanced Analytics (DSAA), 2017, DOI: 10.1109/DSAA.2017.31


E-Book