

LDA Based Classification of Video Surveillance Sequences Using Motion Information

Abdou Khadre DIOP*, Serban MEZA**, Mihaela GORDAN**, Aurel VLAICU**

*Department of Information and Communications Technology (ICT), Alioune DIOP University Bambey, Senegal

**Research Center for Information Multimedia and Telecommunications Technical University Cluj Napoca, Romania

abdoukhadre.diop@uadb.edu.sn, serban.meza@com.utcluj.ro, mihaela.gordan@com.utcluj.ro, aurel.vlaicu@com.utcluj.ro

Abstract— Video surveillance is one of the key components in today's public security. The possibility to identify abnormal events in such sequences is a difficult problem in computer vision with the aim of providing automatic means of analysis. The use of Latent Dirichlet Allocation (LDA) provided encouraging results for topic classification in text documents and extensions to the video range have already been presented in the literature. The paper approaches video sequence classification considering the extension of the LDA model by building a vocabulary based on motion information “words” that are used to isolate events/topics present in the video. The implementation is tested on the PETS datasets and results are compared with state of the art.

Keywords— Latent Dirichlet Allocation (LDA), video surveillance, motion information, probability distribution.

I. INTRODUCTION

The analysis of the video surveillance in order to identify abnormal events is a difficult problem in computer vision with the declared aim to include such method in automatic/autonomous classification and decision-making ecosystems used for ensuring public safety. Several studies have been carried in the field to optimize the detection of abnormal events in a video surveillance [1].

The Latent Dirichlet Allocation (LDA) model is a generative probabilistic model initially derived to describe a text document collections or other types of discrete data. In [2], LDA posits that words carry strong semantic information, and documents discussing similar topics use a similar grouping of words. Latent topics are thus discovered by identifying groups of words in the corpus that frequently occur together within documents. In this way, LDA models documents as a random mixture over latent topics, with each topic being characterized by its own particular distribution over words. [2] shows that LDA is not only useful in the text domain, but also in the image and music domain. In this paper we show that LDA can present a valid method for video sequence classification as well. To describe our work, the paper is structured in several parts, with the first part briefly introducing the LDA model. The second part describes our contribution to the adaptation for video surveillance sequence of the LDA model's semantic. The third

part presents the implementation of the proposed algorithm and method.

II. LATENT DIRICHLET ALLOCATION (LDA) MODEL

The Latent Dirichlet Allocation is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words [3], [4].

The following terms are defined as basic semantic components [4]:

A word $w \in \{1, \dots, V\}$ is the most basic unit of discrete data. For cleaner notation, w is a V -dimensional unit-based vector. The set of all possible w_i forms the given/chosen vocabulary. If w takes on the i^{th} element in the vocabulary, then $w_i = 1$ and $w_j = 0$ for all $i \neq j$.

A document is a sequence of N words denoted by $w = (w_1, w_2, \dots, w_N)$, where w_N is the N^{th} word in the sequence, and N represents the size of the document in terms of the number of words used.

A corpus is a collection of M documents denoted by $D = \{w_1, w_2, \dots, w_M\}$.

A topic $z \in \{1, \dots, K\}$ represents a given class of documents that have a common/similar probability distribution of words over the vocabulary of V words.

LDA assumes the following generative process for each document indexed by $m \in \{1 \dots M\}$ in a corpus, [2]:

1. Choose a K -dimensional topic weight vector θ_m from the distribution

$$p(\theta|\alpha) = \text{Dirichlet}(\alpha)$$

The Dirichlet distribution lets take a variable θ such that $\forall i, \theta_i \geq 0$ et

$$\sum_{i=1}^k \theta_i = 1 \quad (\theta \text{ in } (k-1) - \text{simplex});$$

Density form is:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1}; \quad (1)$$

$$\alpha \in \mathbb{R}^k, \alpha_i > 0;$$

$\Gamma(x)$ Gamma function

Therefore, the distribution provides a multinomial distribution parameter θ corresponding for the LDA a mixture of topics of document w .

2. For each word indexed by $n \in \{1 \dots N\}$ in a document:
 - a. Choose a topic $Z_n \in \{1 \dots K\}$ from the multinomial distribution

$$p(Z_n = k | \theta_m) = \theta_m^k$$
 - b. Given the chosen topic Z_n , draw a word w_n from the probability distribution

$$p(w_n = i | z_n = j, \beta) = \beta_{ij}.$$

Fig. 1 shows the graphical model representation of LDA [3]. The boxes are “plates” representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and word within a document.

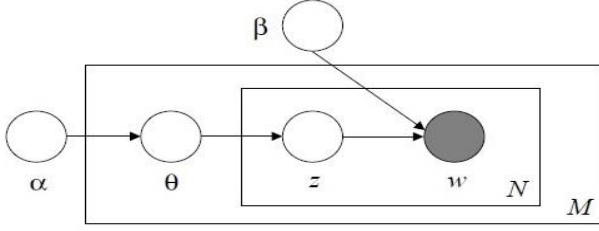


Figure 1. Graphical model representation of LDA.

Analogously, we define the semantic of the LDA model for video sequence.

III. SEMANTIC OF THE LDA MODEL FOR VIDEO SEQUENCE

We first define some terms and notation used commonly in the literature that parallel those defined in the general case, [2], [8].

- A codeword $x \in \{1, \dots, V\}$ is the most basic unit of discrete data in our model.
- A frame is the most basic unit of an input video sequence.
- A codebook is a collection of V learned codewords. This codebook is formed by extracting frames from all video sequences in a corpus, converting them into codewords, clustering these codewords into V clusters, and then selecting the centroid of each cluster as a representative for the V th codeword in the codebook.
- A video sequence is a collection of N frames denoted by $X = (x_1, x_2, \dots, x_N)$, where x_N is the N th codeword in the collection.
- A corpus is a collection of M video sequence denoted by $D = \{X_1, X_2, \dots, X_M\}$, where X_M is the M th video sequence in the corpus.
- A topic $Z \in \{1, \dots, K\}$ is a probability distribution over the vocabulary of V codewords. Topics model particular groups of codewords that frequently occur together in video sequences, and may usually be interpreted as motion information.

To generate a Video Sequence under the LDA model, a topic weight vector is drawn from a Dirichlet distribution. This topic weight vector determines which motion information are most likely to appear in the Video Sequence, [6], [7]. For each frame

in the video sequence, we draw a motion information from the topic weight vector. This motion information given, we generate a codeword by drawing from that motion information distribution. This process can be described more formally as follows.

For each Video Sequence indexed by $m \in \{1, \dots, M\}$ in a corpus:

1. Choose a K -dimensional topic weight vector θ_m from the distribution

$$p(\theta | \alpha) = \text{Dirichlet}(\alpha)$$
2. For each frame indexed by $n \in \{1 \dots N\}$ in the video sequence:
 - a. Choose a motion information $Z_n \in \{1 \dots K\}$

$$p(Z_n = k | \theta_m) = \theta_m^k.$$
 - b. Given the chosen motion information Z_n , draw a codeword x_n from the probability distribution

$$p(x_n = i | z_n = j, \beta) = \beta_{ij}.$$

IV. EXPERIMENTS

A. LDA model training

To inquire about motion information in video transmission, we can apply image differencing in each location (block) to detect changes in a corpus. So we use a motion estimation algorithm which the most popular criteria is based on the displaced frame difference (DFD) [9], [11]. It consists in computing the differences in intensity between every point in every block in the anchor frame and the corresponding point in every block in the target frame. Mathematically, the objective function can be written as:

$$E_{\text{DFD}}(\mathbf{d}) = \sum_{\mathbf{x} \in R} (f_2(\mathbf{x} + \mathbf{d}) - f_1(\mathbf{x}))^p \quad (2)$$

Where $f_1(\mathbf{x})$ is the pixel value at $\mathbf{x}$ in time t_1 , $f_2(\mathbf{x} + \mathbf{d})$ is the pixel value at $(\mathbf{x} + \mathbf{d})$ in time t_2 , $\mathbf{d}$ is the estimated displacement, and R is the region of support.

- When $p = 1$, the error $E_{\text{DFD}}(\mathbf{d})$ is called the mean absolute difference (MAD);
- And when $p = 2$, it is called the mean squared error (MSE).

By calculating the average value of the E_{DFD} in each block in the target frame with $p=1$, we can classify the video sequence by topics represented in Tab. 1. Topics are events observed in the video sequence.

TABLE 1. AVERAGE VALUE OF THE EDDF FOR TOPICS

Topics	Actions	Value of E_{DFD}
Walking	Most moving at low speed	$100 < E_{\text{DFD}} \leq 180$
Local Movement	Little or no movement	$180 < E_{\text{DFD}} \leq 200$
Crowd Formation	Merge of individuals	$200 < E_{\text{DFD}} \leq 220$
Crowd Splitting	Split to 2 or more directions	$220 < E_{\text{DFD}} \leq 240$
Running	Most moving at high speed	$220 < E_{\text{DFD}} \leq 255$

B. LDA model classification

The proposed method is implemented in MATLAB 2015 and has been tested on two datasets: The Crowd_PETS09 [10] dataset has three one video sequence S3 contained set: High_Level. This dataset contains different crowd activities and the task is to provide a probabilistic estimation of each of the following events: Walking, Local Movement, Crowd Formation, Crowd Splitting, Running at different time instances.

The set comes with different timings (such as 14-16, 14-27, 14-31 and 14-33) and account 378 frames numbered from 000 to 377. The timings refer to the hour (“hh”)-minute (“mm”) of data collection. There are four different views (001, 002, 003 and 004) for each dataset time. In This work we consider S3/High_Level (with time 14-33, View001) and S3/High_Level (with time 14-33, View002).

Furthermore, we are interested in systems that can identify the start and end of the events considered as topics.

In the LDA model, S3/High_Level (with time 14-33, View001) and S3/High_Level (with time 14-33, View002) are considered as the corpus. By calculating the number of document, we consider a sliding window of 30 frames, comprising several periods and shifted of 5 frames [12].

The sliding window is illustrated in Fig. 2.

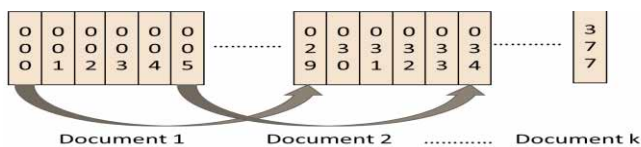


Figure 2. Sliding window of 30 frames, comprising several periods and shifted of 5 frames

For establishing codeword, each frame of each document (video sequences) is divided in 8x8 blocks [5], [6] and [7]. E_{DFD} mean value of each block is considered as codeword. Fig. 3 and Fig. 4 represent the number of Documents (Video Sequence) in terms of number of Codeword (Block) respectively for S3/High_Level (with time 14-33, View001) and S3/High_Level (with time 14-33, View002).

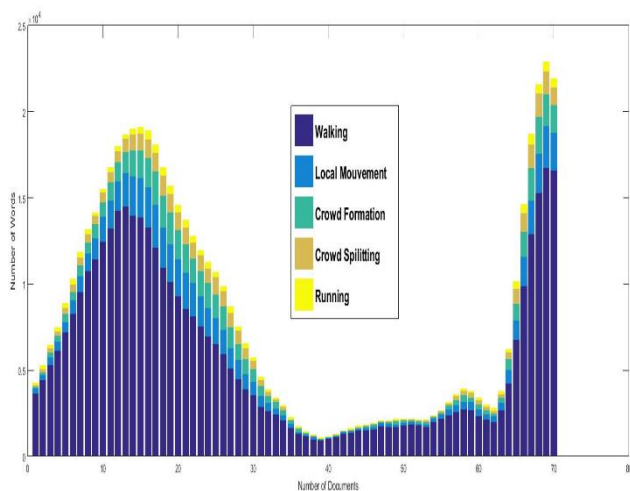


Figure 3. Histogram of number of video sequence in terms of number of blocks in all topics. S3/High_Level (with time 14-33, View001) is considered.

For observing the “Running” action in the video sequence, Fig. 6 shows respectively the first two frames in the detected video sequence in a.) and b.) and Fig. 7 shows respectively the last two frames in the detected video sequence in c.) and d.). We also consider S3/High_Level (with time 14-33, View002).

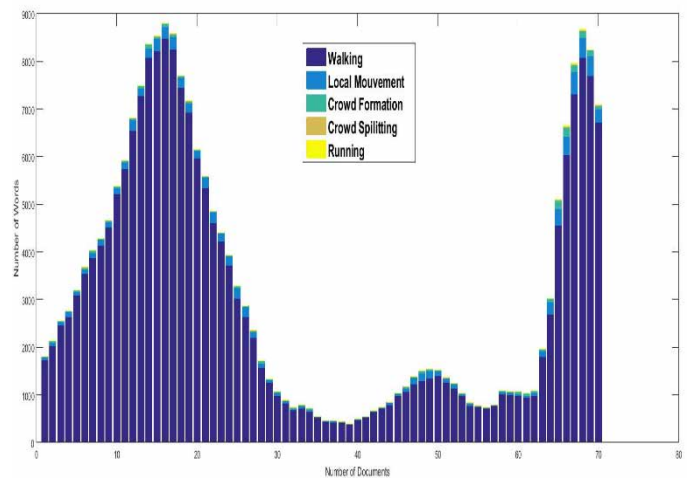


Figure 4. Histogram of number of video sequence in terms of number of blocks in all topic. S3/High_Level (with time 14-33, View002) is considered.

For S3/High_Level (with time 14-33, View001) and S3/High_Level (with time 14-33, View002), our implementation shows that Topics “Running” is most important in the corpus at the 67th Document (Video Sequence) corresponding at the range [335-364] of frames that we can observe people run (Fig. 6 and Fig. 7).

The number of codewords in terms of the average value of E_{DFD} is illustrate in Fig. 5 by considering the topic “Running”.

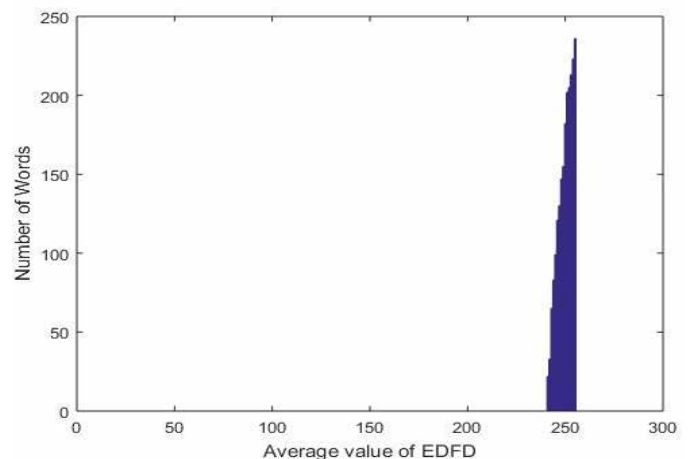


Figure 5. Number of codewords in terms of the average value of E_{DFD} using the topic “Running”.

V. CONCLUSIONS

The Latent Dirichlet Allocation (LDA) is a versatile generative model that can be used to discover underlying from large data collections, be it of texts, images, or even music notes [2]. The proposed approach shown that LDA model is useful in the video domain. Events observed in a video sequence is classified by using LDA model. As anomalous event detection

is one of the important applications in crowd monitoring, our implementation is limited to topics "Running" considered as anomalous event.

In terms of future work, we propose the detection of abnormal events in a surveillance video based on Latent Dirichlet Allocation (LDA) model in the compressed domain.

REFERENCES

- [1] Lorena Calavia, Carlos Baladrón, Javier M. Aguiar, Belén Carro and Antonio Sánchez-Esguevillas: "A Semantic Autonomous Video Surveillance System for Dense Camera Networks in Smart Cities", *Sensors Journal* 2012; doi:10.3390/s120810407, pp. 10407-10429.
- [2] Diane J. HU, "Latent Dirichlet Allocation for Text, Images, and Music", Research Exam, Department of Computer Science, University of California, San Diego, 2009.
URL : <http://seweb.ucsd.edu/~dhu/research.html>. Retrieved on 20/07/2016.
- [3] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research* 3 (2003) pp. 993-1022.
- [4] Nicholas Vretos, Nikos Nikolaidis, Ioannis Pitas, "Video fingerprinting using Latent Dirichlet Allocation and facial images", *ScienceDirect, Journal Pattern Recognition Elsevier* 45 (2012) pp. 2489-2498.
- [5] Razie Kaviani, Parvin Ahmadi, Iman Gholampour, "Incorporating Fully Sparse Topic Models for Abnormality Detection in Traffic Videos", 4th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE (2014) pp. 586-591.
- [6] Razie Kaviani, Parvin Ahmadi, Iman Gholampour, "Automatic Accident Detection using Topic Models", 23rd Iranian Conference on Electrical Engineering (ICEE), IEEE, (2015) pp. 444-449.
- [7] Parvin Ahmadi, Mahmoud Tabandeh, Iman Gholampour, "Abnormal event detection and localization in traffic video based on group sparse topical coding", *ScienceDirect Journal Image Processing Institution of Engineering and Technology*, 2016, Volume 10, Issue 3, pp. 235-246.
- [8] Manika Kar, Sérgio Nunes, Cristina Ribeiro, "Summarization of changes in dynamic text collections using Latent Dirichlet Allocation model", *ScienceDirect Journal Information Processing and Management* 51 (2015) pp. 809-833.
- [9] Oge Marques, "Practical Image and Video Processing Using MATLAB", Chapter 22 Digital video processing techniques and applications, John Wiley & Sons, Inc., (2010) pp.561-589.
- [10] <http://www.cvg.reading.ac.uk/PETS2009/a.html#s2>.
- [11] DIOP, A. K.; Tall, K.; Farssi, S.M.; Diop, I.: "Performances of the estimation and motion compensation for the reconstruction of motion areas in a sequence video Motion JPEG 2000", In *Proceedings of IEEE ICACT 2014 16th International Conference Advanced Communication Technology*. February 2014, pp 92-95.
- [12] Manuel de-Frutos-Lopez, José Luis González-de-Suso, Sergio Sanz-Rodríguez, Carmen Peláez-Moreno, Fernando Díaz-de-María: "Two-level sliding-window VBR control algorithm for video on demand streaming", *Sciencedirect, Signal Processing: Image Communication*, Volume 36, August 2015, Pages 1-13.

First A. Author : Dr. Abdou Khadre DIOP was born the 10th of July 1979 in Pikine, Dakar, Senegal. He received the degree in Telecommunications PhD thesis from the Cheikh Anta DIOP University of Dakar (Senegal) in 2013. From 2012 to 2014, he was employed as ATER at the Department of Technology and Information Communication of the Alioune DIOP University. Since 2014 he was employed as permanent Teacher and Researcher at that department. His current research areas cover transmission of the video sequence in Video Surveillance.

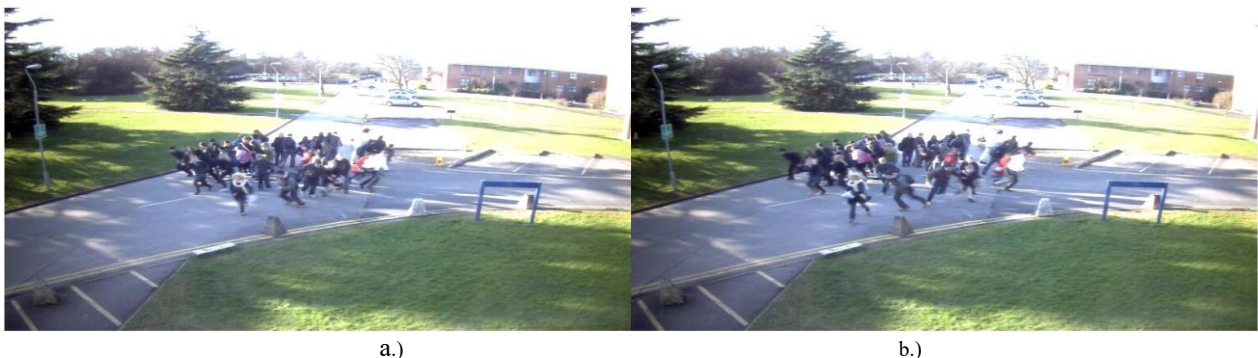


Figure 6. First two frames in the detected video sequence a. and b. respectively



Figure 7. Last two frames in the detected video sequence c. and d. respectively