International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 4, No. 2, 17-34

A Unified System for Processing News and Detecting Advertisements in TV Streams

Abdesalam Amrane[1,2], Abdelkrim Meziane[2], Nour el houda Boulkrinat[2], Ali Atik[2]

[1] Computer Science Department, University of Bejaia, 06000 Bejaia, Algeria
amrane@cerist.dz
[2] Research Center on Scientific and Technical Information, 16030 Algiers, Algeria
ameziane@cerist.dz
nboulkrinat@cerist.dz
atik@cerist.dz

Abstract. Media monitoring plays an important role in the success of companies by protecting their corporate, staff, and brand reputation. With the emergence of ICT, monitoring has emerged as an inevitable means of covering various fields of media: newspapers, online news, broadcast news (TV, radio), and social networks. Companies need a 360-degree view of media sources in near real-time that reports what's going on. In this paper, we proposed a unified system for processing news and detecting advertisements in TV streams. The system architecture is completely described, and the automatic processing has been detailed. Our contributions include a TV stream segmentation method, a news topics segmentation method, and an advertisements classification model. These methods have been evaluated and given promising results. However, more improvements can be made in the future.

Keywords: Deep Learning · TV Stream Analysis · News Processing · Advertisements Extraction · Media Monitoring

1 Introduction

Media monitoring is the process of reading, watching or listening to the editorial content of media sources on a continuing basis [7]. Media monitoring systems are forcing themselves to automate more and more the process of collecting and analyzing news information, while some solutions combine the automatic information gathering from one side and the experts analysis in the field of information and communication sciences in the other. We are mainly interested in relieving experts as much as possible of the automated tasks that are part of the multimedia watch process, namely: monitoring, collecting and disseminating information. The process of media monitoring (news and advertisements) is depicted in Fig. 1.

Through media monitoring, companies aim to find information about competitors and specific issues relevant to their operation. Whereas the old-fashioned press clipping services required 2 to 3 weeks to deliver clips, online media monitoring services deliver clips overnight as a standard service and usually offer near real time delivery [7]. The set of clips are summarized and delivered by e-mail in text or HTML format associated by PDF files generated from clips and sent via FTP. This enables executives in companies to keep up-to-date with a fast and comprehensive overview of their reputation.

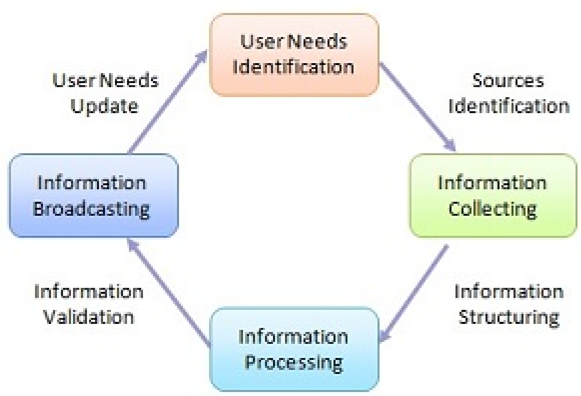

Fig. 1: Information monitoring process

The providers of information operate on various media of communication: newspapers, TV, radio, web, and social networks. The number of media sources is still growing which makes it necessary to have a platform that monitors, processes, and distributes daily press reviews for companies according to their areas of interest. Multimedia data from television channels and generated daily are of the order of gigabytes and push us to exploit the tools of Big Data for storing, processing, and analyzing these data.

The media monitoring system provides companies near real-time reports on all aspects of compliance and competitor activity. Video advertisements (TV commercials) have become an indispensable tool for marketing. Advertisement detection is a classification problem with various training datasets, where every class may have different audio length which vary between 3 and 60 seconds. Some recent research has started to focus on the classification of imbalanced data since real-world data is often skewed [6]. Deep learning-based models are used to overcome some of the limitations of the hand-crafted features [22].

In this work, we present a global architecture of our Multimedia Monitoring System (MMS), detailing some development of specific tools for collecting and processing multimedia news extracted from the television streams of local channels. They include distributed streams acquisition and storage, stream segmentation, content analysis, Arabic text recognition, and selective broadcast of the press review to customers. Monitoring news includes newspaper, online news, broadcast news, and social media. We focus on processing of a ten public and private TV channels where news content is extracted automatically and merged by subject, then analyzed and summarized by specialized users before the step of delivery to the customers.

2 Related Work

In the literature, the problem of extracting and analyzing information from television stream is processed in three stages which are: stream structuring into programs, news program identification, and news topics segmentation. The television stream structuring also called program extraction requires a macro-segmentation of the TV stream. Macro-segmentation algorithms generally rely on detecting inter-programs (IP) which include commercials, trailers, and jingles as in [16]. Pinquier and Andre-Obrecht, detect and locate one or many jingles to structure the audio dataflow in program broadcasts [25]. According to [18], authors proposed an automatic system for TV broadcast structuring. It is based on studying repeated sequences in the TV stream in order to segment it. Segments are then classified using an inductive logic programming-based technique that makes use of the temporal relationships between segments. Metadata are finally used to label and extract programs using simple overlapping-based criteria. Ramires et al. [26] proposed an approach that uses only audio features and centers on the detection of short silences that exist at the boundaries between programs and advertisements.

Wang et al. [30] proposed a multimodal representation of individual programs by using program-oriented informative images, key frames, and textual keywords in a summarization manner for program segmentation in broadcast video streams. Unsupervised TV program structuring is another work proposed in [1], the idea is to automatically recover the original structure of the program by finding the start time of each part composing it. It is based on the detection of separators which are short audio/visual sequences that delimit the different parts of a program.

Several works were proposed for news program identification, among them Zlitni et al. [34] suggested an approach based on video grammar to identify the programs in TV stream and deduce their internal structure. Li et al. [17] proposed a program segmentation system in news broadcast and have used the online electronic program guides (EPG) and closed caption text in TV news. In a recent work, Kannao et al. [14] proposed a two-stage approach to classify news video segments. First, broadcast video shots are classified with multiple labels based on a set of audiovisual features. Second, sequences of these shot features are modeled to detect news programs. Another motivation of program identification is TV commercial detection, as the advertisers spend much money, it is necessary to verify their commercials are broadcasted as contracted. A system for TV commercial monitoring is desired [8].

The topics segmentation for news program has been studied by many researchers such as in [34], [12], [32], [9], [28]. It has processed in two techniques: based on audio transcription methods and based on text detection methods. Text detection methods in the literature can be grouped into texture-based, connected component-based and hybrid methods. Texture-based algorithms scan the image using generally multi-scale sliding windows to extract different texture proprieties and classify image areas as text or non-text based on texture-like features [32].

Many multimedia content analysis systems often require real-time processing and scalability to deliver and store videos. To achieve scalability, programmers need to use distributed programming frameworks such as Hadoop/MapReduce. [15] proposed a high-speed and scalable video server system using PC-cluster for video content delivery. Regarding on development tools, [13] proposed PyCASP, a Python-based content analysis parallelization framework. That is designed using a systematic, pattern-oriented approach with the goal of making it modular, comprehensive and applicable to a wide range of multimedia content analysis applications. Another original approach was proposed by Nedjah et al. [23] that consists to put neural networks in a parallel and distributed setting.

In general, TV broadcast processing can be performed using either the metadata associated with the stream or by directly analyzing the audiovisual stream [5]. Metadata stream like closed caption, electronic program guide, event information table and teletext provides specific textual information that describes the audiovisual streams. Manson et al. [19] proposed a metadata-based system that automatically structures TV stream, the system first detects inter-programs as repeated sequences in the broadcasted stream, and then deduce the boundaries of programs. In audiovisual-based systems, some solutions have been proposed that use speech transcription to extract information like proposed by [12], [24], and [21].

The embedded text in videos is one of the most relevant sources of high-level semantic information [31] that are used in videos indexing and searching. In [33], Zhang proposed a novel unsupervised method to detect and localize the text objects occurring in image and video documents based on a text model, character features and a tracking method to track text event in video documents.

Optical Character Recognition systems (OCR) have been widely developed in past years for different languages and addressing specially scanned documents, also used for TV news video indexing.

Smith et al. [29] described the development of a system for learning of semantic concepts in broadcast video and report on experimental results showing effective automatic detection of semantic concepts in the domain of broadcast news video. Among the works dealing with semantics, [4] proposed a multimedia retrieval system that uses low level information and textual information in indexing process. However semantic concepts alones are not useful for executives in client companies, they needs to be informed by only the most important news that have been analyzed and summarized by experts in communication science. There are another category of works in broadcast news that concerns person identification, such as proposed by Rouvier et al. [27] and who uses scene understanding in order to improve unsupervised people identification.

Escalada et al. [10] proposed an automatic indexing and aggregation web based system (NewsClipping) for the news domain that manages multimedia information without any manual annotation, but their application use metadata associated to the TV stream for indexing content. Some online news service such as CyberAlert[3] monitor the closed caption text of TV stations and deliver a text file of all clips. Y. Aggoun proposed a media monitoring process and reporting for decision-makers [2], [3].

Finally, we conclude that most of the cited methods are based on metadata provided by the broadcasters. In our context, local TV channels don't use metadata and black frames in the broadcasting process. Our contributions are the following:

- A new method for TV stream segmentation,
- A classifier for news program identification
- A new process fo news topics segmentation
- A deep learning model for advertisements classification

Next, we will describe a global architecture of a multimedia monitoring system, and experiment different modules.

3 Proposed System

The media monitoring systems must provide 360-degree view of media sources in real-time and coverage 24/7 support to the clients. Based on the companies needs and according to the context of local media in Algeria, we propose a semi-automatic Multimedia Monitoring System (MMS) where we describe all components to process newspaper, online news, broadcast news, and social networks. It combines the thoroughness of automated news clipping with the accuracy and judgment of human readers. The system must deliver daily alerts to companies, reports, and periodically synthesis of all media. Then, we mainly focus on processing news, and detecting advertistements in TV streams. News have to be

[3] www.cyberalert.com

extracted, stored, summarized, and prepared for delivery. Our proposed system is described in Fig. 2.

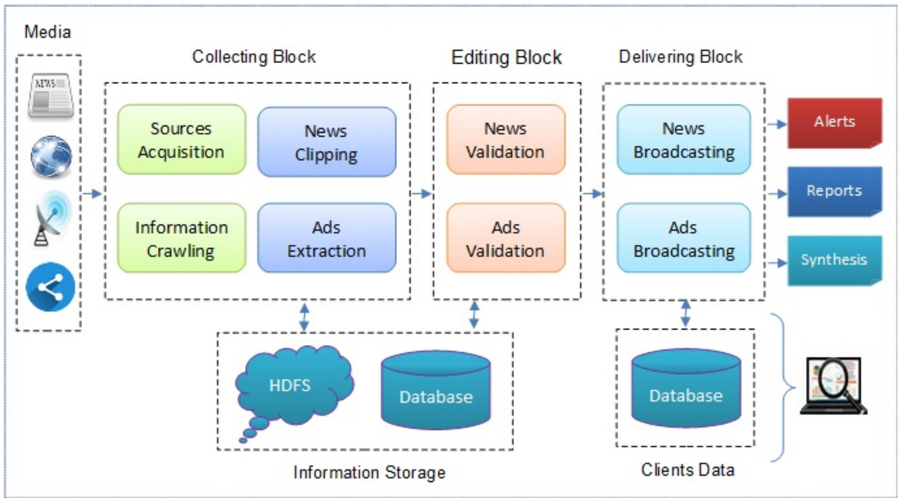

Fig. 2: Overview of the proposed system for processing news and detecting advertisements in TV streams

The proposed system is composed of three functional blocks, grouped in consideration of the users categories: Collecting Block (CB), Editing Block (EB) and Delivering Block (DB). Each block exposes some functionality of the proposed system that we will detail later. Our work is mainly focused in TV news processing. Our system can be used by the following users: readers, journalists, editor-in-chief and broadcasting officer. Companies can access to the news portals on the Web and perform their search in the database.

3.1 Collecting Block

The collecting block consists of four modules: sources acquisition, information tracking, news clipping and advertisements extraction. Some tasks are done automatically and others are performed manually. In the following subsection we will detail each module.

The sources acquisition strategy varies according to the media to be captured: print news are digitized and stored in HDFS[4] or in NAS[5]; the broadcast contents from TV and radio channels are captured, stored by DVB-S card in database (metadata is not provided). Information crawling module is used for data capture from online news and social media, online news are downloaded using RSS feeds and social media are crawled and stored in database.

[4] Hadoop Distributed File System
[5] Network Attached Storage

News clipping is a manual method invented by libraries used for newspaper processing; in our system we adopted the same method using the new ICTs. Users can produce a digital copy of news associated by some data (title, author, media source, time or page number and date of publication) from different media sources. In the past, users listen/watch audiovisual content and store segments of news. For this we introduce audiovisual processing algorithms to automate this task. News received from the crawling module are classified by users in order to keep only the news that concern their customers, the rest is deleted. The topics of news from different sources are fused.

To identify the advertisement segments in the video file and to define the boundary (start and end) for each instance of ads, segmentation can be performed using a sliding window, a peaks detector, or a silence detector. In our proposed system, the segmentation is done by splitting the audio content on silence boundaries that correspond to the low signal energy (values less than a threshold).

3.2 Editing Block

The editing block is composed of news indexing and news validation modules, they are done only by experienced journalists. The indexing module offers the possibility to edit, annotate and summarize each news item, also validation is a confirmation that the news is important to be delivered to companies. Every multimedia news record ingested in the system and validated by editor-in-chief must be sent to companies. In the same way, the advertisements extracted automatically should be validated before the delivery process.

3.3 Delivering Block

Like news media services, the system will send daily news alerts and reports via e-mail with articles containing all news and advertisements related to client needs. Also, a summary should be sent every month. News in reports are classified by date, media source and degree of importance. A broadcasting team can customize reports and assure that customers will receive exactly the relevant news to their preferences. Features can be media type, domain, location, language, etc. Customers are classified into two categories (executives and staff). The delivery is done automatically and planned in three stages (8am, 1pm, and 4pm). However, alerts can be sent in real time.

3.4 TV News Processing

In addition to the design of the MMS and in order to enhance our system, we integrate automatic functionalities that capture continuous TV streams and save them to one hour long mp4 files using the RTSP protocol/H.264 codec, segment streams in programs, identify news program and extract text from video frames. The details of the TV news processing steps are described in Fig. 3 and detailed below.

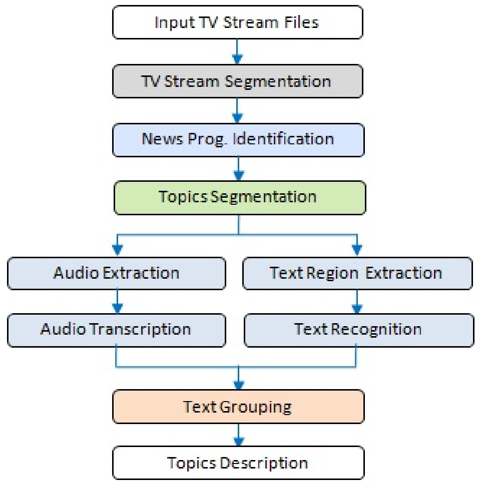

Fig. 3: TV stream processing

3.4.1 TV Stream Segmentation The first stage of news processing is to segment TV stream in programs (called inter-segmentation) using visual features for channels that includes monochrome frames or using audio features for others, because textual metadata are not provided by TV channels. One of the most popular solutions to solve this problem for the first category of channels is to use a simple monochrome frame detector that is based on standard deviation (SD) computation; each frame with SD less than a threshold T_{sd} is considered a monochrome frame and indicates the start or the end of a program. The result of this stage is a set of temporal values Sp = T_1, T_2, , T_n, consecutive values are fused.

Concerning the second category of channels, it is necessary to analyze the stream using audio features and compare them with a dataset of the audio signatures of TV programs. The pseudo code presented in Algorithm 1 shows the steps involved in TV stream segmentation process.

3.4.2 News Program Identification Automatic identification of news program within TV streams is a classification problem and can be resolved using discriminative features and distance metric. One of the best discriminative feature is the video segment called generic that is a marker used to identify the start or the end of a particular program such as news program (Special programs are characterized by an audible and visual illustration of a specific generic).

For that we construct an audio dictionary composed by a set of generic video segments from different TV channels and represented by MFCC[6] audio descriptor and HOG[7] descriptor computed for each frame. Most often, cepstral

[6] Mel Frequency Cepstral Coefficients
[7] Histogram of Oriented Gradient

Algorithm 1 Segment the TV stream and extract the news programs

```
procedure ProcessVideofile(f)
    video = OpenVideo(f)
    start = GetVideo_info(video,' start')
    end = GetVideoInfo(video,' duration')
    cores = CpuCount()
    jobs = []
    ▷ fork all jobs
    while start ¡ end do
        step ← min(start + end/cores, end)
        proc = NewsDetection(f, start, step)
        jobs.append(proc)
        start = step + 1
    end while
    segments = []
    ▷ join all jobs
    for proc in jobs do
        result = proc.join()
        segments.append(result)
    end for
    ▷ save segments
    for item in segments do
        segment_file = GetTmpfile()
        seg_video = GetSubVideo(video, item[0], item[1])
        SaveVideo(seg_video, segment_file)
    end for
end procedure
```

parameters are required and these are indicated by setting the target kind to `MFCC` standing for Mel-Frequency Cepstral Coefficients (MFCCs).

Afterward we compare each segment of video extracted from program at time $T_i\{Sp\}$ using the DTW[8] distance (Eq. 3). If the distance is less than a threshold value $T_{Generic}$ we check if the visual similarity between segment frames and dictionary frames (using Euclidian distance of HOG vectors). If they are similar, we consider that this segment correspond to a news program. Algorithm 2 shows the steps of the news identification process.

Algorithm 2 Process video segments in order to detect news programs

```
function NewsDetection(file, from, to)
    video = OpenVideo(file, from, to)
    s = 0                                                      ▷ start of segment
    e = to                                                     ▷ length of segment
    w = 3                                                      ▷ window = 3 secondes
    l = −1                                                     ▷ last position in segment
    dataset_waves = OpenDataset('generics.txt')
    results = []
    while s ¡ e do
        audio = GetAudio(segment, s, min(s + w, end))          ▷ extract audio
        mfcc = GetMFCC(audio)                                  ▷ compute MFCC features
        dist = 999999
      ▷ compute distance between generic sequences
        for item in dataset_waves do
            dist = DTW(item, mfcc)
            if dist < 90 then
                break
            end if
        end for
    ▷ save and group consecutives positions/slide window over segment with w step
        if dist < 90 then
            if (l == −1) or (s − results[l, 1]) > 1) then
                results.append([s, s, item.class])
                l = l + 1
            else
                results[l, 1] = s
            end if
            s = s + w
        else
            s = s + w/2
        end if
    end while
    return (results)
end function
```

[8] Dynamic Time Warping

We have used the MFCC implementation of McFee et al. [20], DTW implementation of Pierre Rouanet[9], and we identify jingle to do intra-segmentation and classification of news in topics (national, international, sport, business,). The next stage is to process the news program.

3.4.3 Topics Segmentation For each news program extracted in previous steps, we proceed to extract news topics by the following steps :

- Anchorperson identification using a clustering of all faces centered in screen,
- Select cluster that have the frequency of repetition of the same face (we apply TF/IDF techniques used in textual document indexing),
- Extract segments where the anchorperson is located,
- Save extracted segments.

We have defined somme criteria to select faces in video frames such as:

- Face location: Based on the coordinates (x,y) of faces and returned by the detector, we only retain faces that are located in the middle of the screen.
- Face size: The size of the presenter face has a proportional value to the size of the whole image.
- Face distance: We estimate the distance between the camera and the human head by computing the distance between the right eye and the left eye [11].
- Face motion: Both anchorperson and camera are motionless, based on that we eliminate some moving faces shots.

The pseudo code to do this task is described in Algorithm 3.

3.4.4 Audio Extraction For every topics segmented in previous steps, the audio content will be extracted and saved in a temporary file in order to do a transcription of content and get the corresponding text.

3.4.5 Audio Transcription Several online frameworks offer the audio transcription service for non commercial use by using their API. In our context, develop an audio transcription tools is another challenge that only large companies can effectively deal with like Google and IBM. For this, we have opted to use speech to text software to produce transcription of the news.

3.4.6 Text Region Extraction Automatic extraction and recognition of text embed in news video provide an efficient approach to annotate TV news contents, one or several rows of text appear in the screen bottom (in the scope of ¼ of screen height) and express the meaning of news specked. Text region extraction steps are shown in Fig. 4. The final result of this step is a set of rectangular regions bounding text lines that will be submitted to the transcription stage.

[9] https://libraries.io/github/pierre-rouanet/dtw

Algorithm 3 News topics segmentation

```
function TopicsSegmentation(file, from, to)
    video = OpenVideo(file, from, to)
    s = 0                                              ▷ start of segment
    e = to                                             ▷ length of segment
    middle = video.size[0]/2                           ▷ middle of the screen
    fc = [−1, −1, 0, 0, 0, 0]
    prev_face = []
    results = []
    while s ¡ e do
        img = GetFrame(video, s)                       ▷ extract image
        faces = DetectFaces(img)                       ▷ detect all faces
        for face in faces do
            if prev_face == [] then
                prev_face = face
            end if
            cam_dist = face.righteye − lefteye
            face_size = face.w ∗ face.h ∗ 100.0/(video.size[0] ∗ video.size[1])
          ▷ check criteria to retain face position
            if (cam_dist > 25) and (cam_dist < 55) and (face_size > 1) and
             (face.x < middle) and (face.x + face.w > middle) then
                if fc[0] == −1 then
                    fc = [s, s, 0, face.x, face.y, face.w, face.h]
                else if (s − fc[1]) > 1 then
                    results.append(fc)
                    fc = [s, s, 0, face.x, face.y, face.w, face.h]
                else                                   ▷ check face motion
                    face_dist = sqrt(pow(face.x − prev_face.x, 2)+
                        pow(face.y − prev_face.y, 2))
                    if face_dist < 60 then
                        fc[1] = s
                    else
                        results.append(fc)
                        fc = [s, s, 0, face.x, face.y, face.w, face.h]
                    end if
                end if
            end if
            prev_face = face
        end for
        s = s + 1
    end while
    if (fc[1] − fc[0]) > 0 then
        results.append(fc)
    end if
    return (results)
end function
```

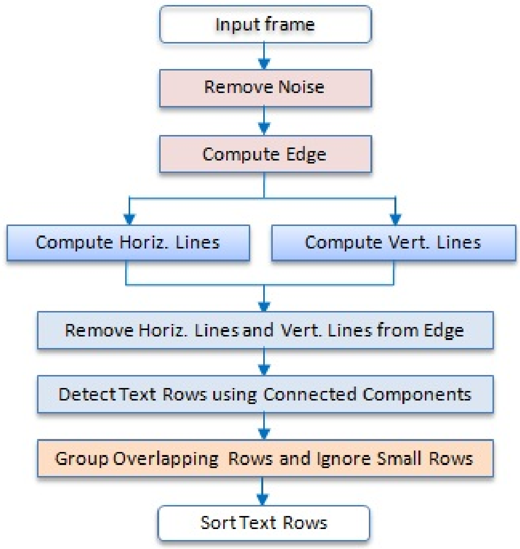

Fig. 4: Text region extraction

3.4.7 Text Transcription The text regions extracted are binarized and transmitted to the OCR (Optical Character Recognition) engine that process each text region and return equivalent textual information. There are few OCR systems capable of recognizing Arabic text, namely: Automatic Reader produced by the Sakhr[10] Software Company; FineReader[11] produced by the ABBYY Company and Tesseract[12] produced originally by Hewlett-Packard and available from Google. We are interested in Tesseract because it is an open-source OCR.

3.4.8 Text Grouping In the final post-processing of news processing, we use combined speech transcription and image understanding technology to represent the TV news by a bag of words that will exploited for describing the content and the news topics.

3.4.9 Advertisements Detection and Classification Given a video file, the first step is video segmentation based on audio content. It allows to obtain a set of small segment that probably contains advertisement. In our proposed system, we use a classifier based on a deep learning model composed of a Batch-Normalization hidden layer for data normalization, with Dense and Dropout hidden layers. We use a Rectified Linear Unit (ReLU) and hyperbolic tangent (TanH) activation functions. The input to the model consists of a combination

10 www.sakhr.com
11 www.abbyy.com
12 code.google.com/p/tesseract-ocr

of MFCC features computed over 25ms of an audio signal with a frame size of 3s. The softmax output layer contains the probabilities of predicted classes. The following Fig. 5 describes the architecture of the classifier for the ads classification.

A common problem in analyzing continuous stream is how to determine which instances of the processed data stand out as being dissimilar to the trained data. Such instances are known as outliers or anomalies. For that, we propose to align the predicted class with the trained data of the same class using the dynamic time warping distance, the predicted class is accepted if the distance DTW is less than a threshold.

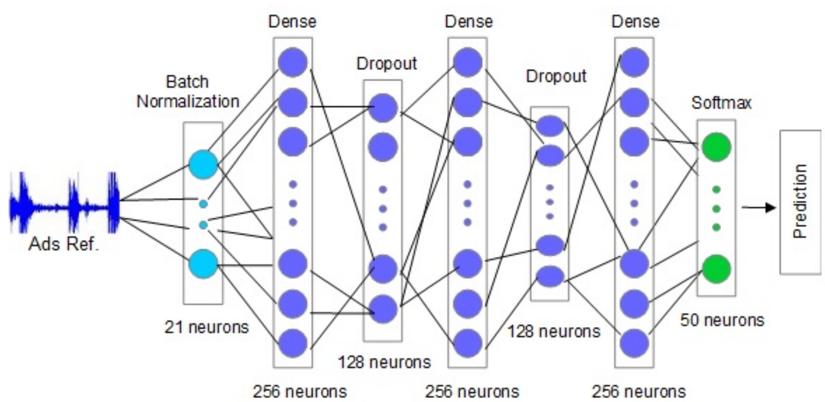

Fig. 5: Architecture of the audio classifier

4 Experimentation

The data used in our experiments has been collected by MediaMarketing[13] company specialized in news monitoring, dataset is composed of five different national channels in Algeria that spoke arabic, french or english. The majority of channels are recorded at a resolution of 720x480 pixels. We have took a sample with 24h times recording for each channel, the recording streams are divided into one-hour portions.

For the experiments, we have created a dataset composed of the generics of the news programs as well as the jingles of advertisements relating to each TV channels and which are represented by the MFCC features and using the library LibROSA, an open-source Python package for music and audio analysis. In order to compute the similarity between the dataset and the TV stream we have used the DTW distance (best threshold = 90), we divided the speech

[13] www.mediamarketing-dz.com

signal into short frames, e.g. 3 seconds segments, and processed each frame as a single unit. We have used a classical evaluation process; we compare the manual annotation provided by MediaMarketing (ground truth) with results obtained by our system.

To choose a discriminative and efficient descriptor used in the learning part, we tested the MFCC (audio), HOG (visual) and MFCC+HOG (Table 1). Finally, we have chosen the MFCC descriptor alone and to increase the relevance, we proceeded to the improvement of our segmentation algorithm by grouping the marks which are close to each other.

Features	**Relevance**	**Execution time**
MFCC	93%	6mn
HOG	95%	14mn
MFCC + HOG	98%	16mn

Table 1: TV Stream Analysis (1 hour) with different descriptors

In order to reduce the time execution, we performed experiments in parallel-processing; we have tested a processor with 4, and 12 cores (Table 2).

Processors	**Execution time**
1 core	25mn
4 cores	14mn
12 cores	6mn

Table 2: Multicores CPU parallel processing

In Topics segmentation level, we tested different models of face detection and description like MTCNN, DLIB, OpenFace. Therefore, we have used DLIB library to do face detection because it is the more accurate in our task.

To evaluate the performance of the proposed system, we have used the precision, recall and F-score metrics (Eq. 4, 5 & 6), that are described as follows:

- Precision (P) is defined as the number of hits over the number of hits plus the number of false alarms.

$$P = \frac{\#hits}{\#hits + \#false\ alarms} \tag{1}$$

- Recall (R) is defined as the number of hits over the number of hits plus the number of misses.

$$R = \frac{\#hits}{\#hits + \#misses} \tag{2}$$

- F-score (F) is defined as the harmonic mean of the precision and the recall, the formula is:

$$F - score = 2.\frac{P.R}{P + R} \tag{3}$$

TV channel	News program detection			Topics segmentation			Text region extraction		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
A3	0.93	0.90	0.91	0.88	0.84	0.81	0.83	0.78	0.80
Canal Algerie	0.91	0.94	0.92	0.85	0.90	0.87	0.85	0.76	0.80
Ennahar News	0.90	0.89	0.89	0.90	0.93	0.91	0.79	0.72	0.75
Dzair News	0.95	0.93	0.94	0.92	0.91	0.91	0.80	0.87	0.83
Echorouk News	0.96	0.95	0.95	0.89	0.92	0.90	0.88	0.83	0.85
All channels	**0.93**	**0.92**	**0.92**	**0.88**	**0.90**	**0.88**	**0.83**	**0.79**	**0.80**

Table 3: System performance results

The experimentation results of each bloc are shown in Table 3.

For the Ads classification model, we exploited 50 classes of advertisements to train our classifier, and we fixed the threshold value (DTW) to 20. We obtained the accuracy for training and testing equal to 1.0 and loss values equal to 0.001. The obtained results for the different metrics are:

- Precision: **0.92**
- Recall: **0.73**
- F-measure: **0.81**

The obtained results shows high precision of our proposed system, but there are still improvements to be done in text extraction from the video frames. The best results are observed in news program detection and advertisements classification. We can say that we can generalize the proposed process on other TV channels and the result will also be relevant. Unfortunately, we can't compare our results with other works because there are no shared datasets.

5 Conclusion & future works

In this paper, we have proposed a global architecture of our multimedia monitoring system, that processes newspaper, online news, broadcast news and social media. This system is a service for a targeted audiences such as large corporations and state institutions. It offers a digital storage of all news clips, faster delivery via e-mail and guaranteed near zero missed news clips. We have tried to automate some tasks of the system's workflow processes, the results obtained are satisfactory, we intend to extend this work on the treatment of radio channels and we wish to parallelize some treatment with Spark framework or implement our algorithms on a different multi-core GPU.

However, we hope to do exhaustive comparison with others methods using a public datasets such as ALIF (A Dataset for Arabic Embedded Text Recognition in TV Broadcast). In future work we plan to do an exhaustive analysis of TV News for the identification and segmentation of topics. Among the future work, we will investigate three tasks which are:

- News story categorisation using NLP processing, each topics must be classified in one of the categories: politics, economics, health, sports, ...),

– Alignment of segmented news topics with textual news collected by the RSS Feeds,
– Design of an autoencoder for outliers detection in advertisements classification.

References

1. Abduraman, A.E.: Unsupervised tv program structuring (2013)
2. Aggoun, Y.: Procd de traitement, de prsentation et de visualisation synoptiques de linformation pour les dcideurs (2011), algeria Patent
3. Aggoun, Y.: La veille presse. de la culture mosaque aux outils d'aide la dcision (2015)
4. Amrane, A., Mellah, H., Aliradi, R., Amghar, Y.: Semantic indexing of multimedia content using textual and visual information. International journal of advanced media and communication **5**, 182–194 (2014)
5. Berrani, S.A., Lechat, P., Manson, G.: Tv broadcast macro-segmentation: Metadata-based vs. content-based approaches. pp. 325–332 (2007)
6. Beyan, C., Fisher, R.: Classifying imbalanced data sets using similarity based hierarchical decomposition. Pattern Recognition **48**, 1653–1672 (2015)
7. Comcowich, W.J.: Media monitoring: The complete guide. Cyber alert Inc, White paper (2010)
8. Duan, L.Y., Wang, J., Zheng, Y., Jin, J.S., Lu, H., Xu, C.: Segmentation, categorization, and identification of commercial clips from tv streams using multimodal analysis. pp. 201–210 (2006)
9. Dumont, E., Qunot, G.: Automatic story segmentation for tv news video using multiple modalities. International journal of digital multimedia broadcasting **2012** (2012)
10. Escalada, J.G., Duxans, H., Conejero, D., Asensio, A., Salinas, D.: Newsclipping: An automatic multimedia news clipping application. pp. 1–6 (2010)
11. Flores, A., Christiansen, E., Kriegman, D., Belongie, S.: Camera distance from face images. pp. 513–522 (2013)
12. Ghosh, H., Kopparapu, S.K., Chattopadhyay, T., Khare, A., Wattamwar, S.S., Gorai, A., Pandharipande, M.: Multimodal indexing of multilingual news video. International Journal of Digital Multimedia Broadcasting **2010** (2010)
13. Gonina, E., Friedland, G., Battenberg, E., Koanantakool, P., Driscoll, M., Georganas, E., Keutzer, K.: Scalable multimedia content analysis on parallel platforms using python. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **10**, 1–22 (2014)
14. Kannao, R., Dandi, D., Yellapu, S., Guha, P.: News program detection in tv broadcast videos. pp. 546–550 (2016)
15. Kimiyama, H., Maruyama, M., Kobayashi, M., Sakai, M., Mandala, S.: An uhd video handling system using a scalable server over an ip network. International Journal of Advanced Media and Communication **7**, 1–19 (2017)
16. Kompatsiaris, Y., Merialdo, B., Lian, S.: TV content analysis: Techniques and applications. CRC Press (2012)
17. Li, H., Jou, B., Ellis, J.G., Morozoff, D., Chang, S.F.: News rover: Exploring topical structures and serendipity in heterogeneous multimedia news. pp. 449–450 (2013)
18. Manson, G., Berrani, S.A.: Automatic tv broadcast structuring. International journal of digital multimedia broadcasting **2010** (2010)

19. Manson, G., Naturel, X., Berrani, S.A.: Online macro-segmentation of television streams. pp. 220–221 (2009)
20. McFee, B., Raffel, C., Liang, D., Ellis, D.P.W., McVicar, M., Battenberg, E., Nieto, O.: librosa: Audio and music signal analysis in python. vol. 8, pp. 18–25 (2015)
21. Meinedo, H.: Audio pre-processing and speech recognition for broadcast news. Universidade Tcnica de Lisboa, Diss (2008)
22. Minaee, S., Bouazizi, I., Kolan, P., Najafzadeh, H.: Ad-net: Audio-visual convolutional neural network for advertisement detection in videos. arXiv preprint arXiv:1806.08612 (2018)
23. Nedjah, N., da Silva, F.P., de S, A.O., Mourelle, L.M., Bonilla, D.A.: A massively parallel pipelined reconfigurable design for m-pln based neural networks for efficient image classification. Neurocomputing **183**, 39–55 (2016)
24. Nouza, J., Zdansky, J., David, P.: Fully automated approach to broadcast news transcription in czech language. pp. 401–408 (2004)
25. Pinquier, J., Andr-Obrecht, R.: Jingle detection and identification in audio documents. vol. 4, p. iviv (2004)
26. Ramires, A., Cocharro, D., Davies, M.E.P.: An audio-only method for advertisement detection in broadcast television content. arXiv preprint arXiv:1811.02411 (2018)
27. Rouvier, M., Favre, B., Bendris, M., Charlet, D., Damnati, G.: Scene understanding for identifying persons in tv shows: beyond face authentication. pp. 1–6 (2014)
28. Sdergrd, C., Aaltonen, M., Hagman, S., Hiirsalmi, M., Jrvinen, T., Kaasinen, E., Kinnunen, T., Kolari, J., Kunnas, J., Tammela, A.: Integrated multimedia publishing: combining tv and newspaper content on personal channels. Computer Networks **31**, 1111–1128 (1999)
29. Smith, J.R., Campbell, M., Naphade, M., Natsev, A., Tesic, J.: Learning and classification of semantic concepts in broadcast video (2005)
30. Wang, J., Duan, L., Liu, Q., Lu, H., Jin, J.S.: A multimodal scheme for program segmentation and representation in broadcast video streams. IEEE Transactions on Multimedia **10**, 393–408 (2008)
31. Yousfi, S., Berrani, S.A., Garcia, C.: Deep learning and recurrent connectionist-based approaches for arabic text recognition in videos. pp. 1026–1030 (2015)
32. Zayene, O., Seuret, M., Touj, S.M., Hennebert, J., Ingold, R., Amara, N.E.B.: Text detection in arabic news video based on swt operator and convolutional auto-encoders. pp. 13–18 (2016)
33. Zhang, J.: Extraction of text objects in image and video documents. University of South Florida (2012)
34. Zlitni, T., Bouaziz, B., Mahdi, W.: Automatic topics segmentation for tv news video using prior knowledge. Multimedia Tools and Applications **75**, 5645–5672 (2016)