Automatic Topics Segmentation for News Video by Clustering of Histogram of Orientation Gradients Faces

Mounira Hmayda, Ridha Ejbali, and Mourad Zaied
RTIM: Research Team in Intelligent Machines,University of Gabes, National Engineering School of Gabes (ENIG), Tunisia

Abstract: *TV stream is a major source of multimedia data. The proposed method aims to enable a good exploitation of this source of video by multimedia services social community, and video-sharing platforms. In this work, we propose an approach to the automatic topics segmentation of news video. The originality of the approach is the use of Clustering of Histogram of Orientation Gradients (HOG) faces as prior knowledge. This knowledge is modeled as images which governs the structuring of TV stream content. This structuring is carried out on two levels. The first consists in the identification of anchorperson by Single-Linkage Clustering of HOG faces. The second level aims to identify the topics of news program due to the large audience because of the pertinent information they contain. Experiments comparing the proposed technique to similar works were carried out on the TREC Video Retrieval Evaluation (TRECVID) 2003 database. The results show significant improvements to TV news structuring exceeding 96 %.*

Keywords: *Anchorperson, clustering, face detection, features extraction, news program*.

Received December 28, 2018; accepted April 10, 2020
https://doi.org/10.34028/iajit/18/3/2

1. Introduction

Due to the presence of large quantities of content videos, effective and rapid access to multimedia information has become more challenging. Therefore, there has been an acute need for appropriate methods that permit quick access to the content of unstructured video contents, which may involve automatic video segmentation and indexing methods. The inter-segmentation of video streams is an essential step for a multimedia indexing system. However, for it to be robust, it must deal with the internal structuring of the programs to promote their audiovisual content.

Therefore, we focus particularly on the analysis and automatic identification of television news structure. This choice can be justified by many reasons: The emergence of mass-communication devices and services, such as Television Replay or video on demand (TV-REPLAY or VoD). The ubiquity of huge volumes of digital videos, especially TV streams. Thus, the use of TV streams for communication is becoming a staple in everyday life and therefore poses a substantial financial durden. In addition, effective TV stream structuring could be the objective of various applications. For example, the identification of TV programs could serve to develop control mechanisms for TV channels broadcasting.

The present work focuses on the automatic identification of News Program (NP) structure in order to segment the content of news into different topics. The segmentation of TV streams permits a more feasible exploitation of their contents via an easier access and distribution on communication devices dedicated to TV channels, large public (Internet Protocol Television, Television Replay) or sharing platforms (YouTube, Dailymotion, Vimeo, Facebook,…).

The remainder of this paper is organized as follows. Section 2 deals with related works. In section 3, the different structures of TV news are introduced. We present the proposed approach in section 4 and the news topics segmentation in section 5. In section 6, we describe the experiments and evaluate the results. To conclude, we discuss the solution and suggest a framework for future work.

2. Related Work

According to an analysis conducted in October 2010 by the National Center for Cinema and Animated Image, France (CNC2), television news accounts for more than 50% of the TV-replay television offer. It constitutes an important field of multimedia indexing. Extensive research has been dedicated to the indexing of TV news [12] according to different modalities. In fact, in the literature, a major part of TV news indexing uses the tools for signal processing to divide different topics into segments. These processing methods principally focus on the elicitation of low-level features, namely colors, forms, intensity and

texture or high-level features, such as anchorperson detection or shots detection. Video segmentation is the first important step to video content analysis. It aims at dividing the video stream into a series of significant and controllable segments (shots) that can serve as essential elements for indexing [8]. In fact, video shots segmentation is the primary process of anchorperson detection which has an central role in further video processing.

O'Hare *et al*. [13] adopted a framework in which a news broadcast can be partitioned into separate stories according to the position of the anchorperson shots in the program. After segmenting a program into distinct shots, a number of analyses are run on it to identify the features that characterize each shot. The outcomes of these feature extraction tools are afterwards combined using a Support Vector Machine trained to perceive anchorperson shots. A system based on dividing a news video into stories and classifying the spotted stories into certain categories (world news, national news, sports, political news, weather, advertising, etc…,) was presented in [1]. The system uses Markov chains and Bayesian networks for partitioning and topics classification. The entire process is conducted using data extracted from video and audio tracks using techniques of superimposed text recognition, speaker identification, speech transcription and anchor detection. The segmentation of news is based on feature recognition, such as anchorperson shots or interviews.

Goyal *et al*. [6], however, suggested a frame-work for the segmentation of semantic stories on the basis of anchorperson detection. They opted for a mechanism of split-and-merge in order to detect topic boundaries. This technique focuses on visual characteristics and transcripts of texts. Dumont and Quénot [2] developed a sensor using Multiple Modalities for Systematic segmentation of stories for News Videos. This system is developed based on classification techniques and machine learning methods. It combines both audio descriptors (silence segments and parole) with visual features such as anchors or logos. Poulisse *et al*. [14] proposed an approach using multiple multimedia features for segmenting news video. It was inspired from the approach based on text segmentation. The authors opted for different methods to achieve topic segmentation with different approaches: text, video, audio and layout features. At an advanced step of the analysis, those features are used to detect story breaks after training a maximum entropy classifier. To structure TV streams, Hmayda *et al*. [7] presented an approach in which deep learning is used to identify programs in TV stream.

Two main reasons may justify the choice of news programs treatment. First, thanks to the important content they have, news programs are produced for and followed by many TV viewers. In fact, as they are frequently published on the web, users get easy access to them. Thus, they have become a means of communication.

News has an operating structure defined as a syntactic rule that governs the organization of the content. This rule is founded on the studio/Topic notion, in which the anchorperson is the main performer announcing the changes in topics. The Topic is the demonstration of what the anchorperson says in the studio. Each subject usually has an average duration between 1 and 3 minutes [9] and is preceded by the appearance of an anchorperson.

3. Proposed Approach

We suggest an approach of structuring news content based on segmentation into topics. This approach is based on two major steps: First; the identification of anchorperson by Single-Linkage Clustering of the Histogram of Orientation Gradients (HOG) faces (Figure 1); Then, the segmentation of news into topics. The first step consists in modeling the indices by image processing techniques, whereas the second step relies on the direct exploitation of the structures ensuing from the first step to segment the content of news into different topics.

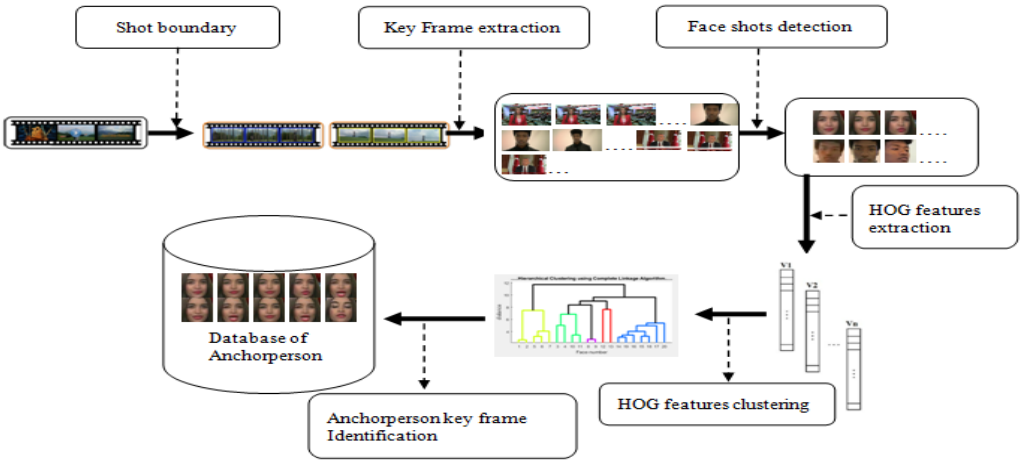

Figure 1. Identification of anchorperson by single-linkage clustering of HOG faces.

3.1. Identification of Anchorperson by Single-Linkage Clustering of HOG Faces

At that stage, no prior information about the structural features of NP is provided. It is, therefore, essential to implement a fully automatic process for modeling this structure by means of image processing characteristics. This is realized in three steps: step one is dedicated to shot boundary detection. The second one involves the identification of shots containing the face of anchorperson using optical flow to extract the keyframe from these shots. The last step is based on unsupervised classification to identify the anchorperson by Single-Linkage Clustering of HOG faces.

3.1.1. Shot Boundary Detection

Segmentation into shots is considered among the first contributions to the analysis and structuring of the video by content. This task is the starting point of any process of macro-segmentation of video content. Indeed, the resulting units (shots) of this segmentation constitute the inputs of any phase of content structuring. They are also the basis for identifying other units with a semantic level of higher granularity, such as topics for the case of News.

The variety of segmentation into shots techniques proposed in the literature is strongly linked to the diversity of the types of changes (abrupt transition, fade, etc.,). Yuan *et al*. [18] conducted an analytical study of the main approaches proposed in the literature. Based on this study, we opte in the present paper for the Edge Change Ration (ECR) [9] technique based on its performances in shots detection. This method consists in changing the edges of the objects (Figure 2) in the frames definitely across a border. Thus, structural disjointedness is accompanied by temporal visual discontinuity.

a) Original picture. b) Picture with edges.

Figure 2. The edges in two consecutive video frames.

On the basis of this assumption, the technique starts with the calculation of the percentage of incoming and outgoing contours between two images.

Thus, the value of ECR (*n*, *k*) between the images *n*-*k* and n is calculated as in (Equation (1)):

$$ECR(n,k) = \max\left(\frac{x_n^{in}}{\sigma_n}, \frac{x_{n-k}^{out}}{\sigma_{n-k}}\right) \quad (1)$$

Where σ_n is the number of edge pixels in the frame *n* and x_n^{in} and x_{n-k}^{out} are the entering and exiting edge pixels in frames *n* and *n*-*k*, respectively.

During the shots' recognition, the ECR technique is used with k=10 for a temporal distance of 10 images. For the detection of hard cuts, we calculate two values. First, we compute the near-far ratio which describes the ratio between the ECR values of two successive images (near ECR) and the ECR value between the current image and the 10th image (far ECR). Then, the far last-far ratio, which is the ratio between the current value of far ECR and the previous value of far ECR, is calculated.

This phase consists in the assembly of all the shots composing the news. In order to delimit the topics, these shots will then undergo two levels of filtering: the first aiming at extracting the keyframe from these shots and the second at the identification of anchorperson.

3.1.2. Key Frame Extraction Using Optical Flow

Once the news video is segmented into shots, the keyframes are retrieved from these shots. The extraction of keyframes (representative frame) has an important impact on the performance of content multimedia, like the extraction of the keyframes in the video and the selection of typical view for objects [4]. The shot of anchorperson lasts more than 2 s, thus shots with a lifetime less than 2 s cannot be anchorperson shots. In addition, the frames in one anchorperson shot are comparable, whereas news report shots are different in terms of camera and object movement. Therefore, if there are many changes in one shot, it cannot be considered as an anchorperson shot. The frames of an anchorperson shot are very analogous (Figure 4), whereas those of the news report shot (Figure 5) are different due to movement of the camera. This difference can be detected by the difference between the first and the fourth frame based on optical flow in order to examine the variation in one shot.

The extraction of keyframes algorithm is depicted as follows:

- *Step* 1: we examine the duration of each shot and cast those shots with a lifetime lower than 2s.
- *Step* 2: using the differential method of Lucas and Kanade applied by Gnouma *et al*. [5] and Khondaker *et al*. [10], we calculate the moving area value between the first and the fourth frame positions for each shot.

This method calculates the minimum of the quadratic matching function; this function can estimate the parameters of a transformation affecting all parts of the frame. The movement calculation is to associate each pixel *(x,y,t) with* a vector (v_x^t, v_y^t) representing the projection onto the plane of the velocity vector (v_a^t, v_b^t, v_c^t) of the objects in the frame. The derivatives

of the matching function are canceled in relation to d_a and d_b

$$(v_a^t, v_b^t) = \arg \min_{(\delta a, \delta b) \in K} A(\delta a, \delta b) = \arg \min_{(\delta a, \delta b) \in K} \sum_{(a,b) \in B} \left(I(a,b,t) - H(a+\delta a, b+\delta b, t+1) \right)^2 \quad (2)$$

Supposing that the movement (d_a, d_b) is small, then the Taylor expansion to order 1 of H becomes:

$$H(a+\delta a, b+\delta b, t+1) \approx H(a,b,t) + \frac{\partial H}{\partial a}\delta a + \frac{\partial H}{\partial b}\delta b + \frac{\partial H}{\partial t} \quad (3)$$

To explain the representation of the route of the optical flow field, the vectors in the original image are superimposed speeds. The color map can also be used to characterize the direction of flow as well as its strength (see Figure 3).

In the color map, the velocity vectors are denoted by the colors contained in the circle. Each vector is determined by the color that designates its origin at the center of the circle. The intensity differs from black to full color until a maximum speed to exhibit white.

a) Video picture. b) Next picture.

c) Intensity of white is very low to designate an anchorperson shot and The intensity of white is very important to indicate a report shot.

Figure 3. Real-time estimation of Optical Flow for keyframe extraction.

If the intensity of light between the first and fourth frames position is less than 30, then it can be concluded that this is an anchorperson shot, if not, i.e., if the value of the intensity is very important and higher then 50, we can consider it as a report shot.

a) Frame 1. b) Frame 2. c) Frame 3.

Figure 4. Anchorperson shot frames.

a) Frame 1. b) Frame 2. c) Frame 3.

Figure 5. News report shot frames with object movement.

3.1.3. Face Shots Detection

Once the step of keyframe extraction is conducted, the various keyframes of all shots are identified. The identification of the anchorperson shots consists of identifying all the shots in which the face of the anchorperson is visible. Then, to detect the face shots, the Viola–Jones algorithm is used. This algorithm consists in using detectors depending on the overall features of objects (faces). This is the case of Haar descriptor [17], in which a set of functions is used to identify the difference in the contrast between multiple contiguous rectangular regions in an image. Fourteen descriptors are therefore computed relative to fourteen filters of rectangular shapes for coding the different contrasts existing in a face as well as their spatial relationships (Figure 6).

These filters make it possible to estimate the variance between the entirety of the pixels in the white areas as well as that of the black areas. The value of a descriptor is calculated as follows:

$$x_i = \Sigma \text{ (pixels in black area)} - \Sigma \text{ (pixels in white area)} \quad (4)$$

a) Edge features.

b) Line features.

c) Center-surround features.

Figure 6. Extended set of Haar features used as simple filters (Image source: OpenCV manual).

3.1.4. Identification of Anchorperson Shots

The principal contribution of our method is to classify all the faces detected according to an unsupervised classification. Consequently, the anchorperson face is part of the largest cluster.

To achieve this classification, we go through the step of extracting the features of all the faces, as inputs to the classification process using the HOG technique.

Features extraction: every keyframe is split up into cells size of 4 pixels and 8 orientation bins for the cell histograms by means of HOG descriptor, for every cell gathering a histogram of gradient direction over the cell pixels. Histogram labeling can be achieved through the accumulation of the energy measure of the local histogram over blocks to get the best invariance

to illumination. The results are used to standardize all cells found in block, whose sequence grants the descriptor vector to be further used.

f is an intensity function that describes the image under examination. The image is composed of cells with a size of N×N pixels (Figure 7-a), while the orientation $\theta_{x,y}$ of the gradient in every pixel (Figure 7-b, 7-c) is calculated using the following rule:

$$\theta_{x,y} = \tan^{-1} \frac{f(x, y+1) - f(x, y-1)}{f(x+1, y) - f(x-1, y)} \quad (5)$$

Subsequently, the orientations θ_i^j i=1...N^2, i.e., belonging to the same cell j, are quantized and gathered into an M-bins histogram (Figure 7). All the obtained histograms are ordered and integrated into a unique HOG histogram that is the final product of this algorithmic step, i.e., the features vector to be considered for the following processing.

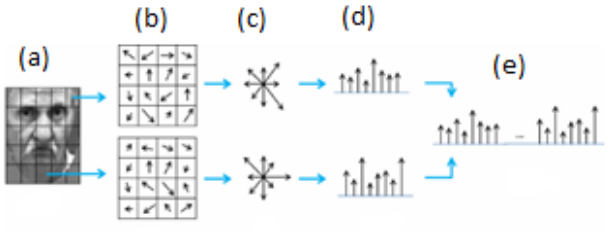

Figure 7. HOG features extraction process.

b. Classification using Hierarchical Clustering Algorithm HOG features vectors are then given as input to Hierarchical Clustering Algorithm. This algorithm partitions the related dataset of key frame by building a hierarchy of clusters.

It utilizes the distance matrix requirements for grouping the key frame and builds clusters gradually. In hierarchical clustering [16], in single step, the information is not separated into a particular cluster. It takes a sequence of partitions, which may run from a single cluster covering all objects to 'n" clusters, each comprising only one object.

In the present work, we have a series of N key frame (faces) to be clustered, thus these steps should be followed:

- *Step* 1: Start with assigning each face to a cluster so that for N faces, we now have N clusters, each of which displays only one face. Let the distances between the clusters be the same as those between the faces they comprise.
- *Step* 2: Determine the nearest pair of clusters and combine them in one cluster. Thus, there is one cluster less.
- *Step* 3: Calculate the distances (similarities) between the new cluster and every ancient cluster.
- *Step* 4: Replicate steps 2 and 3 until all faces are clustered into one cluster of size N

Single-linkage clustering is used in step 3: we consider that the distance between two clusters is equal to the shortest distance from any face of one cluster to any face in the other.

Algorithm of Single-Linkage Clustering:

```
L : Level of clustering
m : sequence number
n: number of clusters
Ci : a cluster, i belongs to {1,..,n}
r, s, j belong to {1, .., n}
D : the proximity matrix, D(i,j)=d(Ci,Cj)
Start
L(0)=0, m=0, min = 10**20 (very big number ~ infinity)
while (n not equal to 1) :
    for i in range(n-1) :
        for j in range(i+1,n) :
            if d(Ci,Cj)<min :
                min = d(Ci,Cj)
                (r,s) = (i,j)
    m = m+1
    L(m) = d(Cr,Cs)
    k=r
    Ck = merge(Cr,Cs)
    for i in {1,..,n} :
        remove D(i,s)
        remove D(s,i)
    for i in {1,..,n-1} :
        D(k,i)=min(d(Cr,Ci),d(Cs,Ci))
        D(i,k)=D(k,i)
        n=n-1
```

After the hierarchy of clusters is provided, the optimal number of clusters is presented based on a data representation tree. Therefore, the cluster which contains the largest number of faces is considered an anchorperson. The data representation tree in blue (Figure 8) is considered an anchorperson.

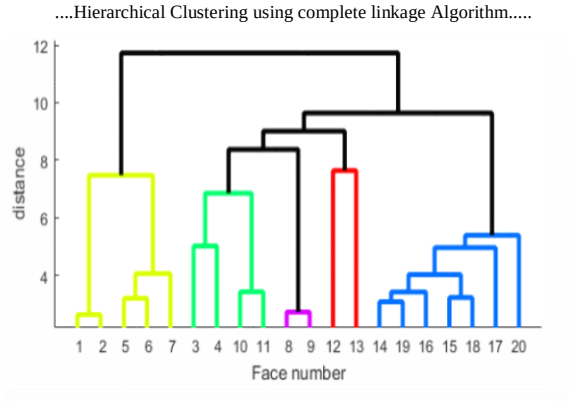

Figure 8. Hierarchical c+lustering of face.

4. News Topics Segmentation

The structuring of TV news content by segmentation into topics is conducted following two modes: Stand-alone mode or Rewire mode. 0The former is operated in off-line and is employed when the NP we want to structure is presented for the first time as input to the proposed scheme. In this case, all the steps of the learning phase of the structure are performed.

Segmentation into topics is, therefore, an immediate use of the outcomes reached at the output of the anchorperson identification step using the technique of face recognition (Figure 9). The latter can be operated on-line. Indeed, in the case where the NP has already been processed once by our system, it is obvious that it is stored in a database. This database is composed of a set of anchorpersons. Thus, when the NP is presented a second time as input to the system, the frames of anchorperson are deployed from the database and the problem of structuring is reduced to discover the presenter in the video stream. Anchorperson detection can be achieved on-line using the technique of face recognition [3]. If the anchorperson is detected, a change of subject is reported.

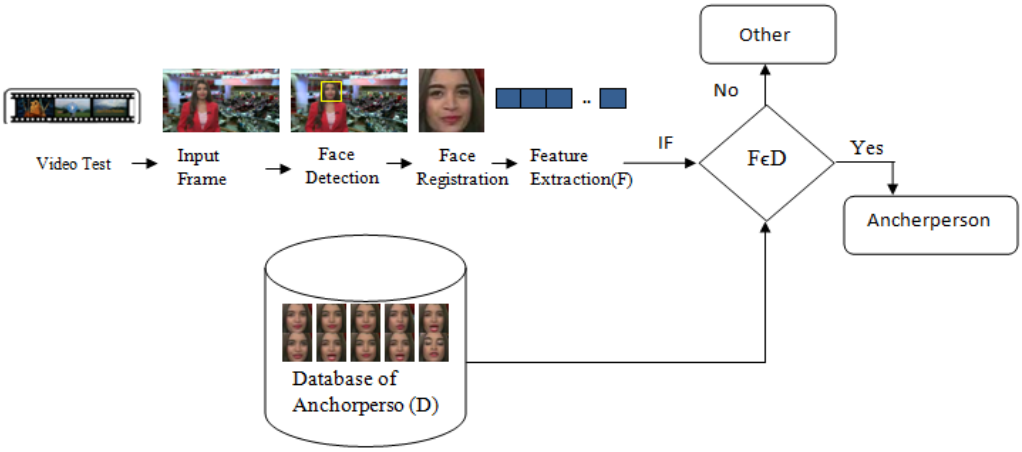

Figure 9. Steps in the face recognition for topics segmentation.

5. Experiments and Results

At this level, two experiments were conducted. We built a dataset of news program from different channels to achieve proportional results with previous works and we carried out a second experiment on the TREC Video Retrieval Evaluation (TRECVID) dataset.

The performance of the proposed algorithm was assessed in terms of precision rate, recall rate and F1. The terms are defined as follows:

$$\text{Precision rate} = \frac{\text{Correct}}{\text{Correct+False}} \times 100\%$$

$$\text{Recall rate} = \frac{\text{Correct}}{\text{Correct+Missed}} \times 100\%$$

$$\text{F1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

With:
Correct: Number of programs identified correctly
False: number of false identifications
Missed: number of missed identifications

5.1. Experiments on Dataset News from TV Streams

In the first experiments, we used ten recordings of news programs from different channels. Among which, two news program recordings were used for each channel: the first one for the segmentation in stand-alone mode and the other for the rewire mode. We also founded a ground truth which contained the number of shots, face shots and anchorperson shots for each news program (Table 1). First, we conducted the recognition of all shots and the face identification. Good rates for shots and faces shot (Figure 10) enhanced anchorperson shot identification and therefore topic identification.

Table 1. Ground truth of TV news segmentation.

	TF1		LCI		France24		Itele		M6	
	NP1	*NP2*	*NP1*	*NP2*	*NP1*	*NP2*	*NP1*	*NP2*	*NP1*	*NP2*
Shot	353	290	350	310	414	370	360	270	446	442
Face Shot	250	160	186	120	231	198	210	135	280	275
Anchor person Shot	50	46	58	35	117	106	55	30	75	68

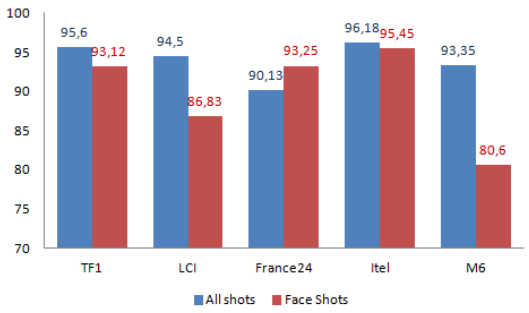

Figure 10. Face and shot detection.

We then carried out the evaluation of our approach using "stand-alone" and "Rewire" modes (Figure11)

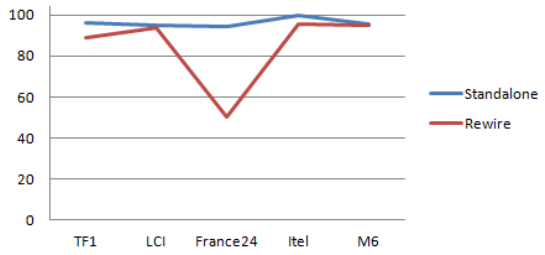

Figure 11. Rate of topics detection.

Results demonstrate very acceptable performance in both modes since with an average detection rate of about 90%, there are only 10% missed detections. This

rate of missed detections can be explained, first, by the missed detections of the face shots and then by the low detection rate attained in the case of the France 24 NP with "Rewire" mode. In fact, in this case the anchorperson in NP2 is dissimilar from the presenter in NP1. As a result, the face tracking method stored in the dataset and the video stream of the NP2 is undetermined. One of the immediate improvements to surmount this problem is to take into account several anchorpersons' faces corresponding to news program in the dataset

5.2. Experiments on TRECVID Dataset

The second experiment was carried out on TRECVID 2003 benchmark. This dataset, which is the only available benchmark and the most adopted by scholars, was selected to compare our results to those of previous works.

The TRECVID 2003 collection include over 2900 story boundaries [15]. TV news programs of this dataset are selected from various channels (CNN, ABC, etc.,).

In these experiments, we compared the results obtained by our approach of structuring the news program with other recent works of the state of the art [2,11, 19]. These works for the structuring of the news program are based on the exploitation of the anchorperson as anchor points of reference for the identification of topics. In addition, all these works used the TRECVID 2003 benchmark during the evaluation process.

To obtain an exact comparison with previous works, we made use of the same metrics as those defined in [5]. We therefore assess the performance of news segmentation of stand-alone mode as well as rewire modes by means of the precision, recall and F1 metrics. The proportional results (Figure 12) demonstrate the efficiency of the proposed scheme. The rates achieved by the stand-alone mode were the best. Dumont *et al*. [2] In the rewire mode, we performed an experiment based on the direct matching between descriptors of the background of the decor stored in the grammar and these extracted from current news program. Another experiment was carried out without considering the texture. So, although the extraction of anchor frame increased significantly the detection rate that confirms the importance of Clustering of histogram of orientation gradients faces.

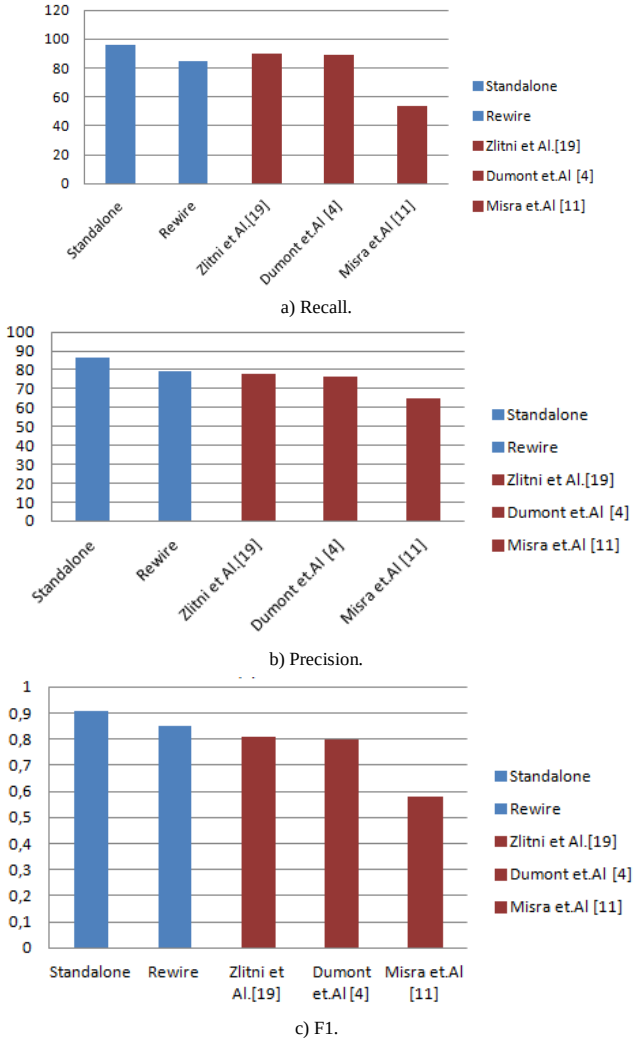

Figure 12. Comparative experiments results.

6. Conclusions and Future Work

In this work, we presented an approach for automatic segmentation of TV streams into topics. We described the necessary steps for this structuring. This contribution draws its originality from using both the contextual and operational features that regulate the organization of the contents of a news program. The modeling of these characteristics was realized by image processing techniques and statistical models. From this structuring, we also showed that we could extract the topics of a news program according to two modes: Stand-alone mode and Rewire mode. The proposed approach focuses on the analysis of television news only. In further research work, we will try to adapt this approach to deal with a continuous stream of programs. One of the interesting perspectives in this context is the identification of Talk Show in the topics of news program.

References

[1] Colace F., Foggia P., and Percannella G., "AProbabilistic Framework for TV-News Stories Detection and Classification," *in Proceedings of*

IEEE International Conference on Multimedia and Expo, Amsterdam, pp. 1350-1353, 2005.

[2] Dumont E. and Quénot G., "Automatic Story Segmentation for TV News Video Using Multiple Modalities," *International Journal of Digital Multimedia Broadcasting*, vol. 12, no. 1, pp. 1-11, 2012.

[3] Ejbali R., Zaied M., and Ben Amar C., "Face Recognition Based on Beta 2D Elastic Bunch Graph Matching," *in Proceedings of 13th International Conference on Hybrid Intelligent Systems (HIS*), Gammarth, pp. 88-92, 2013.

[4] Gao Y., Wang M., Zha Z., Tian Q., Dai Q., and Zhang N., "Lessismore: Efficient 3D Object Retrieval with Query View Selection," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1007-1018, 2011.

[5] Gnouma M., Ejbali R., and Zaied M., "Detection of Abnormal Movements of A Crowd in A Video Scene," *International Journal of Computer Theory and Engineering*, vol. 8, no. 5, pp. 398-402, 2016.

[6] Goyal A., Punitha P., Hopfgartner F., and Jose J., "Split and Merge Based Story Segmentation In News Videos," *in Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, Toulouse, pp. 766-770, 2009.

[7] Hmayda M., Ejbali R., and Zaied M., "Program Classification in A Stream TV Using Deep Learning," *in Proceedings of the 18th International Conference on Parallel and Distributed Computing, Applications and Technologies*, Taipei, pp. 123-126, 2017.

[8] Hu W., Xie N., Li l., Zeng X., and Maybank S., "A Survey on Visual Content-Based Video Indexing and Retrieval," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 41, pp. 797-819, 2011.

[9] Jacobs A., Miene A., Ioannidis J., and Herzog O., "Automatic Shot Boundary Detection Combining Color, Edge, and Motion Features of Adjacent Frames," *in Proceedings of the TRECVID Workshop Notebook Papers*, Gaithersburg, pp. 197-206, 2004.

[10] Khondaker A., Khandaker A., and Uddin J., "Computer Vision-based Early Fire Detection Using Enhanced Chromatic Segmentation and Optical Flow Analysis Technique," *The International Arab Journal of Information Technology*, vol. 17, no. 6, pp. 947- 953, 2020.

[11] Misra H., Hopfgartner F., Goyal A., Punitha P., and Jose J., "TV News Story Based Segmentation one Semantic Coherence and Content Similarity," *in Proceedings of the 16th International Conference on Advances in Multimedia Modeling*, Chongqing, pp. 347-357, 2010.

[12] Mohamed A., Issam A., Boussa M., and Abdellatif B., "Real-Time Detection of Vehicles Using the Haar-like Features and Artificial Neuron Networks," *Procedia Computer Science*, vol. 73, pp. 24-31, 2015.

[13] O'Hare N., Smeaton A., Czirjek C., O'Connor N., and Murphy N., "A Generic News Story Segmentation System and its Evaluation," *in Proceedings of IEEE International Conference Acoust Speech Signal Process*, Montreal, pp. 1028-1031, 2004.

[14] Poulisse G., Moens M., Dekens T., and Deschacht K., "News Story Segmentation in Multiple Modalities," *Multimedia Tools and Applications*, vol. 48, no.1, pp. 3-22, 2010.

[15] Smeaton A., Kraaij W., and Over P., "The TREC VIdeo Retrieval Evaluation (TRECVID): A Case Study and Status Report," *in Proceedings of 7th International Conference, Computer-Assisted Information Retrieval*, France, pp. pp. 26-28, 2004.

[16] Vijaya., Sharma S., and Batra N., "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering," *in Proceedings of International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, Faridabad, pp. 568-573, 2019.

[17] Viola P. and Jones M., "Rapid Object Detection Using A Boosted Cascade of Simple Features," *in Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 511-518, 2001.

[18] Yuan J., Wang H., Xiao L., Zheng W., Li J., Lin F., and Zhang B., "A Formal Study of Shot Boundary Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 2, pp. 168-186, 2007.

[19] Zlitni T., Bouaziz B., and Walid M., "Automatic Topics Segmentation for TV news Video Using Prior Knowledge," *Multimedia Tools and Applications*, vol. 75, no. 10, pp. 5645-5672, 2015.

Mounira Hmayda received the M.S Degree in multimedia computer in 2011 from the University of Gabes, TUNISIA, where she is pursuing the Ph.D. degree in computer science. His research interests focus on video and image processing and analysis, multimedia indexing, and content-based video segmentation and structuring.

Ridha Ejbali received the HDR, the Ph.D degree in Computer Engineering, Master degree and computer engineer degree from the National Engineering School of Sfax Tunisia (ENIS) respectively in 2012, 2006 and 2004. He joined the faculty of sciences of Gabes Tunisia (FSG) where he is an assistant in the Department computer sciences since 2012. Since now, he is assistant professor in faculty of sciences of Gabes Tunisia (FSG). His research area is now in pattern recognition and machine learning using Wavelets and Wavelet networks theories. He is IEEE senior Member.

Mourad Zaied Professor received the HDR, the Ph.D degrees in Computer Engineering and the Master of Science from the National Engineering School of Sfax respectively in 2013, 2008 and in 2003. He obtained the degree of Computer Engineer from the National Engineering School of Monastir in 1995. Since 1997 he served in several institutes and faculties in university of Gabes as teaching assistant. He joined in 2007 the National Engineering School of Gabes (ENIG) as where he is currently an associate professor in the Department of Electrical Engineering. He is a member of the REsearch Team on Intelligent Machines (RTIM) in the National Engineering School of Gabes (ENIG) since 2001.