GENETIC ALGORITHM FOR SUMMARIZING NEWS STORIES

Mehdi Ellouze, Hichem Karray
Research Group on Intelligent Machines, Sfax University, ENIS, Tunisia
Sfax Superior Institute of Technological Studies

Adel M.Alimi
Research Group on Intelligent Machines, Sfax University, ENIS, Tunisia

Keywords: News broadcast, summary, summarization, browsing, Genetic Algorithms.

Abstract: This paper presents a new approach summarizing broadcast news using Genetic Algorithms. We propose to segment the news programs into stories, and then summarize stories by selecting from every one of them frames considered important to obtain an informative pictorial abstract. The summaries can help viewers to estimate the importance of the news video. Indeed, by consulting stories summaries we can affirm if the news video contain desired topics.

1 INTRODUCTION

TV is one of the most important sources of information. The number of news broadcast channels is in perpetual increase. Besides, the techniques of digitizing have greatly progressed and the storage capacity of computers has become very important. For this reason, all people are digitizing enormous quantity of video sequences. However, this has caused a problem in manipulating this mass of information.

The importance of this problem has engendered a large number of works done in the field of video summarization and video browsing. In this paper, we will be interested to the problem of the summarization of the news broadcast.

The news program is a specific type of video. It's usually structured as a collection of reports and stories. A story as it was defined by U.S. National Institute of Standards and Technologies (NIST) is a segment of a news broadcast with a coherent news focus, which may include political issues, finance reporting, weather forecast, sports reporting, etc. (NIST, 2004). In our approach we profited from the several works proposed for news segmentation to detect and to extract news stories. Then we propose a genetic solution to summarize every story by keeping only important frames (Figure 1). The selection is done according to some criteria to keep only informative ones.

Many approaches have been proposed to summarize video. We can distinguish essentially two classes: Image Processing perspective (IP) (section 3.2) and Natural Language Processing perspective (NLP) (Qi *et al.*, 2000) (Maybury, Merlino, 1997). The first perspective is based on the extraction of key frames according to low level features. The major drawback of this perspective is the absence of high level features in the criterion of selection. Indeed, text transcriptions and faces play a key role in content analysis in some kinds of video, especially news broadcast. In the second perspective, researchers work on the extraction of text from video frames and the automatic speech recognition to formulate a textual summary. However, such methods are necessarily limited by the quality of the speech transcription itself and the efficiency of the proposed method to extract text from frames.

Nowadays, we are speaking about multi-modal summarization approaches pioneered by Informedia (Informedia Project, 2006). In such type of approaches, we combine low level features and high level features to accomplish video navigation and browsing systems.

As part of this new tendency, we propose in this paper a genetic solution for summarizing news

Ellouze M., Karray H. and M.Alimi A. (2007).
GENETIC ALGORITHM FOR SUMMARIZING NEWS STORIES.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IFP/IA*, pages 303-308

broadcast which uses both of low level features and high level features to generate pictorial summaries of news stories. As high level features, we choose to work with textual information. However, we have escaped the problem of text extraction from which NLP approaches are suffering.
The rest of paper is organized as follows. In section 2, we discuss works related to news segmentation. Section 3 describes our genetic algorithm proposed to summarize stories. Results of our approach are shown in section 4. We conclude with directions for future work.

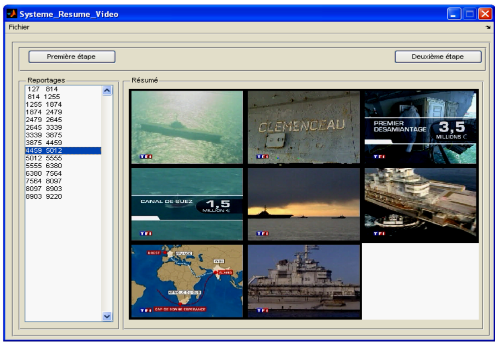

Figure 1: User interface of news segmentation and stories summarization. Stories are selected from the list on the left side. Summary of every story appears on the right side.

2 NEWS SEGMENTATION

Many works have been done in the field of extracting stories from news video. The idea is to detect anchor shots. The first anchor shot detector dates back to 1995. It was proposed in (Zhang *et al.*, 1995) suggesting to classify shots basing on the anchorperson shot model .As part of Informedia project (Informedia Project, 2006), authors in (Yang *et al.*, 2005) use high level information (speech, text transcript, and facial information) to classify persons appearing in the news program into three types: anchor, reporter, or person involving in a news event.

News story segmentation is also well studied in the TRECVID workshops (TRECVID 2004, 2003), in news story segmentation sessions. As part of TRECVID 2004, we can refer to the work proposed in (Hoashi *et al.*, 2004) in which authors proposed SVM-based story segmentation method using low-level audio-video features. In our work, we used the approach proposed in (Zhai *et al.*, 2005) in which the news program is segmented by detecting and classifying bodies to find group of anchor shots. It is based on the assumption that the anchor's wears are the same through out the entire program.

3 STORIES SUMMARIZATION

3.1 Problem Description

Summarizing a video consists in providing an other version which contains pertinent and important items needful for quick content browsing. In fact, our approach aims at accelerating the browsing operation by producing pictorial summaries helpful to judge if the news video is interesting or not. In the web context, it indicates to persons who are connected to online archives, if a given news video is interesting and if it may be downloaded or not.

3.2 Classical Solutions for Key-frames Extraction

Many works have been proposed in the field of video pictorial summary. As it is defined in (Ma, Zhang, 2005) a pictorial summary must respond to three criteria. First, the summary must be structured to give to the viewer a clear image of the entire video. Second, the summary must be well filtered. Finally, the summary must have a suitable visualization form. Authors in (Taniguchi *et al.*, 1997) have summarized video using a 2-D packing of "panoramas" which are large images formed by compositing video pans. In this work, key-frames were extracted from every shot and used for a 2-D representation of the video content. Because frame sizes were not adjusted for better packing, much white space can be seen in the summary results. Besides, no effective filtering mechanism was defined. Authors in (Uchihachi *et al.*, 1999) have proposed to summarize video by a set of key-frames with different sizes. The selection of key-frames is based on eliminating uninteresting and redundant shots. Selected key-frames are sized according to the importance of the shots from which they were extracted. In this pictorial summary the time order is not conserved due to the arrangement of pictures with different sizes.
Later, in (Ma, Zhang, 2005) authors have proposed a pictorial summary, called video snapshot. In this approach the summary is evaluated for 4 criteria. It must be visually pleasurable, representative, informative and distinctive. A weighting scheme is proposed to evaluate every summary. However this approach suggests a real filtering mechanism but it uses only low level features (color, saturation …).

Indeed, no high level objects (text, faces …) are used in this mechanism.

3.3 Genetic Solution

The major contribution of this paper is the integration of low and high level features in the selection process using genetic algorithms. Indeed, there are several advantages of genetic algorithms. First, they are able to support heterogeneous criteria in the evaluation. Second, genetic algorithms are naturally suited for doing incremental selection, which may be applied to streaming media as video.
Genetic algorithms are a part of evolutionary computing (Goldberg, 1989) which is a rapidly growing area of artificial intelligence. A genetic algorithm begins with a set of solutions (represented by chromosomes) called population. The best solutions from one population are taken and used to form a new population. This is motivated by a hope, that the new population will be better than the old one. In fact, the selection operator is intended to improve the average quality of the population by giving individuals of higher quality a higher probability to be copied into the next generation.

In this paper, we aim to summarize a given story using genetic algorithm. We suggest generating randomly a set of summaries (initial population). Then, we run the genetic algorithm (crossing, mutation, selection, crossing…) many times on this population with the hope to ameliorate the quality of summaries population. At every iteration, we evaluate the population through a function called fitness. Evaluating a given summary means evaluate the quality of selected shots. For this reason, we are based on three assumptions.

First, long shots are important in news broadcast. Generally the duration of a story is between 3 and 6 minutes, which is not an important duration. So, the producer of the news program must attribute to every shot the suitable duration. So, long shots are certainly important and contain important information. Secondly, shots containing text are also crucial because text is an informative object. It's often embedded in news video and it is a useful data for content description. For broadcast news video, text information may come in the format of caption text at the bottom of video frames. It is used to introduce the stories (War in Iraq, Darfour conflict) or to present a celebrity or an interviewed person (Kofi Anan, chicken trader, Microsoft CEO).Finally, to insure a maximum of color variability and a maximum of color coverage, the selected shots must be different in color space.

3.3.1 Summary Size

The summary size is computed through a summary rate which is a manually fixed parameter to indicate the number of selected shots. If we raise the summary rate, the number of selected shots will be greater and then the summary size will be larger and so the browsing speed will decrease. The number of selected shots is computed as follows:

$$Ns=N*R \tag{1}$$

For example, if the summary rate is 20% and the story is composed of 15 shots then the number of selected shots will be equal to 3.

3.3.2 Binary Encoding

We have chosen to encode our chromosomes (summaries) with binary encoding because of its popularity and its relative simplicity. In binary encoding, every chromosome is a string of bits (0, 1). In our genetic solution, the bits of a given chromosome are the shots of the story. We use 1's to denote selected shots (Figure 2).

3.3.3 Evaluation Function

Let PS be a pictorial summary composed of m selected shots. PS= $\{S_i, 1 \leq i \leq m\}$.

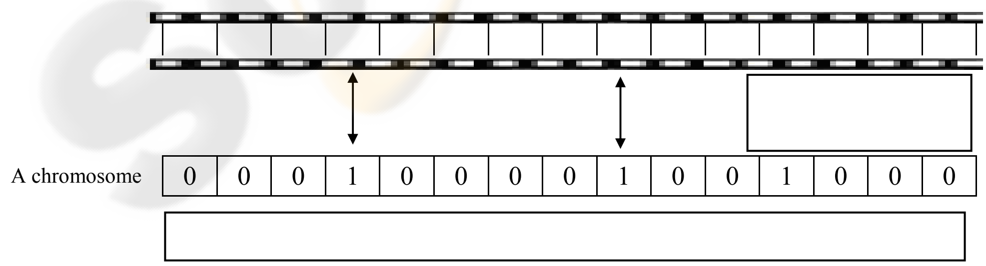

Figure 2: The encoding of the genetic algorithm. The shots which are present in the summary are encoded by 1. The remaining shots are encoded by 0.

We evaluate the chromosome C representing this summary as follows:

$$f(C) = \frac{AvgHist(C) + AvgText(C) + AvgDuration(C)}{3} \quad (2)$$

AvgHist(C) is the average color distance between slected shots. AvgHist is computed as follows:

$$AvgHist(C) = (\sum_{i=1}^{m}\sum_{i=j+1}^{m} Hist(S_i, S_j)) / (m(m-1)/2) \quad (3)$$

The distance between two shots A and B is defined as the complementary of the histograms intersection of A and B:

$$Hist(A,B) = 1 - (\sum_{k=1}^{256} \min(H_A(k), H_B(k)) / \sum_{k=1}^{256} H_A(k)) \quad (4)$$

AvgText and AvgDuration are respectively normalized average of text score and normalized average of duration of selected shots.

$$AvgText(C) = (\sum_{i=1}^{m} Text(S_i)) / (m * T_{\max}) \quad (5)$$

$$AvgDuration(C) = (\sum_{i=1}^{m} Duration(S_i)) / (m * D_{\max}) \quad (6)$$

T_{max}= max {Text (Si), $S_i \in$ PS}.
D_{max}=max {Duration (Si), $S_i \in$ PS}.

More the shots are different in the color space more AvgHist is increasing. In fact, AvgHist is related to Hist which is increasing when the compared shots are more different. It's obvious that more selected shots contain text (respectively have long durations) more AvgText (respectively AvgDuration) is increasing.
The proposed Genetic Algorithm tries to find a compromise between the three heterogeneous criteria. That's why all the criteria are normalized and unpondered in the fitness function.

3.3.4 Computation of Parameters

Our genetic solution is based on three parameters: text, duration and color. To quantify these parameters for every shot we define 3 measures. The first is the text score of the shot, the second is its duration and the third is its color histogram. The duration of a shot can be easily computed. Color and Text parameters are computed for the middle frame of the shot. To compute the color parameter of a given shot we compute the histogram of its middle frame.

The computation of the text parameter is more complicated. In our approach we work only with artificial text, it appears generally as a transcription at the foot of the frame. To compute the text parameter, we select the middle frame of the shot, then we divide it into 3 equal parts and like in (Chen, Zhang, 2001) we apply a horizontal and a vertical Sobel filter on the third part of the frame to obtain two edge maps of the text. We compute the number of edge pixels (white pixels) and we divide it by the total number of pixels of the third part. The obtained value is the text parameter (Figure 3). We have avoided in the computing of text parameters the classic problems of text extraction which may reduce the efficiency of our approach. We have been based on the fact that frames containing text captions are certainly containing more edge pixels in their third part than the others. So, their scores will be greater.

(a)Text score=0.0936 (b)Text score=0.04

Figure 3: Shot text scoring. The text score of a frame containing text (a) is greater than a frame not containing text (b).

3.3.5 Genetic Operations

The genetic mechanism works by randomly selecting pairs of individual chromosomes to reproduce for the next generation.

a) The crossover operation
The crossover consists in exchanging a part of the genetic material of the two parents to construct two new chromosomes. This technique is used to explore the space of solutions with the proposition of new chromosomes to ameliorate the fitness function. In our genetic algorithm the crossover operation is classic but not completely random. In fact, like their parents, the produced children must respect the summary rate. For this reason, the crossing site must be carefully chosen (Figure 4).

b) The Mutation operation After a crossover is performed, mutation takes place. Mutation is intended to prevent the falling of all solutions in the population into a local optimum. In our genetic solution, mutation must also respect the summary rate. For this reason, mutation operation must affect two genes of the chromosome. Besides, these genes must be different ('0' and '1') (Figure 5).

4 EXPERIMENTS

4.1 Dataset and Test Bed

To validate our approach, we choose to work with French channels "*TF1*" and "*France2*" as general channels and the English Channel "BBC" as news broadcast channel. We have recorded 20 hours of news from "*TF1*" and "*France2*" (night and midday news) and 10 hours from "BBC". Our system was implemented with matlab and tested on a PC with 2.66 GHZ and 1GB RAM. An example of a story summarization is illustrated by Figure 6

4.2 Users Evaluation

Our system aims at producing summaries helping users to judge if the news sequence is interesting or not. In this section, we will try to prove if we are successful in achieving our goal. We have invited 10 test users to search through our dataset 4 topics speaking about: "*Avian flu*", "*Aids*", "*Iraqi war*" and "*Israeli-Palestinian conflict*".

Table 1 shows the results of this test. We notice that the best rates of precision and recall correspond to topics like "*Iraqi war*" or "*Israeli-Palestinian conflict*" and in general topics related to wars and conflicts. In fact these topics are related to special key objects as tanks, victims, damage, bombing, soldier, etc. The majority of stories speaking about these topics are showing at least one of these objects. In this type of topic, textual information comes to enhance

the comprehension of the summary by giving indications about places (Iraq, Palestine, Darfour, Afghanistan) or persons and leaderships (Olmert, Arafat, Abbas, Nouri Meliki, Ben Laden).

However, topics like "*Avian flu*" or "*Aids*" are not related to special objects. Indeed, we may have a story speaking about aids which is not showing aids patients because it may speak only about finding funds to fight against this epidemic. In that case, textual information is important for the summary viewer to understand the story context. Names of the interviewed persons may play a key role for the comprehension of the context.

5 CONCLUSION AND FUTURE WORK

In this paper, we have presented a novel multimodal approach to generating pictorial summaries of news stories. Generated summaries help viewers to browse rapidly video archives by showing only informative frames. We have integrated easily low level features and high level features through a genetic algorithm. One of the advantages of the use of genetic algorithms is their extensibility. Indeed, we plan in the immediate future to add other low level features (motion, audio features) and other high level features (faces) to improve the quality of the summary.

The encouraging results obtained for news broadcast corpus, motivate us to think of extending the use of genetic algorithms to summarize other corpus. In fact, we have begun to investigate adapting the architecture of our genetic algorithm (encoding and fitness) to films and documentary videos.

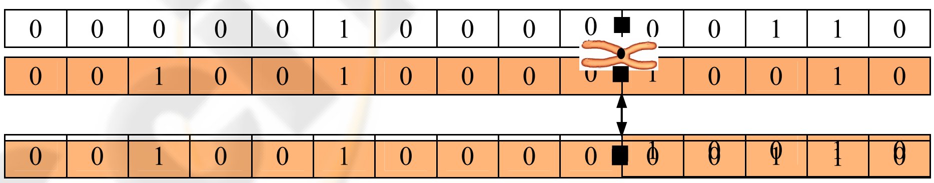

Figure 4: The crossover operation. We must carefully select the crossover site to respect the summary rate.

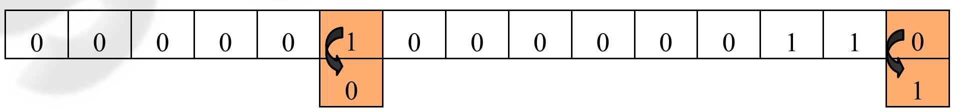

Figure 5: The mutation operation. To respect the summary rate, the mutation must affect two different bits.

Table 1: Evaluation results for four topics on our dataset.

Topics	Manual relevant stories	Mean of true detected stories	Mean of false detected stories	Recall (%)	Precision (%)
Avian flu	8	6.7	1.4	83.75%	82.71%
Aids	5	3.6	1.3	72%	73.46%
Iraqi war	17	14.7	1.6	86.47%	90.18%
Israeli-Palestinian conflict	13	12.1	1.6	93.07%	88.32%

Figure 6: An example of a news story summarized by our genetic algorithm.

REFERENCES

Chen X., Zhang H., 2001 Text area Detection from video frames, *In Proceedings of the IEEE Pacific-Rim Conference on Multimedia* (222-228)

Goldberg D.E., 1989 Genetic Algorithms in Search, Optimization, and Machine Learning

Hoashi K., Sugano M., Naito M., Matsumoto K., Sugaya F., Nakajima Y., 2004, TRECVID Story Segmentation based on Content-Independent Audio-Video Features, *2004 TRECVID Workshop*

Informedia Project, 2006 *http://www.informedia.cs.cmu.edu/dli2/index.html*

Ma Y.F., Zhang H.J, 2005 Video Snapshot: A Bird View of Video Sequence, In Proceedings of *International Multimedia Modeling Conference* (94- 101).

Maybury M.T., Merlino A.E., 1997 Multimedia summaries of broadcast news In Proceedings of *Intelligent Information Systems Conference* (442 – 449).

NIST, 2004 http://www-nlpir.nist.gov/projects/tv2004/tv2004.html#2.2

Qi W., Gu L., Jiang H, Rong X.C., Zhang H.J., 2000 Integrating Visual, Audio and Text Analysis For NewsVideo, *IEEE International Conference on Image Processing Conference*.

Taniguchi, Y., Akutsu, A., Tonomura, Y., 1997 PanoramaExcerpts: Extracting and Packing Panoramas for Video Browsing, In Proceedings of *ACM Multimedia* (427-436)

TRECVID, 2003, 2004, TREC Video Retrieval Evaluation: http://www-nlpir.*nist.gov/projects/trecvid/*.

Uchihashi S., Foot.J, Girgensohn.A, Boreczky.J, 1999 Video Manga: Generating Semantically Meaningful Video Summaries, In Proceedings of *ACM Multimedia* (383-392)

Yang J., Hauptmann A. G., 2005 Multi-modal Analysis for Person Type Classification in News Video, In Proceedings of *Symposium on Electronic Imaging* (165-172).

Zhai Y., Yilmaz A., Shah M. , 2005 Story Segmentation in News Videos Using Visual and Text Cues, In Proceedings of the *International Conference on Image and Video Retrieval* (92-102).

Zhang, H.J., Low, C. Y. S., Smoliar, W., Zhong, D., 1995 Video parsing, Retrieval and Browsing: An integrated and Content-Based Solution, *I*n Proceedings of *ACM Multimedia* (45-54).