

Discovering Knowledge for Better Video Indexing based on Colors

Marcin Detyniecki
CNRS, Pole IA, LIP6
Université P. et M. Curie
8, rue du Capitaine Scott
75015 Paris - France
Marcin.Detyniecki@lip6.fr

Christophe Marsala
Pole IA, LIP6
Université P. et M. Curie
8, rue du Capitaine Scott
75015 Paris - France
Christophe.Marsala@lip6.fr

Abstract — In this paper, we present the discovery of rules for different challenges encountered in video indexing. These rules should be considered as knowledge that can be used as a guideline for the development of better indexing tools. We use a fuzzy decision tree to extract the rules based on color proportions of key-frames extracted from one single video-news. Experimental results and comparisons with other data mining tools are presented.

Index Terms — Data Mining, Multimedia Indexing, Fuzzy Decision Trees.

I. INTRODUCTION

On the one hand, the growth of video data has caused a need to analyze and exploit them. Hints of this increase are the availability of video news on the web or the appearance of hard drive video recorders in the market. In such a context it appears that the user expects to interact with the video object, for instance by finding a specific shot or video sequence. In order to provide access to the video, indexing is needed. Unfortunately, today's indexing is generally done manually. The growth of video data and the requirement of new applications for finer grain access push for an automation of the indexing process.

On the other hand, fuzzy data mining introduces new methodologies to extract and discover fuzzy knowledge from either classical or fuzzy data repositories. The advantage of using fuzzy algorithms is that they enable us not only to offer the discovered knowledge in a more comprehensive way, but also to handle numerical and/or fuzzy data.

Since data mining is known to improve the knowledge of the domain from where the data comes from, it appears natural and promising to link fuzzy data mining with multimedia data that leads to fuzzy multimedia mining. In fact, the extracted knowledge can be used to improve the indexing process (e.g. helping the development of indexing tools). The intuitive idea behind this association is to let the

computer find (mine) the knowledge based on what it (the computer) is able to distinguish and not based on our (human) perception.

Besides the problem of the quantity, dealing with multimedia introduces a new difficulty related to the polymorphism of the data [2]. In fact the data that can be extracted from a video are texts, images, sounds, temporal data, metadata, etc. Thus, it is important to work with a flexible and automated data-mining tool, which will induce knowledge from all kinds of data simultaneously. A particular instance of such tools is the fuzzy decision tree algorithm [1].

The structure of this paper is the following. In Section II, we present the problem of discovering indexing rules. Then before exposing the details of each of the experiments we introduce some references of our data-mining software that constructs the fuzzy decision trees, in Section III. In the following section, we present the discovery of rules for three different problems directly related to the structuring of video news. In the last section, we discuss the results and we explain their application.

II. DISCOVERING INDEXING RULES

In this paper we consider three different mining problems. All related to the extraction of knowledge associated to the identification of the general structure of the video news (macro-segmentation). The idea is to be able to separate the different thematic parts (structure) of the video news automatically, offering the possibility to navigate or automatically personalize the news (for instance under time constraints). One way to tackle the problem is to base the work on the script of the news (text), which makes you dependant on a transcript service (manual or automatic). Here, we try to provide some help on the visual aspects of the problem. We focus on the discovery of knowledge, which should help improving the visual indexing aspects:

- (A). The first problem we focus on is the detection of the appearance of inlays, which usually appear in crucial moments of the news. We try to discover the relevant features that tackle this problem.
- (B). The structure of video is strongly based on the detection of shots. In this second experiment we are interested in finding hints on how to improve the shot detection algorithms, based on the observed errors.
- (C). Finally, the discovery of the discriminant features that should be selected for the detection of a host, a diagram or a correspondent. Note that all three are very important for the general structure of the news.

Before entering into the details of each of these points, we briefly present our mining tool in the next section.

III. FUZZY DECISION TREES

Knowledge Discovery from Data (KDD) was introduced by Fayyad, Piatetsky-Shapiro and Smyth at the beginning of the nineties [4]. Due to the polymorphism of multimedia data, Multimedia Data Mining (MDM) was recently proposed as a new topic of research [5].

In fact in a multimedia framework, versatile data-mining tools are necessary. One particular tool is the fuzzy decision tree learning algorithm, which provides rules that summarize and explain the data. We use the Salammbô software, which is able to handle typical numerical input (non fuzzy), and it constructs a fuzzy decision tree without human intervention [6].

Another advantage of using fuzzy decision trees resides in the fact that they represent a natural and understandable knowledge: a fuzzy decision tree is equivalent to a set of fuzzy "if...then" rules.

In order to construct these rules, you need a set of examples. The fuzzy decision trees will then summarize the information based on these examples. Therefore the examples have to be representative of the problem and in a sufficient number. We used traditional machine learning techniques to make sure that we learn rules general and correct (as for instance learning and testing on different example sets).

The trees are built by means of a fuzzy entropy measure, which translates a certain order. In other words we can automatically discover which features are the most important (discriminant) and what are the values to be considered for these features.

IV. MINING COLORS FROM KEY-FRAMES

As we said above we focus here on the visual aspects. We could use complicated image signature to extract the knowledge, but since this is one of the first works of this type, we prefer to focus on a well-known feature: the colors. This restriction allows us to simply interpret and understand

the results. But there are no fundamental difficulties in directly applying our procedure to more complex signatures.

The general scheme for all the experiments is presented in the following. We start from a set of key-frames extracted (per shot) from a single video news [7]. In a second step, the set of colors of each key-frame is vectorized and "projected" to a given reference-palette (for instance a palette of 64 or 256 colors obtained by equally sampling the RGB space). We obtain like this a common basis to compare the key-frames. Then for each key-frame a histogram of frequencies (of the colors) is computed. This provides us with a vector in a reference space (defined by the colors of the reference palette) for each key-frame. See Figure 1.

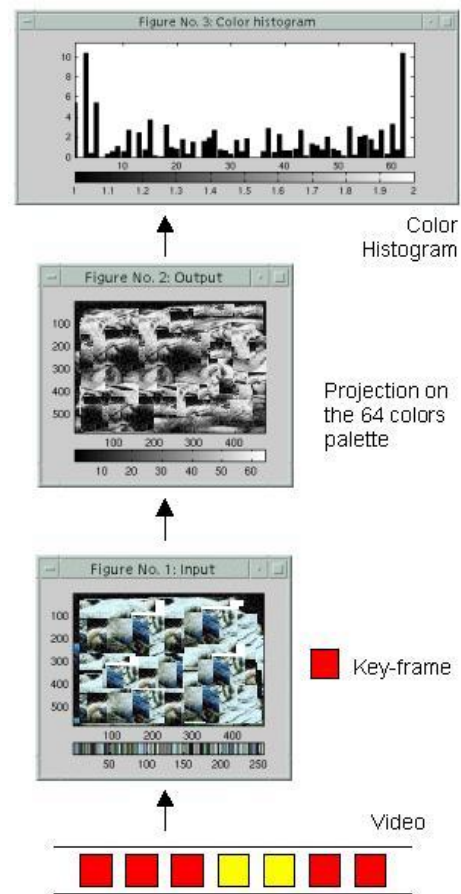


Fig. 1. Color histogram extraction. (not actual key-frames) *

From each key-frame we obtain a vector, which is then considered as a training example. Thus, a training set can be composed from an original set of key-frames extracted for one video. Finally we build the fuzzy decision tree using the Salammbô software.

* The original images could not be presented here, because of copyright issues.

Based on this general scheme, three different experiments are conducted: *A*, *B* and *C*.

A. Discovering the presence of inlays

Inlays that appear on the TV screen are very often hints for the structure of the video news. They also appear either when a new person is presented or when a report ends. They usually consist in a square or a rectangle that frames some text (e.g. name of the journalist, name of the place, etc).

We have conducted several mining experiments in order to determine if colors are discriminant for the detection of inlays. A training set was composed with 176 vectorized key-frames, to each vector was assigned one class (type) of the key-frame: either *with* or *without* inlays.

A first experiment was conducted with the whole training set and based on a reference palette of 64 colors. The resulting fuzzy decision tree is not very deep and has a root node on which the presence of the white color is requested (white is the major background color of inlay key-frames in the training news report). We notice that the accuracy, recall and precision of this "root-rule" are extremely high (for details refer to [3]). This result points out that only a few number of colors is needed to discriminate the presence of inlay key-frames. And more generally, we confirm the empirical observation that the use of colors is a suitable feature for discriminating inlays.

This rules confirms the intuition. Nevertheless if we look closely at the fuzzy sets built by the system, we notice that the proportion of the main color of the inlay has to be inside a fuzzy range. In other words the system not only tells us what the rule is, but also that a specific fuzzy range has to be respected. In our case the percentage of white has to be *less than 12%* (fuzzy membership, see Figure 2). Another interesting observation is that other colors (than white) are used to determine the presence of inlays. By studying this carefully we found out that this is due to a bad projection (or detection) of the colors on the reference palette. In other

usually wrong projected colors. This is a typical example where a data mining (learning) system, will construct rules based on what it sees and not what we (human beings) perceive. And this is one of the reasons why we think that knowledge discovery may be very useful in multimedia indexing.

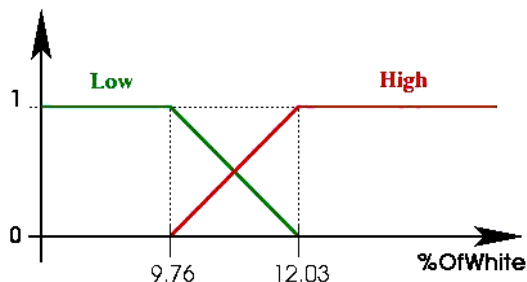


Fig. 2. Membership function describing low and large percentage of white color

A second experiment was conducted in order to compare the Salammbô algorithm with regard to other learning algorithms (see Table 1). For the other algorithms we used Weka software [8]. Notice that in this case recall and precision of the model are as important as the accuracy of the model. In fact, it is important to perfectly recognize at least one kind of key-frames (with or without inlays in this case).

We can observe that the fuzzy decision tree method is not only among the methods with the higher accuracy, but also, presents high recall and precision rates for inlays recognition. Moreover, it appears that the construction of fuzzy decision trees by Salammbô is also among the lowest time consuming methods, an important property for multimedia applications, where for instance a stream of frames must be handled in a short time. In addition to the understandability of the fuzzy decision tree model,

TABLE I. RESULTS FOR INLAYS RECOGNITION (64-COLORS PALETTE)

Algorithm	Accur. (%)	With inlays		Without inlays		Bld. Time (s)
		Recall	Precision	Recall	Precision	
Salammbô (FDT)	81.3	0.88	0.78	0.75	0.86	1
Naive Bayes	54.6	0.39	0.57	0.71	0.53	0.4
Voted Perceptron	71.6	0.71	0.72	0.73	0.71	0.6
Weka J48 (C4.5)	78.4	0.75	0.81	0.82	0.77	1
Decision table	82.9	0.87	0.81	0.8	0.85	3.3
Neural Network	80.1	0.84	0.78	0.76	0.83	322

words using the fuzzy decision tree we discover potential problems and the system provides a solution by using the

Salammbô presents better accuracy and quickness ratio than any other of the tested methods.

B. Discovering errors in the shot detection

Shot-detection algorithms have been proposed, in order to reduce the number of key-frames and to structure the video. Unfortunately, they produce a lot of false detections. Already by looking at the key-frames extracted from the video, we noticed that a lot of key-frames were very similar, due to the errors of the shot detection tool.

Thus, an experiment was conducted to discover any hints for improving the shot-detection based on examples of successive key-frames. The training set was constructed as follows: A first group of training examples was composed by two successive key-frames from the same shot (class "same shot"). Then each key-frame was vectorized in a 64 colors palette (as in the previous section). Finally the training vector was built by merging the two vectors. We obtain a single training example with 128 features. Notice that there is no information about the relationship between colors (for instance color i and color $i+64$). A second group (of counter examples) was composed by two successive key-frames, but this time from different shots (class "different shots").

At this point, we expect to find rules translating the similarity. But the built fuzzy decision tree points out that the most discriminant approach is to detect the increase of the proportion of a particular color: white (the dominant color of inlays). For this experiment, the mean accuracy with cross-validation of fuzzy decision trees was 78.26%.

This rule suggests that the most common shot detection error is due to the appearance of inlays. It also highlights, by looking at the other rules, that after a classical shot detection (based on similarities), the accuracy of the shot detection can be enhanced using the color dissimilarity between successive key-frames.

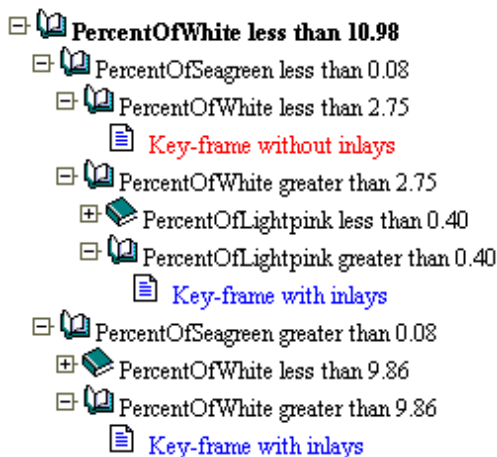


Fig. 3. Example of three rules extracted by Salammbô.

C. Discovering host, diagram, correspondent

An important hint about the structure of the news is to detect the appearance of the host, of an explanatory diagram, or of a journalist (not in the studios). Thus, an experiment was conducted in order to discover if the presence or non-presence of the host speaker in a key-frame can be discriminated just based on colors.

As in previous experiments, a training set was constructed. A first group of training examples was composed by key-frames where the host appears. A second group was composed of key-frames without host. Each key-frame is vectorized in a 256 colors palette.

We extracted from these data that the most effective fuzzy rule to recognize if we are in presence of the host (or not) is based on a color of the host's background (presently, one specific blue color). By looking at the host key-frames for this specific color, we noticed that it corresponds to the colors of the background. This rule points out the following knowledge: the best way to know if we are in presence of the host is to look if the scene takes place in the channel studio (by looking at the background). The accuracy of the fuzzy decision tree here is 88%.

V. DISCUSSION AND APPLICATION

With the previous experiments we show how a data mining tool (fuzzy decision tree) can be used to extract knowledge for better indexing in view of a macro-segmentation. We extracted a set of rules for each of the three different problems. These rules should be considered as guidelines (knowledge) for improving the indexing based on colors. It is important to note that these rules are not indexing rules, which means that they are intended to be applicable directly to a new video. There are different reasons for this. For instance, it is obvious that in order to obtain a general rule (i.e. for any video), there is a need of a representative database of this general case (i.e. all type of video news). This brings to light the importance of the choice of the examples.

The extracted knowledge is in the form of three unexpected seminal rules. Note that we did not introduce any prior knowledge, which shows the potential of this approach. In the following we summarize the extracted knowledge:

- (A). In order to recognize inlays, the system suggests detecting a large proportion of a single color, putting forward that all inlays are relatively large surfaces of a set of similar colors (in our experiments large means more than 3% and less than 12% of the screen).
- (B). In order to improve the shot detection (which is naturally based on similarity) it is recommended to look at the differences between key-frames. More precisely, the extracted rules point out that the difference is the presence of a large proportion of a

single color in the second frame, suggesting that the cause of the errors is usually the appearance of inlays.

- (C). In order to detect the presence of the host, we should focus on the background in order to discern if the scene was taken in the studios. So, this rule suggests not looking at the host, in order to detect his presence.

VI. CONCLUSION

In this paper, we present an example of multimedia knowledge discovery applied to the video-news structuring. We used fuzzy decision trees that extract simple and understandable rules.

We focus on the mining of the color feature of per-shot-extracted key-frames. The extracted rules are intended to provide hints for better video indexing in a structuring perspective. In fact, these rules deal with the appearance of important information on the screen (inlays), the presence of the host, and correcting possible mistakes in the shot detection.

This is a first step in the multimedia mining of video news. The next step could be to continue the mining of visual contents, based on more complex signatures and features as for instance the texture and also the mining of structural content.

Another perspective is to consider simultaneously other medias (sound, text, etc.). We anticipate that in this case we will be able to take advantage of their interaction. This can be particularly interesting, since cooperation of different medias is usually very difficult to take into account. In fact the per-media approaches are extremely specific.

REFERENCES

- [1] C. Marsala and B. Bouchon-Meunier, An Adaptable System to Construct Fuzzy Decision Trees, *Proceedings of NAFIPS'99* (New-York, June 1999), pp. 223-227.
- [2] M. Detyniecki and C. Marsala, Fuzzy inductive learning for multimedia mining, *Proceedings of EUSFLAT'01* (Leicester, UK, Sept. 2001), pp. 390-393.
- [3] M. Detyniecki and C. Marsala, Fuzzy Multimedia Mining Applied to Video News, *Proceedings of the IPMU'02 Conference* (Annecy, France, July 2002), pp. 1001-1008.
- [4] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From Data Mining to Knowledge Discovery, *Databases in AI Magazine*, 17:3, pp. 37-54, 1996.
- [5] First Workshop of Multimedia Data Mining. http://www.cs.ualberta.ca/~zaiane/mdm_kdd2000/
- [6] B. Bouchon-Meunier, C. Marsala and M. Ramdani, Learning from Imperfect Data, *Fuzzy Information Engineering: a Guided Tour of Applications*, D. Dubois, H. Prade and R. R. Yager eds, chapter 8, pp. 139-148, 1997.
- [7] R. Ruiloba and P. Joly, Framework for evaluation of video-to-shots segmentation algorithms, *Video Data, special issue of Networking and Information Systems*, Vol. 3, pp. 46-57, 2000.
- [8] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 1999.