M. Detyniecki, C. Seyrat, and R. R. Yager, "Interacting with Web Video Objetcts," *Proceedings of the International Conference of the North American Fuzzy Information Society - NAFIPS'99,* New York, USA, pp. 914-917, June, 1999.

Interacting with Web Video Objects

Marcin Detyniecki

Fuzzy Logic Team
LIP6 - CNRS -University of Paris VI
4, place Jussieu
75230 Paris Cedex 05, France
Marcin.Detyniecki@lip6.fr

Claude Seyrat

Indexation Multimedia Team
LIP6 - CNRS - University of Paris VI
4, place Jussieu
75230 Paris Cedex 05, France
Claude.Seyrat@lip6.fr

Ronald R. Yager

Machine Intelligence Institute
Iona College
New Rochelle, New York 10801, USA
Yager@panix.com

Abstract: We use ideas in the spirit of *computing with words* in order to answer time related queries about a single video. We introduce a time-related dictionary and some fuzzy logic tools in order to be able to construct the fuzzy queries. The system is based on fuzzy annotation. The model presented here was implemented on a Java based Video Search Engine.

Introduction

Video as format of computer related material is becoming more and more common [1,2]. Each day new multimedia information systems appear in the market containing more and more video. Also the format of information on the Internet is clearly evolving to a video form. Initially we saw the embedding of images on text pages, now we see simple animation on almost every web page. We also know that the amount of information stored in computers is growing. So, the question that naturally arises is: "How to get the information you want?" We propose here a new tool for retrieving information inside a video.

In this paper, building on the work of Yager [10], we focus on how to construct and answer queries on a single annotated video. The annotation [10] may be in a database with other information. We assume that the annotations are precisely time-indexed, but their attached information may be uncertain. In other words, we know precisely in which minute of the video something happens, but we are not completely sure about everything associated with the event. We note that this information may come from automatic or manual indexing. In order to handle this inevitable uncertainty we propose to use fuzzy methods.

We introduce, following the spirit of Zadeh's idea [3] of "Computing with words", a dictionary with the basic concepts and methods for constructing new ones. Computing with words allows us to have a human friendly interface. With this vocabulary and the logic tools introduced, the user will be able to realize human type queries. We focused on the fuzzy time related queries [4,5,10].

We remark that our dictionary contains notion such as time positioning, time descriptors and time relationships. We show how to combine them in a uniform way in order to construct the queries.
Let us begin, by explaining the most common way, information is attached to a video: annotations. And what we exactly understand by a fuzzy annotation.

Fuzzy annotations

The present works on query-systems for video are based on the use of annotations [6-10]. These annotations can be considered as information contained in a database associated to the video and indexed by the time. These annotations can be extracted from the video manually or automatically. With the automatic way it is clear that a lot of uncertainty arises. Just to give an example, the automatic face recognition is not, at this time, a hundred percent process.
But even the manual indexing, which we can consider as the most reliable way of obtaining the annotations, is not without uncertainty, not because of human errors, but because of the complexity of the world. For example when annotating night and day scenes, we can have a smooth passage from the day into the night. So we propose here to use fuzzy annotation to enrich the descriptions.
We call fuzzy annotation a classical annotation accompanied by a degree of certainty of the information (and not of the time indexing this annotation). This degree is usually a value between 0 and 1 (zero for completely uncertain and one for completely certain). So, for example an annotation can be: "At minute 6 the actor on the scene is Robert with a degree of certainty of 0.75". Which means that we are not totally sure if it is Robert, but rather yes. We notice here, that we assume that the indexing time (6 minutes) is certain.
We can now represent this information on a graph, where the *x*-axis is the indexing time of the film and the *y*-axis is the degree of certainty. Note that the actor appears for a period of time so that we have a curve and not a point.

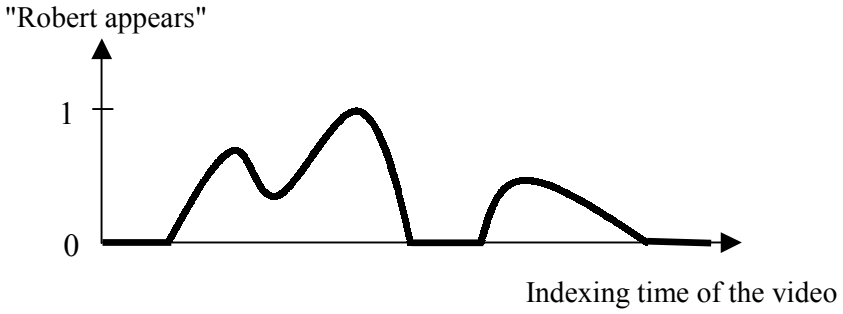

Figure 1. Fuzzy Annotation: "Robert appears".

In an analog way we can represent on the indexing time axis annotation like: Between the minute 9 and 13 we have a dialog. And where the beginning and the end of the dialog are not precise.

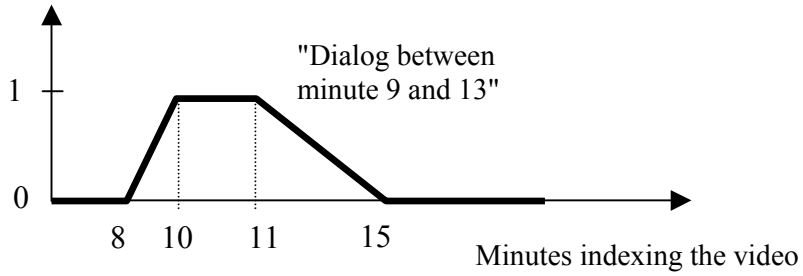

Figure 2. Fuzzy Annotation of a dialog

We remark that fuzzy annotations are a generalization of classical ones. We can obtain classical annotation by using only the degrees 0 and 1 of certainty. Now, that we have these annotations (fuzzy or not), we want to use them in order to extract information. So in the following we propose a fuzzy time related query system. For that we introduce at first what we call the fuzzy vocabulary.

Fuzzy time vocabulary

In the spirit of Zadehís idea of computing with words [3] and Yagerís work in [10], we propose to construct a fuzzy time related dictionary. This thesaurus will allow us to handle imprecise time querying. Using this we will be able to use time positioning definitions such as *beginning*, *end* and *middle*, to use imprecise time durations such as *about five minutes*, *long* and *short time* and to use time relationship like *after*, *before* and *close*. We will also see how to modify and combine them.

Time positioning

One of the things, which we may want to express, is a fuzzy positioning (of an event) in the time. This concepts gives us an imprecise location in the time, to which we refer. Examples of this kind of concept are *beginning, middle* and *end*.

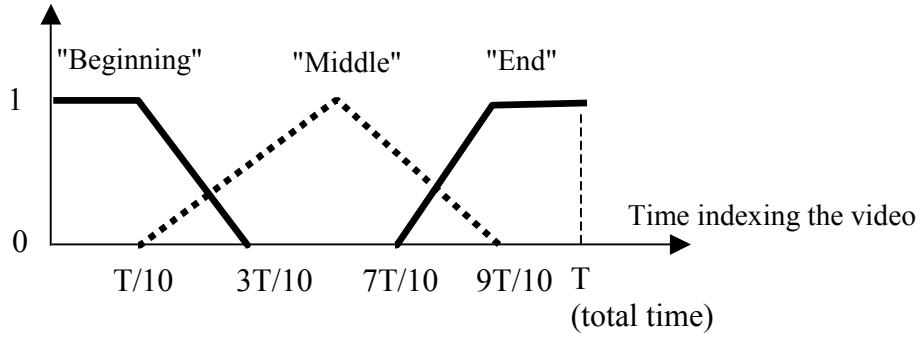

Figure 3. Beginning, Middle and End.

We remark here that the fuzzy *beginning*, *middle* and *end* are defined in relationship to the total reference time. We note that these notions are personal and so the user should personalize his translations, but always on a reference scale.

Duration descriptors

We may also want to describe the duration of some event. Here notions like long time and short time appear. Here once more, even if it is not explicit, we refer to a reference. It is not the same a *long time* in a 5 minutes video and a long time in a two hours film. We can then define this notion in a time scale having the total reference time.

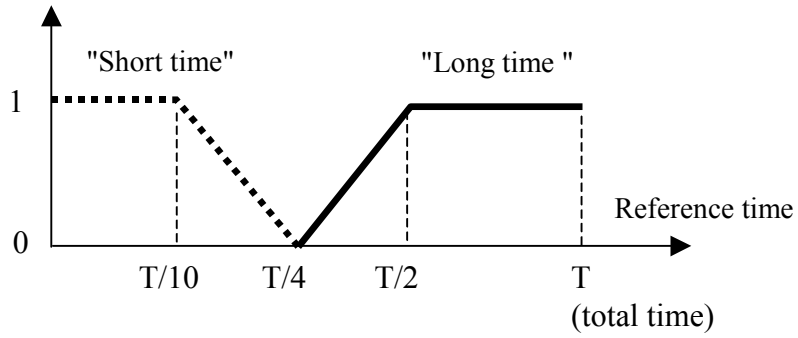

Figure 4. Long and Short time

We can also define on the time reference scale the notion "*about X minutes*". Here, once again we think that this notion is relative. We believe that the uncertainty grows with the value *X*. So, for example if we consider that the precision on "about five minutes" is a single minute, then the precision on "about 60 minutes" is at least 5 minutes (certainly more that 1 minute). We propose to translate "*about X minutes*" into "*between* (*X* minutes less 20 %) *and* (*X* minutes plus 20 %)" with the certainty of 1 for *X* minutes and a linear decreasing certainty between the extremes and the *X* minutes (see figure).

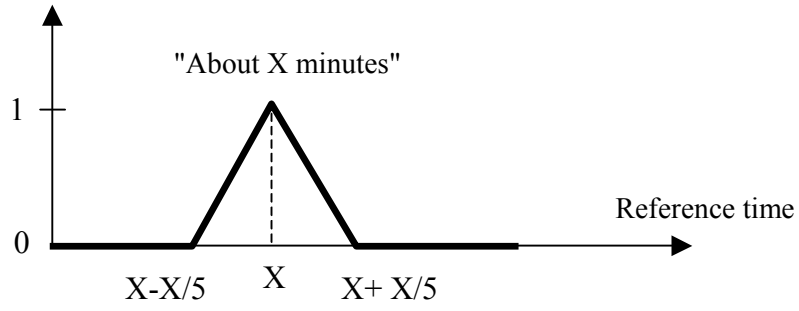

Figure 5. About *X* minutes

Time relationship

We may also want to define relationships between time events, like for example *after* and *before*. In this spirit Yager introduced in [10] a general framework: the relative temporal relationship (RTR). On this framework we have for example that the definition of "*after*" will be: If X-Y<0

then the degree of satisfaction of the concept "*after*" is 0 and if X-Y≥0 the degree will be 1. In a symmetric way we can define the notion of "*before*". We can also use the time descriptors introduced before to generate notions as for example: "*About 10 minutes after*" or "*About 10 minutes before*".

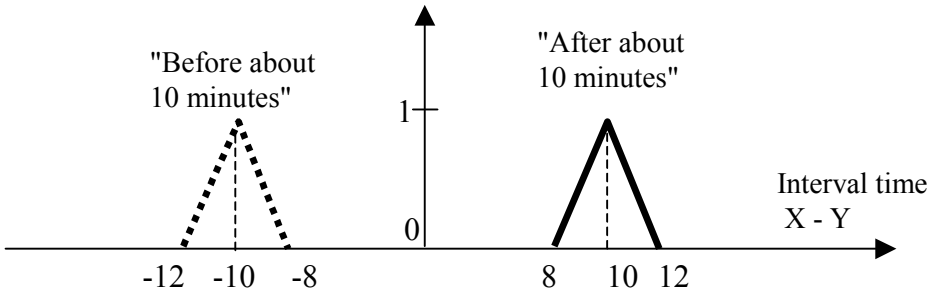

Figure 6. About 10 minutes after (or before)

Now, that we have defined the basic relationships we may want to use them to indicate a particular moment in the video. So we may say: "*About 10 minutes after the crash*".
Let *Crash*(*y*) be the membership degree of the event "*the crash*" at the time *y*. Then the membership function of "*About 10 minutes after the crash*" indexed by the time *x* will be obtained by the formula:

$$About\_10m\_after\_Crush(x) = \max_y\left(after\_about\_10m(x-y) \wedge Cruch(y)\right) \qquad (1)$$

Where $\wedge$ is the classical fuzzy *'and'* operator (i.e. the minimum).
In general let *R* is the membership function of a time relationship. Let *T* be the membership function of an event, then we can point to a new moment of the video using the general formula:

$$R \circ T(x) = \max_y\left(R(x-y) \wedge T(y)\right) \qquad (2)$$

We remark that the event *T* can also be a time positioning, like *beginning*, *middle* or *end* of the video.

Combining.

We may also want to use the richness of the logic operators to create more precise notions. We remark that we can only combine notions having the same time scale definition. We can in these way combine two different relationships, two different times descriptors, two different events, two different times positioning, but also an event and a time positioning.
We propose to use the common fuzzy logic operators: Zadeh's operators. We will have that the logic '*and*' operation will be computed by the **min**; the '*or*' by the **max** and negation by function $x \to not(x) = 1 - x$. Now that we defined the dictionary and a sort of grammar based on the logical operators, we can construct any query.

Conclusions

In this paper we presented a model for fuzzy query system based on fuzzy annotations. This model is in the spirit of Zadeh's idea of "Computing with words". We introduced a dictionary with the basic concepts and the way to construct new ones. Computing with words allows us to have a human friendly interface. With this vocabulary and the logic tools introduced, the user will be able to realize human kind queries. We are implementing this model on a Java based Video Search Engine. For more information about this program contact the authors.

Acknowledgment

The Java based Video Search Engine is being developed by the multimedia indexing group at LIP6 (University Paris 6). The group is partially funded by the Esprit / Avir project.

Bibliography

[1]. Nwosu, K. C., Thuraisingham, B. and Berra, P.B., *Multimedia Database Systems*, Kluwer Academic Publisher: Boston, 1996.

[2]. Davis, M., Knowledge representation for video, in *Proc. of the 12th National Conference on Artificial Intelligence,* AAAI Press, 120-127, 1994.

[3]. Zadeh, L. A., Fuzzy logic = computing with words, *IEEE Transactions on Fuzzy Systems* 59, 125-148, 1993.

[4]. Vitek, M., Fuzzy information and fuzzy time, *Proc. IFCA Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis,* Marseille, 159-162, 1983.

[5]. Cobb, M. and Petry, F., Fuzzy querying binary relationships in spatial databases, *Proceedings of the 1995 IEEE,* International Conference on Cybernetics and Society, Vancouver, 1452-1458, 1995.

[6]. Hjelsvold, R. and Midstraum, R., Modeling and querying video data, in *Proc. the 20th VLDB Conference,* Satiago de Chile, 686-694, 1994.

[7]. Mackay, W.E., EVA: An experimental video annotator for symbolic analysis of video data, *SIGCHI Bulletin 21*, 68-71, 1989.

[8]. Hibino, S. Rundenstein, E. A., A visual multimedia query language for temporal analysis of video data, in *Multimedia Database Systems*, Kluwer Academic Publisher: Norwell, MA. 123-159, 1996.

[9]. Snodgrass, R., The temporal query language TQUEL, *ACM Transactions on Database Systems* 12, 247-298, 1987.

[10]. Yager, R. R., "Fuzzy temporal methods for video multimedia information systems", *Journal of Advanced Computational Intelligence 1*, 37-45, 1997.