M. Detyniecki, "Accelerating Imprecise Temporal Queries for Video Navigation," *Proceedings of the International Workshop on Fuzzy Logic and the Internet - FLINT'2001,* Berkeley, USA, pp. 131-135, August, 2001.

Accelerating Imprecise Temporal Queries for Video Navigation

Marcin Detyniecki
Berkeley initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
United Sates of America
marcin@eecs.berkeley.edu

Abstract
In this paper we present first a new method for computing temporal queries based on fuzzy time vocabulary. And secondly, since temporal analysis is usually heavy, we propose a parameterized pseudo-associative t-norm that reduces the computational time, without loosing generality.

1. Introduction

In this paper, we focus on how to navigate in an annotated video by making temporal queries. The annotations 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 at what time (of the video) something happens, but we are not completely sure about everything associated with the event.

We introduced in [4], following the spirit of Zadeh's idea [14] of "Computing with words", a dictionary with the basic time related concepts. With this vocabulary and the logic tools introduced, the user is able to realize human type queries. Here, we focus on the time related queries [11],[1],[12]. We show a general way of how to compute a solution. However, the operators involved are usually heavy from the computational point of view. We propose a solution that reduces computational effort by relaxing some basic properties, but without loosing generality.

Let us start by explaining how the classical video query systems work and what we exactly propose.

2. Fuzzy Continuous Annotations

The actual works on query systems for video are based on the use of annotations (see [7], [8], [6], [10], [12]). These annotations can be considered as information contained in a database associated to the video and indexed by the time.

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 0.75"*. Which means that we think that the actor is Robert but we are not totally sure. We notice that the indexing time (6 minutes) is considered as certain.

We speak about *continuous* annotations because we have the information for every time. Now, we can 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.

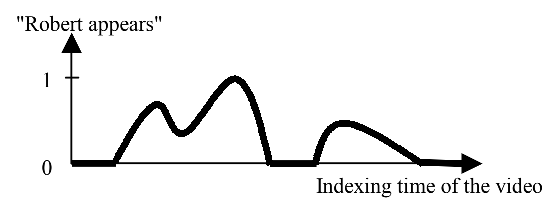

Figure 1. Fuzzy Annotation "Robert appears".

3. Placing the video player

Placing the video player at the starting time is not a trivial issue. Since we have fuzzy annotations, we do not know exactly when the event starts. Let us assume that we want to see when Robert appears. If we just use the certain information, the video player will always be placed after the real start (i.e. not when Robert appears but some seconds after). This will force the user to rewind in order to see the beginning, and it is not something we want. Taking this into account, we may think that a good solution is to start at the point where the certainty is not null (i.e. where the membership function starts). This time the video player will be placed too far in advance and the user will have to wait until the event happens, which is also not a good solution.

The action of indicating the exact start time can be seen as a defuzzification process. We use an approach, based on the alpha-cuts. The idea is to work by alpha-cuts. Here, we propose to simply take as starting time(s) the minimum(s) of the (intervals) of the 1/2 cut. This gives us a point(s) where we are more or less sure that it starts. We pre-select this alpha-cut, but we leave the possibility to the advanced user to change its attitude for the defuzzification, by of increasing and decreasing the alpha value (see figure 2).

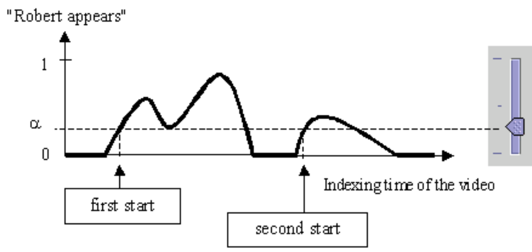

Figure 2. Placing the player.

For more details on this kind of defuzzification method in more general framework refer to our paper [2].

4. Fuzzy time vocabulary

In the spirit of Zadeh's idea of "computing with words" [14], we propose in [4] to construct a fuzzy time related dictionary. This thesaurus allows us to "precisiate" natural time querying. Using this we are able to use time positioning definitions such as *beginning*, *end* and *middle*, to use imprecise time duration such as *about five minutes*, *long* and *short time* and to use time relationship like *after*, *before* and *close*. In [4] we also present how to modify and combine this notions.

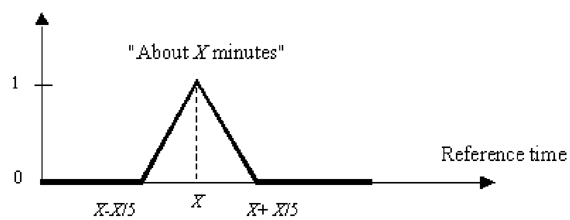

Figure 3. Example of fuzzy time vocabulary.

5. Fuzzy time relationship

We also defined relationships between time events, like for instance *after* and *before*. We based our approach on Yager's general framework for relative temporal relationship (see [12]). In 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 "*X after Y* " is 0 and if X-Y ≥0 the degree will be 1. In a symmetric way we can define the notion "*before*". We proposed in [4] to use the time descriptors in order to generate new notions as for instance: *"About 10 minutes after"* or *"About 10 minutes before*".

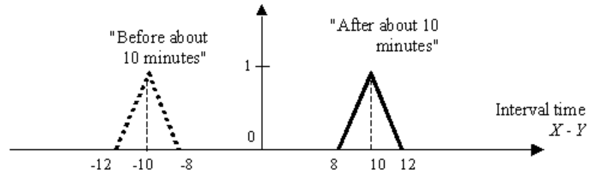

Figure 4. About 10 minutes after (or before).

6. Resolving time relationships

Now, using these relationships we may want to point to a particular moment in the video. For instance to answer the query: "*About 10 minutes after the crash*". Let $Crash(y)$ be the membership degree of the annotation "*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 following formula, where T is a t-norm :

$$\begin{aligned} &About\_10m\_after\_Crash(x) = \\ &\quad \max_y \left[T\left(after\_about\_10m(x-y), Crash(y) \right) \right] \end{aligned} \tag{1}$$

Formula (1) computes the "best answer" (max) for the logical conjunction (t-norm) of *"after about 10 minutes"* AND the *"crash"* event.

Let us generalize this result and let R be the membership function of a time relationship and E the membership function of an event, then we can point to a new moment of the video by using the general formula:

$$R \circ E(x) = \max_y \left[T\left(R(x-y), E(y) \right) \right] \tag{2}$$

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

7. Choosing the t-norm

Formula (1) allows calculating simply the degree of membership at the time x of specific temporal query. However, the choice of the t-norm T is not clear. So we initially proposed to use a parameterized t-norm, so that the user can adjust the attitude of his aggregation.

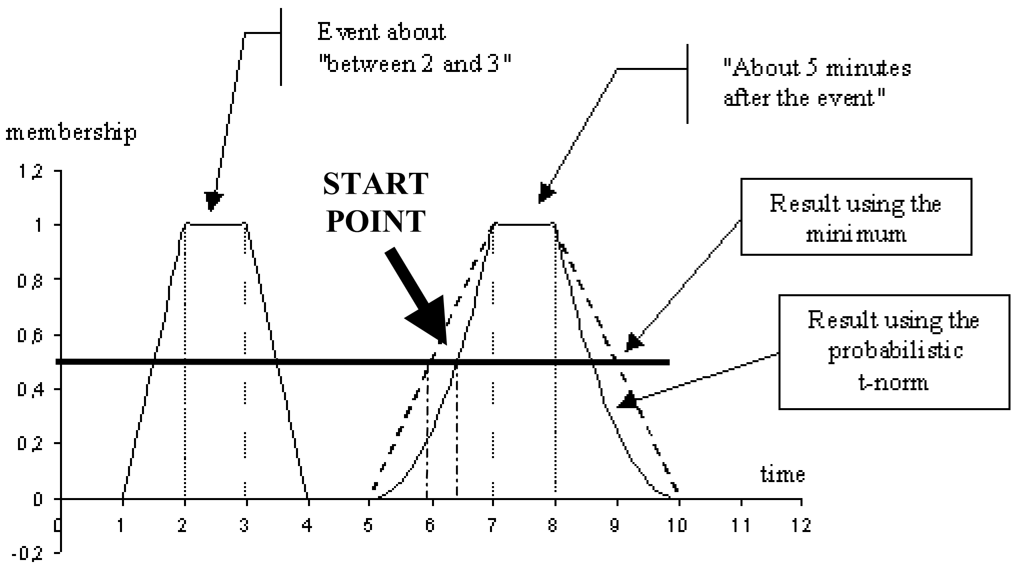

Figure 5. Pointing to "about 5 minutes after an event that happened more or less between 3 and 5 minutes"

In figure 5 we are looking for "about 5 minutes after an event that happened more or less between 3 and 5 minutes". Here "about 5 minutes" is a triangular fuzzy number being tolerant for more or less one minute. We notice that for a fixed defuzzification alpha-cut, we can have different starting points when using different t-norms. For instance when using the minimum we obtain a considerable earlier start than when using the product.

It is to notice that we are differentiating the t-norms by their attitude when aggregating the uncertainty. Since this is a personal choice we left the possibility of changing the attitude by using a parameterized t-norm. In [3] we present a methodology that is intended to help the user in the choice of the attitude by changing the parameter.

It is also important to note that all the t-norms have the same behavior when aggregating a certain value (see [3]). This induces that no matter what t-norm we use, we will always obtain the same certain interval (i.e. the same core). We can conclude that if we do not care about the uncertainty, then we can use any t-norm. This may happen if we just want to point the certain areas. However, the practice shows that this approach is deficient, since it is too strict. With just certain information the video-player is usually placed after the beginning of the event, forcing the user to rewind. Also a "only certain" approach will ignore all the imperfect annotations, in particular most of the automatic ones.

8. Using fast operators

It is clear that an interesting solution for the previous choice of the t-norm is to pick out a parameterized t-norm with a large attitude range. Like this we have always the possibility of choosing (from a large spectrum) the attitude with respect to the full uncertainty. For instance, we may select the Yager t-norm [13] (see [5] more complete justifications):

$$Y(u,v) = 1 - \left[(1-u)^p + (1-v)^p\right]^{1/p} \tag{3}$$

However, looking at (2) we notice that we have to compute for every time x of the video, the aggregation (by the t-norm) of every time y in order to take the maximum. Taking into account that at least we have 25 frames per second, we have very quickly a great number of calculi. For instance just for *one* time relationship in a *one-hour-video*, we will have to compute 8.100.000.000 times the t-norm.

Examining the proposed algorithm and equation (2) we observe that we never use the associativity property of the t-norms. In fact each time we just aggregate two values. Here a fast operator that having similar properties as the Yager t-norm may be interesting. In [8] we study a family of operator with a relax type of associativity (pseudo-t-norms). From this work and obtain the following operator:

$$R(u,v) = \max\left((1-2t)\cdot\left(\max(u,v)-1\right)+\min(u,v),\;\; 0\right) \quad (4)$$

Where t is a parameter in the range [0,1/2]. Note that this parameter translates the attitude with respect to the total uncertainty. In fact it is the result of the aggregation at the 1/2 alpha-cut:

$$R^{\min-\max}\left(\frac{1}{2},\frac{1}{2}\right) = t \quad (5)$$

It is also to notice that this pseudo t-norm generalizes some basic t-norms. In fact, if we choose that the attitude with respect to the "total fuzziness" should be relaxed (i.e. t =1/2) then the operator (4) becomes the minimum (the largest t-norm). And for the strictest attitude (i.e. t =0), (4) becomes the Lukasiewicz t-norm. It is possible to obtain stricter operators than the Lukasiewicz t-norm by using negative parameter t. These particular cases will be bigger than the limit drastic t-norm, but we consider that their attitude is too strong. In fact all these last particular cases cut off everything with a degree under 1/2.

Now, coming back to the computational issues. Here for illustration proposes we are going to compare operator (4) to the Yager t-norm. But note that the same comparison can be done to any parameterized t-norm and we would obtain the same result.

***Yager t-norm: Y*(*u*,*v*)**	***Operator: R* (*u*,*v*)**
1 division - 1/p	1 product
1 additions	1 addition
3 subtractions	1 subtraction
3 power operations	1 comparison (*min*, *max*)

Table 1. Effort comparison between Yager t-norm and the reduced min.

Looking at Table 1, we notice that the new operator is computationally lighter. In fact, we observe that Yager t-norm has a division that has the same complexity as a product.
We immediately see that the calculus of the Yager t-norm is more arduous. In fact, the two operators have then same amount of additions and product, but Yager t-norm has one more subtraction. This does not make a large difference. The main difference appears when we observe that Yager t-norm needs 3 power operations, while the new operator only needs one comparison: the bigger number will be used for the max and the other one for the min.
It is to notice that in the n-ary case (see [5]) aggregation using Yager t-norm is heavier. In other words the new operator (4) is interesting when aggregation several times just couple of values, but also for the calculus of large number of arguments.

The fact that the reduced minimum is computationally faster than the Yager t-norm is interesting, but we may think that the price for this is that we lost the associativity property. In fact, the reduced minimum is not associative, but it is pseudo-associative as explain in [5]. The idea is that we can still aggregate by packages only by keeping in memory the maximum and minimum.

9. Application

A Java based Video Search Engine was developed by the multimedia indexing group at LIP6 (University Paris 6) for the Esprit / Avir project (see [9]).

Figure 6. Java based Video Search Engine.

10. Conclusions

In this paper we presented a model for answering to imprecise temporal query. The system uses fuzzy annotations and a time related dictionary. This thesaurus "precisiates" what the user understands for a particular concept. This allows us to achieve a human friendly interface.
In this paper we focused on the fact that the calculus of time relationship is extremely heavy, so we propose an operator that lightens this computation. It is to notice that the presented operator is not only interesting for this particular example, but it is also an interesting solution for practical efficient applications, where use of a t-norm is required. Its computational lightness combined to its pseudo-associativity, without

forgetting its generalization property make of it a power-full tool.

11. Acknowledgment

This work was founded by the French Government in the form of grants. I would like to thank Claude Seyrat for the development of the Java Video Search Engine and Ronald Yager with whom I started this work.

12. References

[1] 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.

[2] Detyniecki M. and Yager R., Ranking fuzzy numbers using α-weights, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, to appear in 2000.

[3] Detyniecki M., Yager R. and Bouchon-Meunier B., Specifying t-norms based on the value of T(1/2,1/2), *Mathware & Soft Computing vol.VII*, number 1, 2000.

[4] Detyniecki M., Seyrat C. and Yager R., Interacting with Web Video Objects, *Proceedings of 18th International Conference of the North American Fuzzy Information Society - Nafips'99*, pp.914-917, New York, USA, June 1999.

[5] Detyniecki M., Mathematical Aggregation Operators and their Application to Video Querying, *Doctoral Thesis - LIP6 research reports 2001,* number 002, November 2000.

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

[7] Hjelsvold, R. and Midstraum, R., Modeling and querying video data, *Proceedings of the 20th VLDB Conference,* Santiago de Chile, 686-694, 1994.

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

[9] Seyrat C. and Detyniecki M. On Multimedia Indexing, *Second International Workshop on Knowledge Representation for Interactive Multimedia Systems - KRIMSII*, Trento - ITALY. June 1998.

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

[11] Vitek, M., Fuzzy information and fuzzy time, *Proceeding of IFCA Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis,* Marseille, 159-162, 1983.

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

[13] Yager, R. R., On a general class of fuzzy connectives, *Fuzzy Sets and Systems 4*, 235-242, 1980.

[14] Zadeh, L. A., Fuzzy Logic = Computing with Words, *IEEE Transactions on Fuzzy Systems* 59, 125-148, 1993.