

# Video Content Description Using Fuzzy Spatio-Temporal Relations

Archana M. Rajurkar<sup>1</sup>, R.C. Joshi<sup>2</sup>,  
Santanu Chaudhary<sup>3</sup> and Ramchandra Manthalkar<sup>4</sup>

<sup>1</sup>*Dept. of Computer Sc. and Engg. M.G.M.'s College of Engineering, Nanded – 431 605,*

<sup>2</sup>*Dept. of Electronics and Computer Engg. Indian Institute of Technology Roorkee,  
Roorkee - 247 667,*

<sup>3</sup>*Dept. of Electrical Engg., Indian Institute of Technology, Delhi, New Delhi,*

<sup>4</sup>*Dept. of Electronics and Telecommunication Engg. S.G.G.S. Institute of Engineering,  
and Technology Nanded – 431 605,  
India*

## 1. Introduction

There has been an explosive growth in multimedia data such as images, video and audio in the past few years. The rapid proliferation of the Web made a large amount of video data publicly available. This necessitates a system that provides the ability to store and retrieve video in a way that allows flexible and efficient search based on semantic content.

Most of the existing video retrieval techniques search video based on visual features such as color, texture and shape (Zhang et al., 1997; Chang et al., 1998; Hampapur et al., 1997). A video sequence is first segmented into shots, each shot is then represented in terms of a number of key frames and then the visual features are used for retrieval (Zhang et al., 1993; Zhang et al., 1995; Mottaleb et al., 1996). Key frame based methods do not always consider events occurring in the video that involves various objects and choosing the key-frames is a challenging problem. Furthermore, in these approaches the temporal nature of video is neglected. A few systems have addressed the issue of object-based video retrieval (Deng & Manjunath, 1998; Courtney, 1997) and spatial modeling of video data that involve temporal information (Bimbo et al., 1995; Little & Ghafoor, 1993; Dagtas & Ghafoor, 1999; Vazirgiannis et al., 1998).

The semantic modeling of video content is a difficult problem. One simple way to model the video content is by using textual annotation (Oomoto & Tanaka, 1993) or visual features (Zhang et al., 1997; Chang et al., 1998), but the annotation is tedious and time consuming and simple visual features are not always sufficient to represent the rich semantics of digital video. Another way to model the video content is by using spatio-temporal concepts or events that describe the dynamic spatio-temporal behavior and interaction of video objects. As humans think in terms of events and remember different events and objects after watching a video, these high-level concepts are the most important cues in content-based retrieval (Petkovik & Jonker, 2001). For example, in a football game, we usually remember goals, interesting actions *etc.* Therefore there is a great need for development of robust tools

for modeling and extraction of spatio-temporal relations among objects visible in a video sequence.

Some approaches have been suggested in the literature for modeling the spatial and temporal relations of video objects. In most of the approaches, spatial relationships are based on projecting objects on a two or three-dimensional coordinate systems (Dagtas & Ghafoor, 1999), while temporal relations are based on Allen's temporal interval algebra (Allen, 1983). Some models, which use temporal relations for video retrieval, are proposed in (Oomoto & Tanaka, 1993; Little & Ghafoor, 1993). Very few efforts integrate both spatial and temporal aspects (Dagtas, 1998; Vazirgiannis et al., 1998; Nepal & Srinivasan, 2002; Pissiou et al., 2001). A Spatio-Temporal Logic (STL) language is presented in (Bimbo et al., 1995), which is used to describe the contents of an image sequence. A prototype image retrieval system supporting visual querying by example image sequences is discussed in this but; the problem of modeling higher-level concepts of spatio-temporal events is not addressed. A framework for semantic modeling of video data using generalized n-ary operators is presented in (Day et al., 1995). A graph-based model, called Video Semantic Directed graph (VSDG) is proposed in (Dagtas, 1998) for unbiased representation of video data using directed graph model. It also suggests the use of Allen's temporal interval algebra to model relations among objects. A model for complete declarative representation of spatio-temporal composition is presented in (Vazirgiannis et al., 1998). Nepal and Srinivasan (2002) presented a binary representation-based framework for modeling and querying video content using video object relationships at a semantic level that supports conceptual spatio-temporal queries. A topological-directional model for spatio-temporal composition of objects in video sequences is presented in (Rajurkar and Joshi, 2001). It describes the spatio-temporal relationships among objects for each frame and models temporal composition of an object.

Most of the previous spatio-temporal models do not deal with extraction of spatio-temporal relations rather they use handcrafted precise definitions of spatial relations (Egenhofer & Fanzosa, 1991) and temporal relations (Allen, 1983). It is not always possible to precisely characterize spatial and temporal relations between objects in the video sequence because of the inherent dynamic nature of the media. Further, spatial relations such as LEFT, ABOVE and others defy precise definitions, and seem to be best modeled by fuzzy sets (Matsakis & Wendling, 1999). Furthermore, errors may occur in shot segmentation and object detection due to non-robustness of the image/video processing operations leading to erroneous inference of crisp spatio-temporal relations between objects. By using fuzzy definitions of spatio-temporal relations, we can take care of these errors and ambiguities. Consequently, this approach facilitates video modeling as well as query processing based upon spatio-temporal relations extracted automatically from the actual video data without manual intervention.

Motivated by the above observation, we present a new approach for video content description. A fuzzy spatio-temporal model that is based on fuzzy directional and topological relations and fuzzy temporal relations between video objects is proposed. We use the linguistic definitions of spatial relations using *histogram of forces* presented in our earlier work (Rajurkar & Joshi, 2001). Fuzzy definitions and membership functions of temporal relations are presented and the second order fuzzy temporal relations have been proposed. A video representation scheme is presented using the proposed fuzzy spatio-temporal model. We have illustrated query processing with our fuzzy spatio-temporal model.

The rest of the chapter is organized as follows. Section 2 describes the method used for object detection using motion segmentation. The proposed fuzzy spatio-temporal model is presented in section 3. Query processing and retrieval results are given in section 4 and conclusions are presented in section 5.

## 2. Object detection using motion based segmentation

In our approach, a video sequence is modeled in terms of fuzzy spatio-temporal relations between objects visible in the video sequence. For sake of simplicity it is assumed that the objects of importance in a video sequence are in motion and simple motion based segmentation scheme for identification/detection of the objects is used in this work. A simple approach for motion-based segmentation using difference picture has been described here. More sophisticated approaches can improve the system performance.

In the proposed model, moving objects in every frame are detected in the video sequence using motion segmentation method described in (Courtney, 1997). For each frame  $F_n$  in the video sequence, the motion segmentation stage computes segmented image  $C_n$  as

$$C_n = T_h \bullet k$$

Where  $T_h$  is a binary image resulting from the absolute difference of images  $I_n$  and  $I_0$  at threshold  $h$ .  $T_h \bullet k$  is the morphological close operation on  $T_h$  with structuring element  $k$ . The image  $T_h$  is defined for all pixels  $(i, j)$  in  $T_h$  as

$$T_h(i, j) = \begin{cases} 1 & \text{if } |I_n(i, j) - I_0(i, j)| \geq h \\ 0 & \text{otherwise} \end{cases}$$

Connected component analysis is then performed on the segmented image  $C_n$  resulting in a unique label for each connected region. Each connected component is further recognized and identified manually for constructing the offline video model.

An example of motion segmentation process explained above is shown in Fig. 1. The reference image and the image to be segmented are shown in Fig.1 (a) and (b) respectively. Absolute difference and thresholded image is shown in Fig. 1(c) that detect motion region in the image. Fig. 1(d) shows the morphological close operation that joins small regions together into smoothly shaped objects. The result of connected component analysis that assigns each detected object a unique label is shown in Fig. 1(e). The output of the motion segmentation stage is  $C_n$  after discarding components, which are smaller than a given threshold. Fuzzy spatio-temporal relations between these objects are then defined.

Such a motion segmentation technique is best suited for video sequences containing object motion within an otherwise static scene, such as in surveillance and scene monitoring applications (Courtney, 1997).

## 3. Fuzzy spatio-temporal model

Spatio-temporal relations of the salient objects play a crucial role in characterization of the "content" of video. The spatial characteristics are referred to as spatial relationship between any two objects in a sampled frame, whereas the temporal characteristics are referred as the dynamics of the spatial relationships between every two objects over the frames in a video sequence.

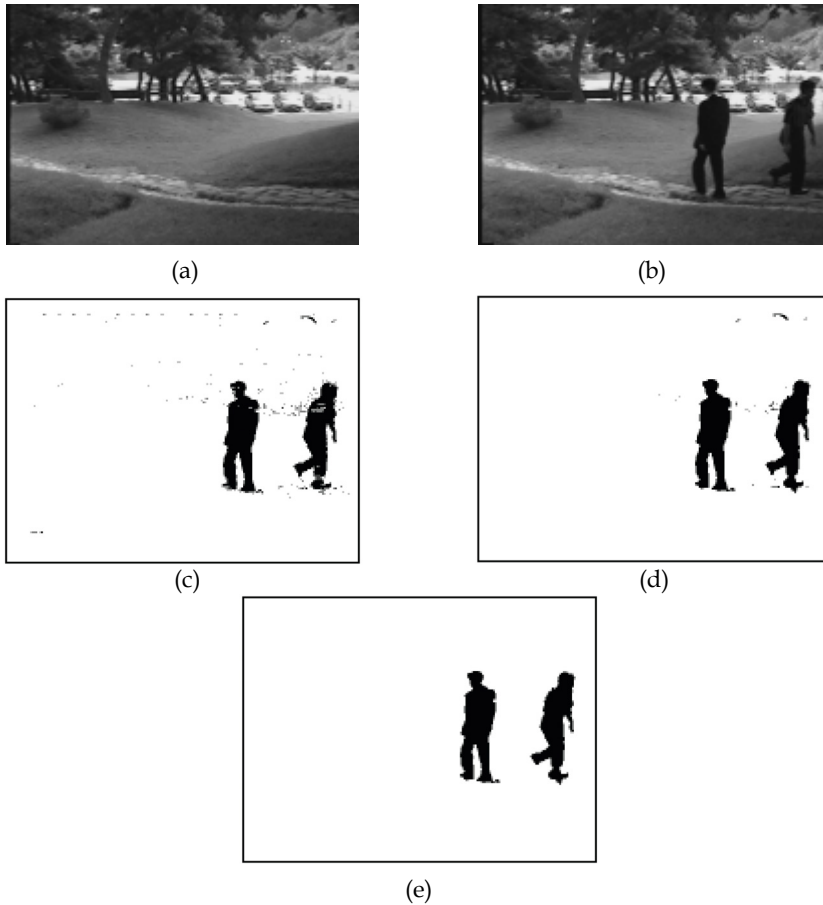


Fig. 1. Motion Segmentation example: (a) Reference Image  $I_0$  (b) Image  $I_n$  (c) Thresholded Image  $T_h$  (d) Result of Morphological Close Operation (e) Result of Connected Component Analysis

In the proposed model the description of a scene is expressed by the mutual spatial relationships between every two objects and the temporal change in their relationships. Frame intervals specify the period for which a particular spatial relationship holds. For each object  $O_i t$  in a frame  $t$ , its fuzzy spatial and temporal relationships,  $O_{ST_{ij}} t$ , with every other object,  $O_j t$ , in the same frame are recorded in a vector of size  $n(n-1)$ , where  $n$  is the number of objects in the frame  $t$ . The fuzzy spatio-temporal relationship between the two objects is defined as the function:

$$O_{ST_{ij}} t (O_i t, O_j t) = (S_{ijt}, T_{ijt}) \quad (1)$$

where  $S_{ijt}$  and  $T_{ijt}$  are the spatial and temporal relationships, respectively, between the objects  $O_i t$  and  $O_j t$  in the frame  $t$ .

### 3.1 Spatial relationships

The relative positions between two objects  $O_i$  and  $O_j$  can be captured as a fuzzy spatial relation using *histogram of forces* (F-Histogram) (Matsakis & Wendling, 1999). The spatio-temporal models presented earlier in (Dagtas & Ghafoor, 1999; Nepal and Srinivasan, 2002; Pissiou et al., 2001) use precise definitions of spatial relations using either angle measurements or minimum bounding rectangles (MBR). Spatial relations such as *left of*, *above* and others defy precise definitions, and seem to be best modeled by fuzzy sets (Matsakis & Wendling, 1999). Miyajima and Ralescu (1994) developed the idea of representing the relative position between two objects by *histogram of angles*. Matsakis and Wendling (1999) introduced the notation of the F-histogram, which generalizes and supersedes that of the histogram of angles is discussed in the following section.

### 3.2 F-histograms

The F-histogram represents the relative position of a 2D object A with regard to another object B by a function FAB from  $\mathbf{R}$  in to  $\mathbf{R}^+$ . For any direction  $\theta$ , the value FAB( $\theta$ ) is the total weight of the arguments that can be found in order to support the proposition "A is in direction  $\theta$  of B". Object A is the argument and object B is the referent. The directional spatial relations between objects are defined from a fuzzy subset of  $\mathbf{R}$ . Its membership function  $\mu$  is continuous, with a period  $2\pi$  that decreasing on  $[0, \pi]$ , and takes the value 1 at 0 and the value 0 at  $\pi/2$  shown in Fig. 2(a). It can be employed to define a family of fuzzy directional relations between points (Rajurkar & Joshi, 2001). We used the typical triangular function graphed in Fig. 2(a). Let  $\alpha$  and  $\beta$  be two reals and A and B be two distinct points. If  $\beta$  is an angle measure, then (Fig. 2(b)):  $R_\alpha(A, B) = \mu(\beta - \alpha)$ . In our earlier work (Rajurkar & Joshi, 2001) we have presented definitions of fuzzy directional spatial relations and topological relations using F-histogram. We make use of these definitions of fuzzy spatial relations as perceived by humans for capturing relative position of a 2D object  $O_{it}$  with regards to another object  $O_{jt}$ . Each pair of objects in every frame in the video sequence is represented by relative position histograms and then the degree of truth of a proposition "A is in direction  $\theta$  of B" is computed. The degree of truth is a real number generated greater than or equal to 0 (proposition completely false) and less than or equal to 1 (proposition completely true).

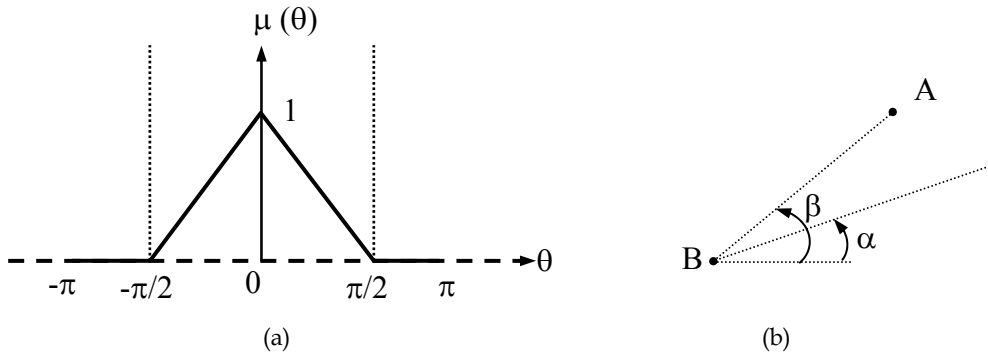


Fig. 2. (a) Example of directional spatial relation between points. (b) The degree of truth of the proposition "A is in direction  $\alpha$  of B" is  $\mu(\beta - \alpha)$

The spatial relationships,  $S_{ijt}$ , between two objects are defined as follows:

$$S_{ijt} = (R_{ijt}, O_{it}, O_{jt}) \quad (2)$$

The  $R_{ijt}$  represents the degree of truth of a proposition “ $A$  is in direction  $\theta$  of  $B$ ” computed as described in (Rajurkar & Joshi, 2001).

### 3.3 Temporal relationships

Allen (1983) introduced temporal interval algebra for representing and reasoning about temporal relations between events represented as intervals. The elements of the algebra are sets of seven basic relations that can hold between two intervals. The relations are not reflective except equals and a set of inverse relations exists for the other six relations. A list and graphical illustrations of the binary interval relations is provided in Fig. 3. A qualitative description of relative position of objects (*MBRs*) when their projections on X and Y-axes are considered as interval operands to the relations are provided by interval relations.

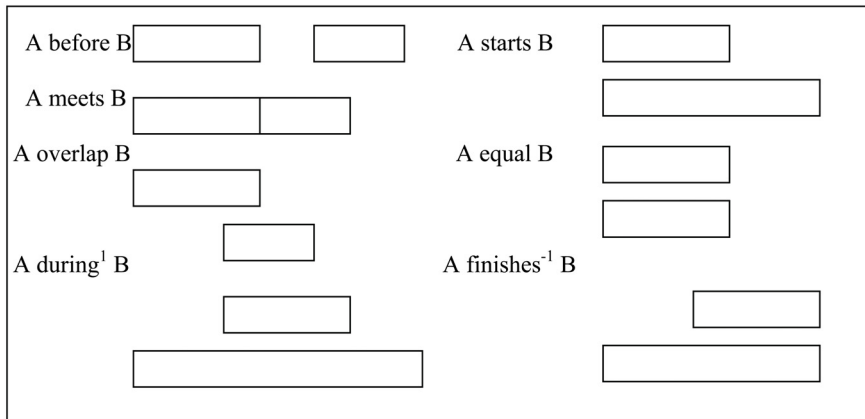


Fig. 3. Binary Interval Relations

Crisp description of temporal relations may not be suited for defining spatio-temporal models of video because of errors in segmentation, event and object detection due to noise and processing errors. We present some examples to illustrate this point. Figure 4 shows an example of error occurred in segmentation. Original image is shown in Fig. 4(a). Thresholded absolute difference image is shown in Fig. 4(b), and the results of motion segmentation after morphological close operation and connected component analysis is shown in Fig. 4(c). It is observed that though the objects *Person 1* and *Person 2* are disjoint after the segmentation they appear to be meet. Further it is also shown that how errors can occur in event detection and segmentation of noisy images. Fig. 5(a) shows an example of noisy image. The thresholded absolute difference image is shown in Fig. 5(b), and results of segmentation are given in Fig 5(c). It can be seen that though the objects *Person 1* and *Person*

2 are disjoint they appear to be meet after segmentation. Fig. 6 shows an example of error occurred in object detection. Fig. 6(a) shows the original image and the results of segmentation are presented in Fig. 6(b). It can be observed that half of the part of *Person2* is missing in the result, which may lead to wrong event and object detection. We present fuzzy definitions of temporal relations in this section that can take care of these errors in event-detection, segmentation and object detection and allows flexible description of video scene (see Table 1). Relations defined in Table 1 are applicable between objects, in particular for the intervals for which two objects are visible. For example, *A before B* with a membership function "mu" means that the *object A* is not visible, in general, in the scene when *B* appears. The temporal relation  $T_{ij}$  between the spatial relationships,  $S_{ijt}$ , of objects  $O_i$  and  $O_j$  can be described at two levels. In the first level, the temporal interval  $\Delta f$  for which a fuzzy spatial relationship between the two objects is valid is determined. In the second level, the second order fuzzy temporal relationships between the two spatial relations are described. The advantage of the second order fuzzy temporal relations is that they are more informative and provide global description of a sequence. The proposed second order fuzzy temporal relations using fuzzy spatial relations are shown in the Table 2. Graphical illustration of two second order fuzzy temporal relations are presented in Fig. 7. The temporal relationship,  $T_{ij}$ , between the temporal intervals of two spatial relationships,  $S_{ijt}$  and  $S_{ijt'}$  is defined as follows:

$$T_{ij}^1 = (S_{ijt} \Delta f)$$

$$T_{ij}^2 = (S_{ijt} \Delta f \langle FTOP \rangle S_{ijt'}) \quad (4)$$

where  $\langle FTOP \rangle$  is one of the temporal operators representing the fuzzy temporal relationship between two intervals and  $\Delta f$  is temporal interval for which fuzzy spatial relationship  $S_{ijt}$  is valid.

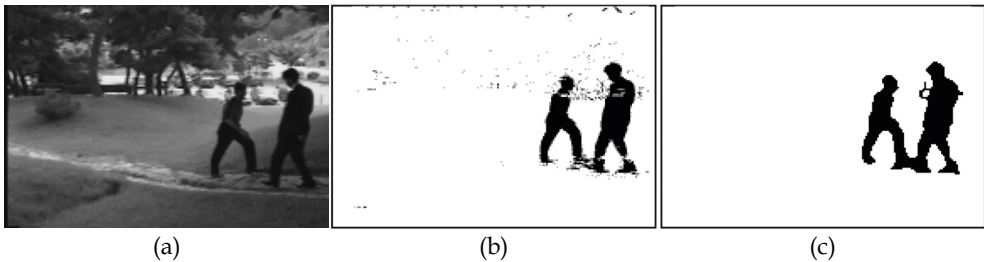


Fig. 4. Error Occurred in Segmentation (a) Original Image (b) Thresholded Absolute Difference Image (c) Results of Segmentation

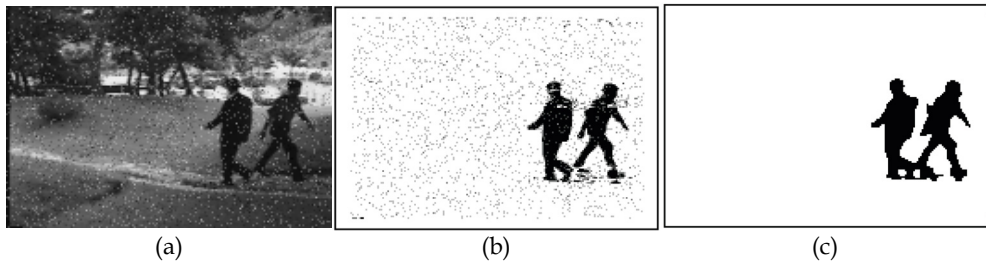


Fig. 5. Error Occurred in Segmentation (a) Noisy Image (b) Thresholded Absolute Difference Image (c) Results of Segmentation

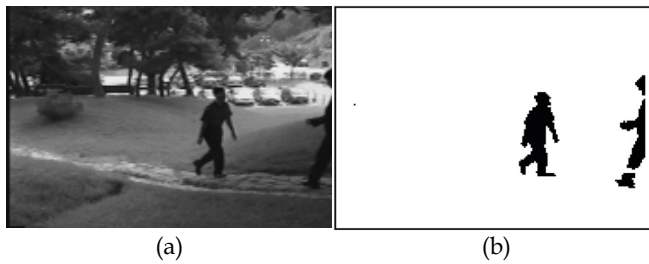


Fig. 6. Error Occurred in Event and Object Detection (a) Original Image (b) Results of Segmentation

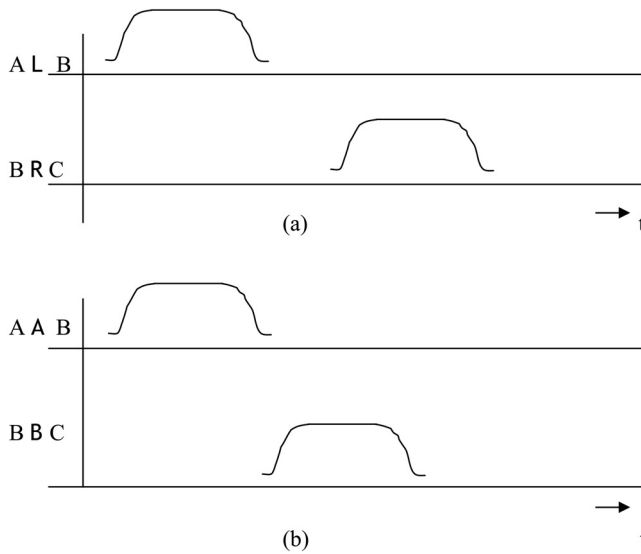


Fig. 7. Graphical Illustration of Second Order Fuzzy Temporal Relations (a) Relation  $A L B$   $\mu_{\text{before}} B R C$  (b) Relation  $A A B$   $\mu_{\text{meets}} B B C$



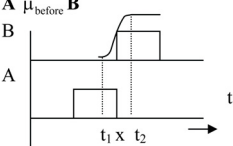
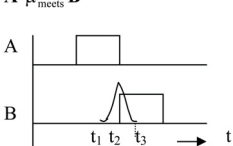
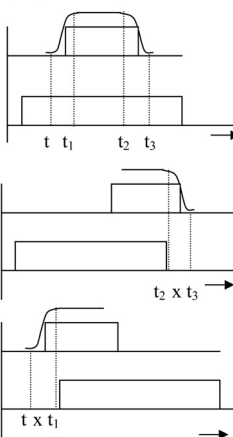
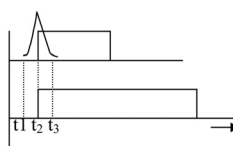
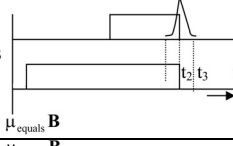
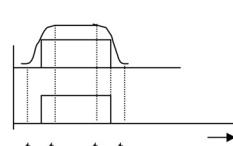
Fuzzy Temporal Relationship	Membership Function	Logical Description
 <p><math>\mu_{\text{before}}</math></p>	$\mu_{\text{before}}(x) = 1 \quad x \geq t_2$ $\mu_{\text{before}}(x) = e^{-(x-t_2)^2} \quad x < t_2$	Object A disappear before the appearance of the object B, but there may be some overlap
 <p><math>\mu_{\text{meet}}</math></p>	$\mu_{\text{meet}}(x) = 1 \quad x = t_1$ $\mu_{\text{meet}}(x) = e^{-(x-t_2)^2} \quad x < t_2, x > t_2$	Object A meet object B, but there may be some overlap or they may be disjoint
 <p><math>\mu_{\text{during}}</math></p>	$\mu_{\text{during}}(x) = 1 \quad t_1 \leq x \leq t_2$ $\mu_{\text{during}}(x) = e^{-(x-t_1)^2} \quad x < t_1$ $\mu_{\text{during}}(x) = e^{-(x-t_2)^2} \quad x > t_2$	Object A is appearing and disappearing during object B, but it may appear before object B or it may disappear after object B.
 <p><math>\mu_{\text{starts}}</math></p>	$\mu_{\text{starts}}(x) = 1 \quad x = t_2$ $\mu_{\text{starts}}(x) = e^{-(x-t_2)^2} \quad x < t_2, x > t_2$	Object A starts exactly with Object B, but it may start little before or after object B.
 <p><math>\mu_{\text{finishes}}</math></p>	$\mu_{\text{finishes}}(x) = 1 \quad x = t_2$ $\mu_{\text{finishes}}(x) = e^{-(x-t_2)^2} \quad x > t_2, x < t_2$	Object A finishes exactly with Object B, but it may finish little before or after object B.
 <p><math>\mu_{\text{equals}}</math></p>	$\mu_{\text{equals}}(x) = 1 \quad t_1 \leq x \leq t_2$ $\mu_{\text{equals}}(x) = e^{-(x-t_1)^2} \quad x < t_1$ $\mu_{\text{equals}}(x) = e^{-(x-t_2)^2} \quad x > t_2$	Object A starts and finishes exactly with Object B, but it may start or finish little before or after object B.

Table 1. Fuzzy Temporal Relationships and Corresponding Logical Descriptions

### 3.4 Representation of video sequence using fuzzy spatio-temporal model

This section presents the representation of a video sequence using the proposed fuzzy spatio-temporal model. The video database  $V_{db}$  contains video sequences  $S_1, S_2, \dots, S_n$  as follows:

$$V_{db} = \langle S_1 \rangle \langle S_2 \rangle \langle S_3 \rangle \dots \langle S_n \rangle \quad (5)$$

A video sequence  $S$  is an ordered set of  $n$  frames ( $t$ ), denoted  $S = \{t_0, t_1, t_3, \dots, t_n\}$ , where  $t_n$  is the  $n^{th}$  frame in the sequence. For each frame  $t$  in the video sequence  $S$  the moving objects are detected and labeled using motion segmentation method outlined in Section 2. Then the attributes of labeled objects are derived.

Spatial Relation 1	Temporal Relation	Spatial Relation 2	Definition of Second Order Fuzzy Temporal Relations
A left of B	Before	B right of C	$A \text{ L B } \mu_{\text{before}} \text{ B R C}$
A above B	Before	B below C	$A \text{ A B } \mu_{\text{before}} \text{ B B C}$
A left of B	Meets	B left of C	$A \text{ L B } \mu_{\text{meets}} \text{ B L C}$
A above B	Meets	B below C	$A \text{ A B } \mu_{\text{meets}} \text{ B B C}$
A left of B	Overlap	B right of C	$A \text{ L B } \mu_{\text{overlaps}} \text{ B R C}$
A above B	Overlap	B below C	$A \text{ A B } \mu_{\text{overlaps}} \text{ B B C}$
A left of B	During	B right of C	$A \text{ L B } \mu_{\text{during}} \text{ B R C}$
A above B	During	B below C	$A \text{ A B } \mu_{\text{during}} \text{ B B C}$
A left of B	Starts	B right of C	$A \text{ L B } \mu_{\text{starts}} \text{ B R C}$
A above B	Starts	B below C	$A \text{ A B } \mu_{\text{starts}} \text{ B B C}$
A left of B	Finishes	B right of C	$A \text{ L B } \mu_{\text{finishes}} \text{ B R C}$
A above C	Finishes	B below C	$A \text{ A C } \mu_{\text{finishes}} \text{ B B C}$
A left of B	Equals	B right of C	$A \text{ L B } \mu_{\text{equals}} \text{ B R C}$
A above B	Equals	B below C	$A \text{ A B } \mu_{\text{equals}} \text{ B B C}$

Note: Set of inverse relations exists for all above relations except *equals*. The symbols used for spatial relations are L - left of, R - right of, A - above, B - below.

Table 2. Second Order Fuzzy Temporal Relations

For each object  $O_i t$  of a frame  $t$ , its fuzzy spatial and temporal relationship  $O_{ST_{ij}} t$ , with every other object,  $O_j t$ , is represented using the proposed fuzzy spatio-temporal model as discussed in Section 3.

$$O_{ST_{ij}} t(O_i t, O_j t) = (S_{ij} t, T_{ij} t) \quad (6)$$

where  $S_{ijt}$  and  $T_{ijt}$  are defined by Eq. (2) and (3) respectively.

For each frame  $t$  fuzzy spatio-temporal relationship between all object pairs (e.g. suppose there are four objects in the frame  $t$ ) is represented as follows:

$$O_{ST}t = (O_{ST_{12}}, O_{ST_{13}}, O_{ST_{14}}, O_{ST_{23}}, O_{ST_{24}}, O_{ST_{34}}) \quad (7)$$

To capture the dynamic change in the fuzzy spatial relationship of two objects  $O_i t$  and  $O_j t$  over the video scene, interval of  $l$  frames in which the corresponding spatial relation is valid is determined. The temporal interval of a spatial relationship is found from the frame  $t$  of the initial appearing of a particular spatial relationship, which represent the beginning of the temporal interval in which that spatial relationship is valid. Then the frame  $t'$  of the initial appearance of the first different relationship is determined. Thus the duration of the temporal interval in which a particular spatial relation is valid is  $t' - t$ . Based on these durations, fuzzy spatio-temporal relations are computed for describing the video sequences. The detailed algorithm for representing video sequences in a video database using the proposed fuzzy spatio-temporal model is presented in Fig. 8.

**Algorithm Represent video<sub>fstm</sub>**

Input:  $V_{db}$ (video database)

Output: Representation of video database sequences.

Procedure:

For every video sequences in the video database do the following

1. Motion segmentation  
For each frame  $t$  in the database video sequence, the motion segmentation stage computes segmented image  $C_n$ . as  $C_n = T_h \bullet k$ . Connected component analysis is then performed on the segmented image  $C_n$  and objects are identified, labeled and their attributes are derived.
2. Compute fuzzy spatio-temporal relations  
For every frame in the video sequence  $S$ , for each object pair  $(O_i t, O_j t)$   
Compute fuzzy spatial relations  $S_{ijt} = (R_{ijt}, O_i t, O_j t)$   
Compute the temporal interval  $\Delta t$  for which the spatial relationship  $S_{ijt}$  is valid

$$T_{ij}^1 = (S_{ijt_{\Delta t}})$$

3. Compute the second order fuzzy temporal relations between every two different spatial relations.

$$T_{ij}^2 = (S_{ijt_{\Delta t}} < FTOP > S_{ij't'_{\Delta t}})$$

Fig. 8. Algorithm Represent\_video<sub>fstm</sub>

We used MPEG-7 video content *ETRI\_od\_A.mpg* to illustrate the proposed video sequence representation scheme. Fig. 9 shows few frames in the video sequence *ETRI\_od\_A.mpg*. Moving objects in the sequence are detected and labeled as *Person1*, *Person2*, *Person3* and *Person4* using the motion segmentation method described earlier in section 2 and their attributes are derived. Initial appearing of two objects *Person1* and *Person2* is detected in frame 173 and their spatial relation is “*Person1 in white shirt is left of Person2 in black shirt*” shown in Fig.9. The temporal interval in which this spatial relationship is valid starts from

frame 173 and ends at frame 290. The same spatial relationship holds for another temporal intervals from frame 2316 to frame 2343 and from the frame 2792 to frame 2900. The fuzzy spatio-temporal relationship is given by

$$O_{ST_{12}} t(O_1 173, O_2 173) = (S_{12} 173, T_{12} 173) \quad (8)$$

$$S_{12} 173 = (\text{left of, Person1, Person2})$$

$$T_{12}^I = 354$$

Similarly initial appearing of the different spatial relationships between objects *Person1* and *Person2* is detected in frame 636 and its temporal interval is computed as described above. The procedure is repeated for the initial appearing of the other object pairs in the video sequence and their fuzzy spatio-temporal relationships are computed.



Fig. 9. Few Sample Frames in Video ETRI\_od\_A.mpg

The second level temporal relationship is determined by computing the second order fuzzy temporal relationships between the two different spatial relationships. In the video sequence *ETRI\_od\_A.mpg* the first two spatial relations are detected in frame 173 and 636 and their first level temporal intervals are 116 and 187. The second level temporal relationship is given by:

$$T_{12}^2 = (S_{12_{116}} 173 \langle \mu_{\text{before}} \rangle S_{12_{187}} 636) \quad (9)$$

$$S_{12} 173 = (\text{left of}, \text{Person1}, \text{Person2})$$

$$S_{12} 636 = (\text{right of}, \text{Person1}, \text{Person2})$$

The second level temporal relationships between all the different spatial relationships found in the video sequence are determined as described above. In this way all the video sequences in the database are represented by the fuzzy spatio-temporal relationships among the objects in the sequence.

#### 4. Querying and retrieval

Formulation of meaningful and clear query is very important for searching and retrieving video. Content-based retrieval of video is rather difficult than that of the image due to the temporal information in the video. The issue of efficient handling queries related to the spatio-temporal relationships among objects is discussed here. We perform the ranking of the database sequences that are similar to the query using the relevance membership function (RMF). The temporal intervals for all spatial relations between two objects over a sequence are computed to decide the maximum number of frames for which a particular spatial relationship between two objects is valid. Depending on the query, the relevance membership function, which is the ratio of total number of frames in the sequence to the maximum number of frames for which the spatial relationship in the query is valid for every sequence in the database, is computed. The detailed algorithm *Search\_Video<sub>stm</sub>* for searching and ranking the video sequences in the database similar to the query  $V_{qr}$  is presented in Fig. 10.

In the following, we present some examples of queries using the spatio-temporal properties.

Query1: Find a video sequence in which a *Person P1* approaches *Person P2* from the left.

Select         $S$   
From        sequences  $S_n$   
Where         $S$  contains  $P1$  and  $P2$  and  $P1$  L  $P2$

Query 2: Find a video sequence in which a *Car* is moving to the right reaches to a *House*

Select         $S$   
From        sequences  $S_n$   
Where         $S$  contains a *Car* and a *House* and  $Car$  R  $House$

Query 3: Find a video sequence in which *Car A* is in the left of *Car B* before *Car C* is in the right of *Car D* in the race.

Select         $S$   
From        sequences  $S_n$   
Where         $S$  contains *Car A*, *Car B*, *Car C* and *Car D* and  $Car A$  L  $Car B$   $\mu_{\text{before}}$   $Car C$  R  $Car D$

Query 4: Find a video sequence in which *Flowers* are right to the *Lake* and a *Jeep* is passing besides the *Lake*.

Select	$S$
From	sequences $S_n$
Where	$S$ contains <i>Flowers, Lake</i> and <i>Jeep</i> and <i>Flowers</i> R <i>Lake</i> $\mu_{\text{overlaps}}$ <i>Jeep</i> .

We used MPEG-7 video content to evaluate the effectiveness of our query processing scheme. Video sequences in the database are represented using the proposed fuzzy spatio-temporal model as described in section 3.4.

#### Algorithm Search\_Video<sub>fstm</sub>

Input:  $V_{qr}$

Output: Ranking of video sequences that are similar to the query  $V_{qr}$

Procedure:

1. For every video sequences in the video database compute the relevance membership function (RMF)

$$RMF = \frac{\text{Number of frames in the database video sequence } S}{\text{Maximum number of frames for which the spatial relation in the query } V_{qr} \text{ is valid}}$$

2. Rank the video sequences in the database in the increasing order of the value of RMF

Fig. 10. Algorithm Search\_Video<sub>fstm</sub>

Now, consider a query “Find video sequences in which Person1 in white shirt is left of Person2 in black shirt”. The query results are shown in Table 3. The temporal intervals representing the number of frames for which a spatial relationship *left of* is valid for every sequence and the RMF values are given in the Table 3. Low value of RMF function corresponds to the more number of frames while higher value of RMF corresponds to the less number of frames that satisfy the spatial relationship *left* in the sequence. The sequences in the database are then ranked depending on the value of RMF. The video sequences having low RMF values are retrieved as similar to that of the spatio-temporal relation described in the query. For the query in question the most similar sequence in the video database is ETRI<sub>od\_A</sub>.mpg, which has maximum number of frames i.e. 273 for which spatial relationship *Person1 in white shirt is left of Person 2 in black shirt* is valid. The database video sequences are ranked as 1, 3, 2 depending on the RMF value for the query in question.

Consider a query “Find video sequences in which a blue car is going towards North”. Table 4 shows the results for this query. The temporal intervals representing the number of frames for which a spatial relationship *blue car is going towards north* is valid for every sequence and the RMF values are given in the Table 4. The video sequences in the database are ranked depending on the RMF values. For the above query the most similar video sequence is speedwa2.mpg, which has the maximum number of frames i.e. 1132 that satisfy the spatial relationship mentioned in the query. The ranking of the database video sequences for this query is 2, 3, 4, 1.

Consider another query “Find video sequences in which red car is standing right side of the road”. The results for this query are shown in Table 5. The most similar video sequence for this query in the database is speedwa5.mpg and the ranking obtained for the database video sequences is 5, 4, and 3.

## 5. Conclusions

We have presented a fuzzy spatio-temporal model for video content description that supports spatio-temporal queries. The proposed model is based on fuzzy directional and topological relations and fuzzy temporal relation. The problems associated with the use of crisp spatio-temporal relations were highlighted. It is shown that errors may occur in segmentation, event detection and object detection due to noisy video data and use of precise spatial relations. In order to minimize these errors fuzzy definitions of temporal relations are proposed. In addition, the second order temporal relations are presented that are more informative and provide global information about the sequence. The proposed model provides a mechanism that represents the fuzzy spatio-temporal relationships among the objects in video sequences in the database and ranks the database sequences based on the query for effective content-based retrieval. We reported the results of our experiment on a sample video from MPEG-7 data set.

Video Sequence	Spatial Relation <i>Person 1 in white shirt is left of Person 2</i>	Relevance Membership Function (RMF) *
1 (ETRI_od_A.mpg)	273	$3086/273 = 11.3040$
2 (ETRI_od_B.mpg)	0	$3455/0 = \text{Positive Infinity}$
3 (ETRI_od_C.mpg)	219	$8910/219 = 40.6849$

\* RMF value 1 indicate maximum relevance and positive infity indicate minimum relevance

Table 3. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find The VideoSequences in Which Person1 in White Shirt is Left of Person 2 in Black Shirt" for the Video Sequences in MPEG-7 Video Content Set

Video Sequence	Spatial Relation <i>Blue car is going towards north</i>	Relevance Membership Function (RMF)
1 (speedwa1.mpg)	87	$1420/87 = 16.32$
2 (speedwa2.mpg)	1132	$13185/1132 = 11.64$
3 (speedwa3.mpg)	388	$14019/388=36.13$
4 (spedwa4.mpg)	100	$7496/100=74.96$
5 speedwa5.mpg)	0	$7497/0=\text{Positive Infinity}$

Table 4. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find the Video Sequences in which Blue Car is Going Towards North" for the Video Sequences in the MPEG-7 Video Content Set

Video Sequence	Spatial Relation <i>Red car is standing right side of the road</i>	Relevance Membership Function (RMF)
1 (speedwa1.mpg)	0	$1420/0 = \text{Positive Infinity}$
2 (speedwa2.mpg)	0	$13185/0 = \text{Positive Infinity}$
3 (speedwa3.mpg)	452	$14019/452=31.01$
4 (spedwa4.mpg)	1506	$7496/1506=4.97$
5 speedwa5.mpg)	1834	$7497/1834=4.08$

Table 5. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find the VideoSequences in which Red Car is Standing Right Side of the Road " for the Video Sequences in the MPEG-7Video Content Set

## 6. References

- Allen J.F., (1983). Maintaining knowledge about temporal intervals. Communications of ACM 26 (11), pp. 832-843.
- Bimbo A.D., Vicario E., Zingoni D., (1995). Symbolic description and visual querying of image sequences using spatio-temporal logic. IEEE Trans. Knowledge and Data Engineering 7 (4), pp. 609-621.
- Chang S. F., Chen W., Meng H., Sundaram H., Zhong D., (1998). A fully automated content-based video search engine supporting spatiotemporal queries. IEEE Trans. on Circuits and Systems for Video Technology 8 (5),pp. 602-615.
- Courtney J.D., 1997. Automatic video indexing via object motion analysis. Pattern Recognition 30 (4), pp. 607-625.
- S. Dagtas, A. Ghafoor, (1999). Indexing and retrieval of video based on spatial relation sequences, Proc. ACM International Multimedia Conf. (part 2) Oriando, FL, USA, , pp. 119-121.
- Dagtas S., (1998). Spatio-Temporal Content-Characterization and Retrieval in Multimedia Databases, Ph.D. Thesis, Purdue University.
- Deng Y., Manjunath B.S., (1998). NeTra-V: Toward an object-Based video representation. IEEE Trans. Circuits and Systems for Video technology 8 (5), pp. 616-627.
- Day Y.F., Dagtas S., Lino M., Khokhar A., and Ghafoor A., 1995. Object-oriented conceptual modeling of video data. Proc. Eleventh Int. Conf. on Data Engineering, Taipei, Taiwan, pp.401-408,.
- Egenhofer M.J., Fanzosa R., (1991). Point-set topological spatial relations. International Journal on Geographic Information Systems 5 (2), pp.161-174.
- El-Kwae E., Kabuka M. R., (1999). A robust framework for content-based retrieval by spatial similarity in image databases. ACM Trans. on Information Systems 17( 2), pp.174-198.



- Gudivada V., Raghvan V., (1995). Design and evaluation of algorithms for image retrieval by spatial similarity. *ACM Trans. on Information Systems* 13 (2), pp.115-144.
- Hampapur A., Gupta A., Horowitz B., Shu C.F., Fuller C., Bach J., Gorkani M., Jain R., (1997). Virage Video Engine. *Proc. SPIE Storage and retrieval for Image and Video Databases V*, San Jose, pp.188-197.
- Little T.D.C., Ghafoor A., (1993). Interval-based conceptual models for time dependent multimedia data. *IEEE Trans. Knowledge and Data Engineering* 5 (4), pp.551-563.
- Li J. Z., Ozsu M. T., Szafron D., (1997). Modeling of video spatial relationships in an object database management system. *Proc. IS & T/ SPIE Int'l Symposium on Electronic Imaging: Multimedia Computing and Networking*, San Jose, USA, pp.81-90.
- Li J. Z., Ozsu M. T., Szafron D., Oria V., (1997). MOQL: A multimedia object query language. *Proc. Third International Workshop on Multimedia Information Systems*, Italy, pp. 19-28.
- Mottaleb M.A., Dimitrova N., Desai R., Martino J., (1996). CONIVAS: Content-based image and video access system. *Proc. ACM Multimedia Conf.*, Boston, MA USA, pp. 427-428.
- Matsakis P., Wendling L., (1999). A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (7), pp. 634-643.
- Miyajima K., Ralescu A., (1994). Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems* 65, pp. 225-236.
- Nabil M., Ngu A., Shepherd J., (1996). Picture similarity retrieval using the 2D projection interval representation. *IEEE Trans. on Knowledge and Data Engineering* 8 (4), pp. 533-539.
- Nepal S., Srinivasan U., (2002). Spatio-temporal modeling and querying video databases using high level concepts. *Proc. Sixth Working Conference on Visual Database System (VDB 6)*, 29-31 May, Brisbane, Australia .
- Oomoto E., Tanaka K., (1993). OVID: Design and implementation of a video-object database system. *IEEE Trans. on Knowledge and Data Engineering* 5 (4), pp. 629-643.
- Petkovik M., Jonker W., (2001). Content-based video retrieval by integrating spatio-temporal and stochastic recognition of events. *Proc. IEEE workshop on Detection and Recognition of Events in Video*, Vancouver, Canada.
- Pissiou N., Radev I., Makki K., Campbell W.J.,(2001). Spatio-temporal composition of video objects: representation and querying in video database systems. *IEEE Trans. on Knowledge and Data Engineering* 13 (6), pp. 1033-1040.
- Rajurkar A. M., Joshi R.C., (2001). Content-Based Image Retrieval: A fuzzy spatial similarity approach. *Proc. of the International Symposium on Artificial Intelligence, ISAI'2001*, Fort Panhala (Kolhapur), INDIA, Dec pp. 18-20.
- Vazirgiannis M., Theodoridis Y., Sellis T., (1998). Spatio-temporal composition and indexing for large multimedia applications. *Multimedia Systems* 6, pp. 284-298.
- Zhang H.J., Wu J., Zhong D., Smoliar S.W., (1997). An integrated system for content-based video retrieval and browsing, *Pattern Recognition* 30 (4), pp. 643-658.

- Zhang H.J., Kankanhalli A., Smoliar S.W., (1993). Automatic partitioning of full-motion video. *Multimedia Systems* 1 (1), pp. 10-28.
- Zhang H.J., Low C.Y., Smoliar S.W., Wu J.H., (1995). Video parsing, retrieval and browsing: an integrated and content-based solution. *Proc. of the ACM Multimedia Conf.*, San Francisco, CA, November 5-9, pp. 15-24.