SPATIO-TEMPORAL MODELLING AND QUERYING VIDEO DATABASES USING HIGH-LEVEL CONCEPTS

Surya Nepal and Uma Srinivasan
CSIRO Mathematical and Information Sciences,
Locked Bag 17, North Ryde NSW 1670, Australia
{Surya.Nepal,Uma.Srinivasan}@cmis.csiro.au

Abstract: Modelling and querying video content involves understanding the spatio-temporal relationships of objects[1] represented in a video sequence. In this paper, we present a binary representation-based framework for modelling and querying video object relationships at a semantic level. The framework supports spatio-temporal queries using high-level concepts. We propose a set of semantic (qualitative) terms to describe the spatio-temporal relations between objects and then show how such qualitative terms can be calculated within the proposed framework using a fuzzy-logic-based approach. We also report the experimental results performed on a video clip from the MPEG-7 video content archive.

Key words: content-based retrieval, spatio-temporal relations, multimedia databases.

1. INTRODUCTION

The growing number of video databases has placed content-based modelling and search tools in high demand. A number of commercial products as well as academic prototype systems have been developed to facilitate content-based search. Prominent examples are QBIC (Flickner 1995), and Virage (Hampapur et al. 1997). These systems generally extract features from raw images/videos and use them for retrieval. Features such as

[1] When using to the terms ***objects or video objects,*** we are referring to objects or region(s) of features that are identified in a video sequence.

X. Zhou et al. (eds.), *Visual and Multimedia Information Management*

colour, texture, shape, motion trajectories, and edges are commonly used. Many modelling and querying techniques have been developed and used for emerging features (Baral, Gonzalez and Son 1998). Initially, most of these modelling and querying techniques were developed for image databases. More recently, multimedia applications have started paying attention to spatio-temporal modelling and querying video content (Bolle et al 1997, Sistla et al. 1997, Hjelsvold and Midstraum 1994). There have been a few techniques for modelling and querying multimedia databases based on spatio-temporal relations (Li et al 1997, Kuo and Chen 2000). These approaches are motivated from the low-level feature-based techniques. In this paper, our aim is to present a framework that enables users to query video content using high-level concepts. These high-level concepts are described using spatio-temporal relations.

Most feature-based systems index content using extracted features, where users need to be familiar with the underlying feature representation schemes to express their queries. We call this a "bottom up approach", as users need to know how to express their queries in terms of low-level features. Our framework is based on "top-down" approach, where users express their queries using high-level concepts that are more natural. The "concepts" we defined are based on spatio-temporal relations at the frame and shot levels.

In section 2, we review various spatio-temporal relations used in geographical information systems and provide a set of semantic terms to describe spatio-temporal relations in video databases. In section 3, we use binary representation-based modelling techniques to model spatio-temporal relations at different levels. Using an object detection algorithm, we illustrate some fundamental relationships and provide experimental results. These are reported in Section 4.

2. BACKGROUND

We present background work in the areas of spatial and temporal relations and provide a set of semantic terms that are useful for expressing spatio-temporal queries in video databases

2.1 Topological relations

Topological relations are spatial relations that are invariant under bijective and continuous transformations that also have inverses. Topological equivalence does not necessarily preserve distances and directions. Instead, topological notions include continuity, interior, and boundary, which are defined in terms of neighbourhood relations. Egenhofer and Franzosa (1991) have specified eight fundamental topological relations that can be applied between two planar regions. These relations are used in geographical information systems and spatial databases (Papadias and Delis 1997, Delis, Christos and Spiros 1999). The eight topological relations are *Disjoint, Meet, Equal, Inside, Contains, Covered_By, Covers,* and *Overlap.*

2.2 Directional relations

Directional relations are used to define the order of objects in the space. Frank (1996) has defined the eight directional relations based on a qualitative approach. These relations are studied under two different groups: strict directional relations such as N, S, W and E, and mixed directional relations such as NE, SE, SW and NW. Qualitative approaches are highly desirable in similarity-based retrieval systems (such as image and video databases), because precision is not always desirable nor available. The eight directional relations are North, NorthEast, East, SouthEast, South, SouthWest, West, and NorthWest.

2.3 Temporal relations

As a video is a continuous medium, temporal interval relations provide important cues for video data retrieval. Allen (1983) has defined 13 temporal interval relations. These are *Before, Meets, Overlaps, Finishes, Starts, Contains, Equals, During, Started by, Finished by, Overlapped by, Met by,* and *After.* Many variations of Allen's temporal interval relations have been proposed and used in temporal and multimedia databases. These relations are used to define the temporal relationships of events within a multimedia document. For example, weather news *after* the sports news in a daily TV news broadcast.

A partial taxonomy of spatio-temporal relations for video databases is shown in Figure 1. The list of conceptual terms and relationships are not limited to the ones provided in the figure. We have shown spatio-temporal relations and concepts that are of interest in this paper. We categorize spatio-temporal relations into two categories based on the number of objects involved: unary and binary relations. For a single object, unary spatio-temporal relations describe location of the object, motion of the object, etc. In binary relations, our focus is on the relative positions and movements of related objects. We next describe the framework and show how it is used in mapping semantic terms to spatio-temporal positions.

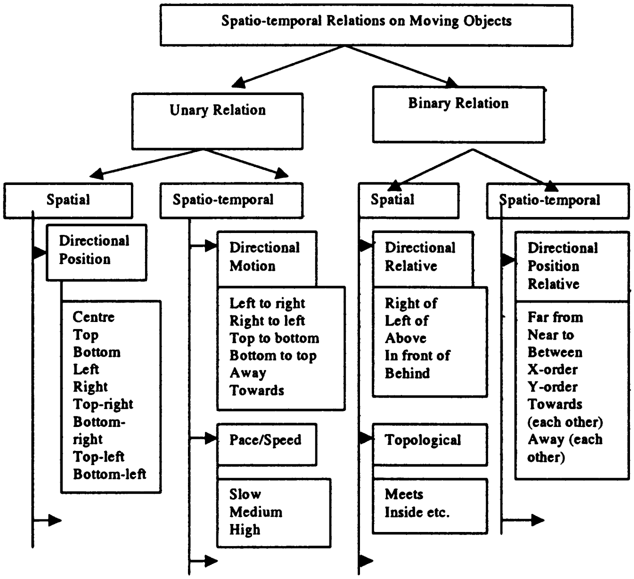


Figure 1: A few semantics describing spatio-temporal relations in video databases.

3. PROPOSED FRAMEWORK

3.1 Binary representation scheme

In this section, we describe a binary representation scheme for positions as defined in (Delis, Christos and Spiros 1999) for spatial reasoning (they are also called relations in literature). We first describe how spatial positions are represented using 1D binary string. The positions are defined using reference points. The reference points partition the given interval into a number of smaller intervals. Each partitioned interval is represented by a single bit. Thus, the number of bits required to represent positions depends on the reference points. A single reference point within a given interval yields two positions: "*left*" and "*right*". We could have infinite number of reference points that yield infinite number of positions. The question is how do we determine the number of reference points. There are no standard rules. We have considered two factors: high-level concepts and fuzzy concepts. The first approximation is that we should be able to describe each region (or partition) using semantic terms. In our case, we have considered nine semantic terms to describe positional relations. So based on the first approximation, we map nine terms (to be described later) to nine partitions shown in Figure 3. There does not exist distinct boundary between two concepts. In order to represent this fuzziness we introduce buffer (fuzzy) regions in between two adjacent semantic regions. The number of buffer regions depends on the level of preciseness (or the granularity of fuzzy membership values). In our case, we introduce one fuzzy region in between two adjacent semantic regions.

Let n be the number of reference points within a given interval. Then we need n+1 bits to represent positions. Let AB = $<X_1, X_2 .. ,X_n>$ be partitions of given interval (a,b) . Given an interval X, let X.l and X.r represent its left and right end points. X.l and X.r determine the boundaries for calculating the binary string.

Definition (Positional Binary String): *Assume a reference interval X. A resolution scheme AB (X) = $<X_1, X_2, .., X_n>$ is a partition of X by (n-1) reference points where a position of an interval (x1,x2) over AB is a binary string $t_1,t_2,..,t_n$ such that t_i belongs to {0,1}, t_i="1" iff $X_i \cap [x_1, x_2] \neq$ Null and t_i = "0" otherwise, i=1,..,n, complying with the following constraints (i) it has exactly one substring of consecutive "1"'s and (ii) there is at least one t_i = "1" in a position "i" such that X_i is a non-zero length interval.*

The positional binary string is defined for a single object. That is, the positional binary string is calculated for each object rather than each image/frame. Thus, a calculated binary string provides a spatial position of an object in a particular dimension.

Definition (Spatio-temporal Positions) *The spatio-temporal relation of an object O consists of spatial and temporal positions $O<P_x, P_y, P_t>$, where $<P_x, P_y>$ gives the spatial positions and P_t the temporal position. $Pos(O<t>) = <P_x, P_y>$ represents the spatial position of an object O at time t.*

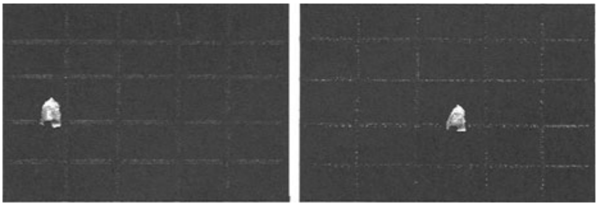


Figure 2: The figure shows objects identified in the frames at time t_1 and t_2.

Figure 2 shows objects identified in a video at time t_1 and t_2, respectively. The frame is partitioned into 25 different parts where each part represents a spatial position significant to human perception. We will elaborate about these partitions. The spatio-temporal relations of an object shown in the figure are represented as follows.

Pos(O<t_1>) = <10000, 11100>, Pos(O<t_2>) = <00100, 11100>

Being raw data these binary strings neither help users to express their queries nor are meaningful perceptually. In the later section, we describe techniques that map high-level concepts that are used to express spatio-temporal relations to these binary strings.

3.2 Similarity measurement

Video/image retrieval systems use similarity-based search which is usually often based on perceptual features. The similarity is measured using distance functions over feature values. In our scheme, spatio-temporal relations are expressed using binary strings as defined above. The question is how do we measure the similarity of binary strings. Our similarity measure is based on fuzzy logic. We first define some functions that are needed to define similarity measures.

rs(X) : this is a right shift function which returns the position of the rightmost "1" in a binary string X.

ls(X) : this is a left shift function which returns the position of the leftmost "1" in a binary string X.

Let *QX* and *QY* be the query binary strings representing positions in *X* and *Y* directions, respectively. Let *OX* and *OY* be the binary strings of an object in the database. The distance between these two are then defined as follows.

Dist(Q,O)=abs(rs(QX)–rs(OX)) + abs(rs(QY)-rs(OY)) + abs(ls(QX)-ls(OX)) + abs(ls(QY)-ls(OY))

The distance function is a combination of four different movements: distance in right movement (abs(rs(QX)–rs(OX))), distance in left movement (abs(ls(QX)-ls(OX))) , distance in upward movement (abs(rs(QY)-rs(OY))), and distance in downward movement (abs(ls(QY)-ls(OY))). In most of the circumstances, the right and upward movement would be equal and opposite to left and downward movement.

The similarity value of the query binary string Q with any database binary string O in fuzzy logic is given by (Nepal, Ramakrishna and Thom 1998)

$$\mu_Q(O) = \frac{(Max - Dist(O, R))}{Max}$$

where Max is the maximum possible distance between the two binary strings representing spatio-temporal positions, and $\mu_Q(O)$ is the similarity of a database binary string O with the query string Q or the membership value of the binary string O in a fuzzy set Q.

3.3 Spatio-temporal queries

In this section we describe how we can map the spatio-temporal relations defined in Section 2 to the positional strings given in binary representation.

3.3.1 Unary relations: spatial: directional Positions

The directional relations such as *North* and *South* discussed in Section 2 are used for spatial reasoning in large-scale spaces, i.e., spaces that cannot be seen or understood from a single point of view. Such relations are useful in application areas such as Geographical Information Systems. In case of still images and video frames, we mainly deal with small-scale spaces, i.e., spaces that can be seen or understood from a single point of view. Nextwe describe our qualitative (concept-based) spatial reasoning approach that can be used in multimedia databases.

In order to support such spatial reasoning, we provide a set of positional relations. We classify positional relations into two groups: (1) absolute and (2) relative. In unary directional positions, we only consider absolute positional relations. We define nine absolute positional spatial relations and their corresponding representations in binary strings are shown in Table 1 and Figure 3.

Relations	Symbol	QX	QY
Top	T	00100	00001
TopRight	TR	00001	00001
Right	R	00001	00100
BottomRight	BR	00001	10000
Bottom	B	00100	10000
BottomLeft	BL	10000	10000
Left	L	10000	00100
TopLeft	TL	10000	00001
Centre	C	00100	00100

Figure 1: Nine basic positional relations and their corresponding symbols, and X and Y binary strings of relations.

Figure 3 shows the partition of spatial space into semantic and fuzzy terms. The white regions correspond to specific regions labelled with appropriate semantic terms. Since the boundary between two semantic regions is not clear, we introduce a fuzzy region. We represent fuzzy regions by shaded areas in the figure. The number of fuzzy regions increases the possible membership values for an object to be at certain semantic position (i.e., the granularity of fuzzy values increases). The possible membership values with the partition shown in figure are 1, 0.75, 0.5, 0.25 and 0. If we introduce two fuzzy regions between adjacent semantic regions, then the possible membership values would be 1, 0.85, 0.68, 0.51, 0.34, 0.17 and 0.

Computation: If a query contains directional position relations such as Top and Bottom, the query string is generated as shown in Table 1. The database strings are then compared with the query string using the similarity measure defined above. For example, the similarity value for a query "Center" is 1 if an object appears at the region specified by "C" in the Figure 7. If the object appears in the shaded region between "C" and "B", then the similarity value would be 0.5. The qualitative semantic terms and their corresponding binary representation scheme are shown in Table 6.

Consider a query Q = Bottom-left (logo), where Bottom-left is a positional spatial qualitative term and logo is a visual feature (or object). This query is then converted to a binary form as Q = <QX, QY> = <00001, 10000>. Consider a database position string where the logo appears at the center is represented as O = <OX, OY> = <00100, 00100>. The similarity of O with Q is then calculated as follows.

rs(QX) = 1, ls(QX) = 5, rs(QY) = 5, ls(QY) = 1
rs(OX) = 3, ls(OX) = 3, rs(OY) = 3, ls(OY) = 3
Dist(O, R) = 2 + 2 + 2 + 2 = 8
Max = 4 + 4 + 4 + 4 = 16
The similarity of Q with X is then given by $\mu_Q(R)$ = (16-8)/16 = 0.5.

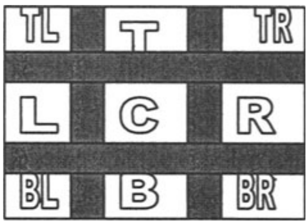


Figure 3: Nine positional relations in a two dimensional space

3.3.2 Unary relations: spatio-temporal: directional

Moving object is an important cue for video analysis and retrieval (Li, Ozsu and Szafron 1997, Sistla et al. 1997). Many applications such as video surveillance have used techniques that deal with moving object detection (Cedras and Shah 1995). The movement of a object can be described in terms of direction of movements. Next we develop a set of semantic query terms to describe the movement of objects and evaluate them under our binary representation scheme. The semantic terms defined here are applicable at the shot level.

Definition (spatial-factor): *this is the factor by which an object changes its spatial position over the duration of that shot. This change could either be induced by camera motion or by the object's intrinsic motion.*

We use this spatial-factor to compute various concepts and terms. We first compute the horizontal and vertical movements: left to right, right to left, top to bottom and bottom to top. Let OX_1 and OX_n be the X binary positional strings of the object O in the first and last frames of the shot. Let OY_1 and OY_n be the Y binary positional strings of the object in the first and last frames of the shot. Then the distances of right, left, top and bottom movements are calculated as follows.

Let the Right movement $D_1 = rs(OX_1) - rs(OX_n)$, Left Movement $D_2 = ls(OX_1) - ls(OX_n)$, Upward Movement $D_3 = rs(OY_1)-rs(OY_n)$, Downward Movement $D_4 = ls(OY_1)-ls(OY_n)$, Sim = the similarity value for the query, w_1 = weight assigned to the horizontal movement, w_2 = weight assigned to the vertical movement, D_1Max = maximum right movement, D_2Max = maximum left movement, D_3Max = maximum upward vertical movement, D_4Max = maximum downward vertical movement.

Consider a query Q = "left-to-right" that requests for objects moving from left to right. D_1 plays an important role in the right movement. Here, we consider that if an object does not have right movement ($D_1 < 1$), then its fuzzy membership value to the result set is 0. If it has a right movement (if $D_1 >= 1$), then the membership value is given by $\mu_Q(R)$.

$$\mu_Q(R) = \frac{abs(w_1.D_1 - w_2\dfrac{(abs(D_3) + abs(D_4))}{2})}{(w_1.D_1Max + w_2\dfrac{D_3Max + D_4Max}{2})}$$

There are various types of left to right movements. An object moving from left to right in a straight line (that is, without having any vertical motion) should have greater membership value than one that has vertical motion. We take this fact into account by considering D_3 and D_4 while evaluating the above membership function. The total effect of horizontal and vertical movements is controlled by their corresponding weights. That is, if w_2 is 0, then the membership function depends only on the total horizontal movement. Similarly, we calculate other directional unary relations such as "Right-to-left", "Bottom-to-top" and "Top-to-bottom" (Nepal and Srinivasan 2001). Next we demonstrate the capability of the proposed framework in representing and querying moving objects.

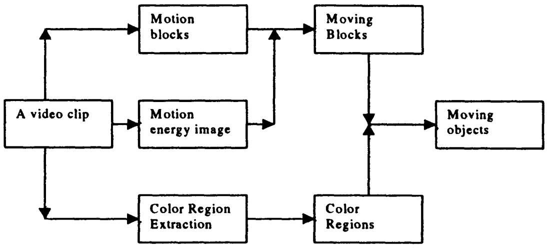


Figure 4: A diagram showing our approach of detecting/capturing moving objects.

4. EXPERIMENTAL RESULTS

4.1 Data set

We used MPEG-7 video content (etri_od_b.mpg) to evaluate the effectiveness of our approach. We identified 17 shots within the video. This can be done using one of the many shot detection algorithms. The first, middle and last frames of the shots are given in the appendix.

4.2 Moving Object detection algorithm

The block diagram of our approach for moving object detection is shown in Figure 4. While moving object detection is not the focus of this paper, this section is introduced to demonstrate how moving objects are modelled in this framework. The pre-processing stage uses a combination of visual features (such as colour and motion vectors) to generate regions for colours and motions and then combine them to obtain the moving objects present in a video clip. The algorithm detects moving object in a stationary background.

The colour regions extracted using 13 perceptually meaningful colours (Carson and Ogle 1996). We use the notion of uniform regions in images and treat each region as an object present in the image. Motion information in a MPEG video is based on 16 x 16 blocks (Gall 1995). We use such motion information to extract motion regions. Let C_{reg} and the M_{reg} be the sets colour and motion regions (or blocks), respectively. The object regions O_{reg} are then given by,

$$O_{reg} = | M_{reg} \cup C_{reg} |$$

where, $C_{reg} = \{C_1, C_2, \ldots, C_m\}$ where m is the number of color regions, $M_{reg} = \{M_1, M_2, \ldots, M_n\}$ where n is the number of motion regions. The moving object is then given by

$$MO_{reg} = \{C_i \mid th > 0.7 \wedge th = \frac{|M_j \cap C_i|}{|C_i|} > 0.7 : i = 1,..,m : j = 1,..n\}$$

where *th* is a threshold value. In our experiments, we calculate the threshold value in such a way that 70% of the colour region has motion, for it to be considered it as a region belonging to a moving object.

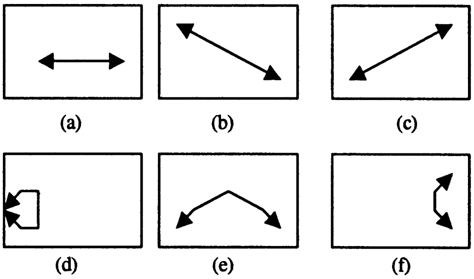


Figure 5: A set of example directions of object movements.

4.3 Results

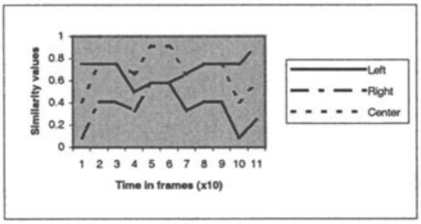
We used Figure 3 as the basis for binary representation for each moving object. The shot number, frame number and their binary representation in both *X* and *Y* dimensions are stored in the database. A set of sample shots with first, middle and last frames are shown in Appendix A. In order to keep it simple, we discarded shots such as shot 3 having more than one object.

We posed a set of directional queries on our experimental data set. A summary of the query results is shown in Table 2. We explain the results in terms of a set of example directional object movements shown in Figure 5. Shots 1, 2, 5, 6, 7, 9, 10, 11, 12 15, 16 and 17 have movements of type Figure 9(a), either the object is moving from left to right or right to left. Shot 3 has multiple objects that we have not considered in this experiment. Shots 4 and 13 have movements of types Figures 9(c) and 9(e), respectively. As can be seen moving objects have both right to left and top to bottom motions. Shot 8 has a motion of type Figure 9(d) with no vertical movements. That is, our current modelling and querying framework does not recognize this kind of movements. This is a limitation of our framework. We plan to improve our query processing techniques in order to recognize such object movements. Our object recognition algorithm could not identify the motion of objects on shot 14. This is due to the fast movement of the object that gives false motion to background objects as well. This is a limitation of our object recognition algorithm rather than the modelling and querying framework presented in this paper. In our experiment, we discarded the effects of vertical movements on horizontal movements and vice versa by setting the weights to 0. We are further investigating the effects of weights and how to optimise such weights to get better similarity values.

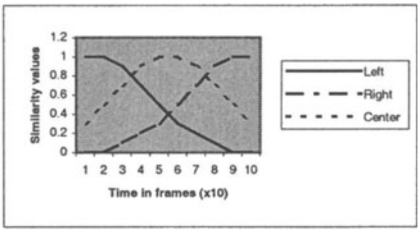
Shot No	Left to Right $W_1 = 1$ $w_2 = 0$	Right to Left $w_1 = 1$ $w_2 = 0$	Top to Bottom $w_2 = 1$ $w_1 = 0$	Bottom to Top $w_2 = 1$ $w_1 = 0$
1	0.8	0.0	0.0	0.0
2	0.0	0.6	0.0	0.0
3	NA	NA	NA	NA
4	0.0	0.8	0.6	0.0
5	0.0	0.8	0.0	0.0
6	0.8	0.0	0.0	0.0
7	0.6	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.6	0.0	0.0	0.0
10	0.0	0.8	0.0	0.0
11	0.0	0.8	0.0	0.0
12	0.8	0.0	0.0	0.0
13	0.0	0.4	0.6	0.0
14	0.0	0.0	0.0	0.0
15	0.8	0.0	0.0	0.0
16	0.8	0.0	0.0	0.0
17	0.8	0.0	0.0	0.0

Table 2: A summary of shot level directional queries on the experimental data set.

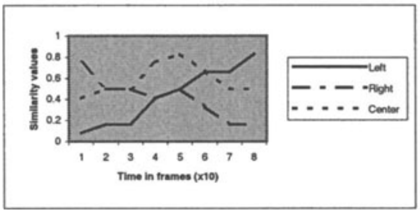
We also posed a set of directional positional queries for all shots. We explain a set of query results here. We posed three directional position queries "left", "right" and "centre" into the database. We recorded the similarity values of each frame from all shots for all three different queries. We choose three shots to explain the results as shown in Figure 6. Figure 6(a) shows the similarity values of each frame in shot 1, where the object is moving from left to right. When an object is having smooth left to right movement, the ideal result would be the one shown in Figure 6(b). Similarly, Figures 6(c) shows the result for the shot 5, where the object is moving from right to left.



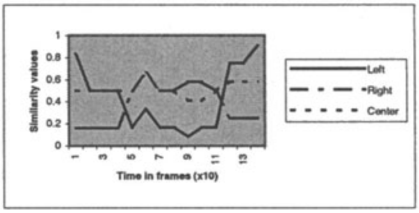
(a)



(b)



(c)



(d)

Figure 11: A set of results for the positional semantic terms left, right and center.

Figure 6 (d) shows the results of three positional queries for each frame in shot 8. We have seen in Figure 10 that shot 8 has similarity values 0.0 for all movement queries like left to right. As we can see in the figure, the similarity values for left and right queries for the first and last frames are almost the same. That means the position of the object in the first and last frame is the same though the position of the object is changed within the shot. This and other similar questions related with movements need to be answered; for example are they important to users? What kinds of terms do people use to express such movements in real life? And how do we evaluate such queries? We leave these queries for our future research.

5. CONCLUSIONS

In this paper we have presented a framework for modelling and querying video databases using spatio-temporal relations. We provided a set of semantic terms that can be used to describe unary and binary spatio-temporal relations. However, our focus in this paper was on a set of unary relations. We proposed a binary string based representation scheme for positions, which offers an efficient technique for mapping high-level semantic terms to low-level feature. We presented a fuzzy logic based approach for computing similarity of spatial positional and spatio-temporal directional semantic terms. We reported the results of our experiment on a sample video from the MPEG-7 data set. Our future work involves mapping other high-level semantics to binary spatio-temporal relations in order to support concept-based querying.

REFERENCES

ALLEN, J.F (1983) Maintaining Knowledge about Temporal Intervals. *Communication of the ACM*. 1983. pp 832-843.

BARAL, C., GONZALEZ, G., AND SON, T (1998). Conceptual Modeling and Querying in Multimedia Databases. *Multimedia Tools and Applications, Vol 7*, pp 37-66, 1998.

BOLLE, R.M., YEO, B.L., AND YEUNG M. (1997) Video Query and Retrieval, in SATTAR, A. (Ed.) *Advanced Topics in Artificial Intelligence,* Lecture Notes in Artificial Intelligence 1342, pp 13-24, 1997.

CARSON, C. AND OGLE, V.E. (1996). Storage and retrieval of feature data for a very large online image collection. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Volume 19, Number 4, Pages 19-27,* December 1996.

CEDRAS, C. AND SHAH, M. (1995) Motion-Based Recognition: A Survey. *Image and Vision Computing,* Volume 13 Number 2 March 1995 pages 129-155.

DELIS, V., CHRISTOS, M. AND SPIROS, S. (1999) A Provably Efficient Computational Model for Approximate Spatiotemporal Retrieval. *ACM GIS* 1999. pp. 40-46.

EGENHOFER, M. and FRANZOSA, R. (1991). Point-set topological spatial relations. International Journal of Geographic Information Systems, 5(2):161—174, 1991

FLICKNER, M. , SAWHNEV, S., NIBLACK, W., ASHLEY, J., HUANG, O., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., and STEELE, D. (1995) Query by Image and Video Content: The QBIC System. Computer. Vol 28, No. 9, 1995, pages 23-32.

FRANK, A.U. (1996) Qualitative Spatial Reasoning: Cardinal Directions as an Example. *International Journal of Geographic Information Systems*, Vol. 10 (3), 269-290

GALL L. (1995) MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM* Volume 34 Number 2 April 1995 pages 47-58, 1996.

HAMPAPUR, A. et al. (1997) Virage Video Engine. *SPIE Vol. 3022*. pp 188-198, 1997.

HJELSVOLD, R. AND MIDTSTRAUM, R. (1994) Modelling and Querying Video Data. *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994.

KUO, T.C.T. AND CHEN A.L.P. (2000) Content-Based Query Processing for Video Databases. *IEEE Transaction on Multimedia Vol 2* No 1, March 2000.

LI, J.Z., OZSU, M. T., AND SZAFRON, D. (1997) Modeling of Moving Objects in a Video Database. ICMCS, 1997, pages 336-343.

NEPAL, S., RAMAKRISHNA, M.V., and THOM, J.A.(1998) A Fuzzy System for Content Based Image Retrieval. *2nd IEEE International Conference on Intelligent Processing Systems* (IEEE ICISP98), Gold Coast, Australia, August 4-7, 1998. Pages 335-339

NEPAL, S., AND SRINIVSAN, U. (2001). Spatio-temporal modelling and querying video databases using high-level concepts, CSIRO Mathematical and Information Sciences, Technical Report Number 2001/66, 2001.

PAPADIAS, D. AND DELIS, V. (1997) "Relation-Based Similarity", In *Proceedings of the 5th ACM Workshop on GIS*, Las Vegas, ACM Press, 1997.

SISTLA, A. P., WOLFSON, O., CHAMBERLAIN, S., AND DAO, S. (1997) Modeling and Querying Moving Objects, ICDE 1997, pages 422-432.

Appendix a