Spatio-temporal Representation and Retrieval Using Moving Object's Trajectories

Choon-Bo Shim
Dept. of Computer Engineering
Chonbuk National University
Chonju, Chonbuk 560-756, South Korea
Tel : +82-63-270-2414
E-mail:cbsim@dblab.chonbuk.ac.kr

Jae-Woo Chang
Dept. of Computer Engineering
Chonbuk National University
Chonju, Chonbuk 560-756, South Korea
Tel : +82-63-270-2414
E-mail:jwchang@dblab.chonbuk.ac.kr

ABSTRACT

In this paper, we propose a new spatio-temporal representation scheme using moving objects' trajectories in video data. In order to support content-based retrieval on video data very well, our representation scheme considers the moving distance of an object during a given time interval as well as its temporal and spatial relations. Based on our representation scheme, we present a new similarity measure algorithms for the trajectory of moving objects, which provides ranking for the retrieved video results. Finally, we show from our experiment that our representation scheme achieves about 20% higher precision while holding about the same recall, compared with Li's and Shan's schemes.

Keywords

moving object's trajectory, spatio-temporal representation, video database.

1. INTRODUCTION

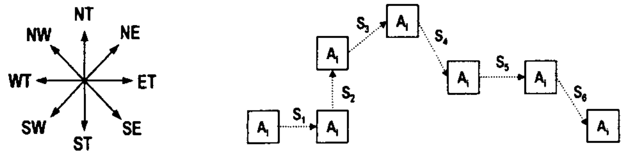
Because a very large number of multimedia data are stored and dealt with in many applications, such as digital libraries, advertisements, video on demand (VOD), digital broadcasting, and electronic commerce, content-based retrieval on multimedia data themselves [1] has recently become a critical research area of multimedia database systems. In video data, the trajectory of a moving object plays an important role in video indexing for content-based retrieval. The trajectory can be represented as a spatio-temporal relationship between moving objects, including both their spatial and temporal properties [2]. User queries based on the spatio-temporal relationship are as follows: "*Finds all objects whose motion trajectory is similar to the trajectory shown in a user interface*" or "*Finds all shots with a scene that two cars approach to each other.*"

To handle the queries, there have been many studies on temporal relationship and spatial relationship between moving objects in video data. The studies on the temporal relationship are based on thirteen temporal relations proposed by Allen [3] while those on the spatial relationship [4] are based on topological and directional relations using spatial coordinates. While most of the studies have concentrated on spatio-temporal relationships between moving objects, they did not consider their moving distance during a given time interval. For modeling moving objects' trajectories in video data, it is necessary to consider their moving distance as well as their spatio-temporal relationships so as to support content-based retrieval on video data very well. For example, in case of soccer game video data, it is very important to decide whether the trajectory of a soccer ball belongs to a long pass or a short pass. This can be determined by using the moving distance of the ball in a given time interval, rather than using the direction of it.

In this paper, we propose a new spatio-temporal representation scheme using moving objects' trajectories in video data. In order to support content-based retrieval on video data very well, our representation scheme considers the moving distance of an object during a given time interval as well as its temporal and spatial relations. Based on our representation scheme, we present a new similarity measure algorithms for moving objects' trajectories, called SDST. This paper is organized as follows. In Section 2, we introduce related work in the area of content-based video retrieval using spatio-temporal relationships. In Section 3, we propose a new spatio-temporal representation scheme using moving objects' trajectories and describe a new similarity measure algorithm to calculate the similarity between a user query and moving objects in video databases. In Section 4, we compare the performance of our scheme with those of the Li's and Shan's schemes. Finally, we draw our conclusions and suggest future work in Section 5.

2. RELATED WORK

There have been some researches on content-based video retrieval using spatio-temporal relationships in video data. First, when assuming a moving object is a salient one moving over time, John Z. Li et al. [5] represented the trajectory of a moving object as eight directions in Figure 1(a), such as North(NT), Northwest(NW), Northeast(NE), West(WT), Southwest(SW), East(ET), Southeast(SE), and Southwest(SW). They represented as (S_i, d_i, I_i) the trajectory of a moving object A over a given time interval I_i where S_i is the displacement of A and d_i is a direction. For a set of time interval <I_1, I_2, ..., I_5>, the trajectories of A can be represented as a list of motions in Figure 1(b), i.e., < (S_1, ET, I_1) , (S_2, NT, I_2), (S_3, NE, I_3), (S_4, SE, I_4), (S_5, ET, I_5), (S_6, SE, I_6) >.



(a) Eight Moving Directions **(b) Trajectory of Object A**

Figure 1. Moving Direction and Example

Based on the representations for moving objects' trajectories, they present a similarity measures to computes the similarity of spatio-temporal relationships between two moving object. Let $\{M_1, M_2, \ldots, M_m\}$ $(M \geq 1)$ be the trajectory of moving object A, $\{N_1, N_2, \ldots, N_n\}$ be the trajectory of moving object B, and $m \leq n$. The similarity measure between the trajectory of object A and that of object B, TrajSim(A, B), is computed by using the similarity distances of directional relations in Table 1 as follows:

$$\text{minDiff}(A, B) = \text{MIN} \sum_{i=1}^{m} \text{distance}(M_i, N_{j+i}) \qquad (\forall_j\, 0 \leq j \leq n-m)$$

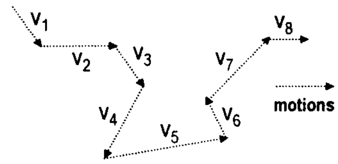
$$\text{TrajSim}(A, B) = \frac{\text{maxDiff}(A,B) - \text{minDiff}(A,B)}{\text{maxDiff}(A,B)}$$

Here, minDiff(A, B) and maxDiff(A, B) are the smallest distance between A and B and the largest distance, respectively. When the moving direction of A is opposite to that of B in all the comparisons, maxDiff(A, B) = 4*m where the maximum number of comparing motions is m. Also, it considered only directional relationship to compute the similarity of a single object's trajectory between video and query.

Table 1. Distances of directional relations

	NT	NW	NE	WT	SW	ET	SE	ST
NT	0	1	1	2	3	2	3	4
NW	1	0	2	1	2	3	4	3
NE	1	2	0	3	4	1	2	3
WT	2	1	3	0	1	4	3	2
SW	3	2	4	1	0	3	2	1
ET	2	3	1	4	3	0	1	2
SE	3	4	2	3	2	1	0	1
ST	4	3	3	2	1	2	1	0

Secondly, Shan and Lee [6] introduced similarity retrieval algorithms for both a single moving object's trajectories in order to support content-based video retrieval. For retrieval based on the single moving object's trajectory, they represented the trajectory of a moving object as a sequence of segments, each being expressed as the slope ranging from 0 to 360 degree. Figure 2 shows the segment of a video trajectory V = (310°, 0°, 310°, 240°, 5°, 95°, 45°, 0°). For the single moving object's trajectory, they proposed an algorithm to measure the similarity between a query trajectory and moving objects' trajectories in video data by using only directional property, called OCM(Optimal Consecutive Mapping).

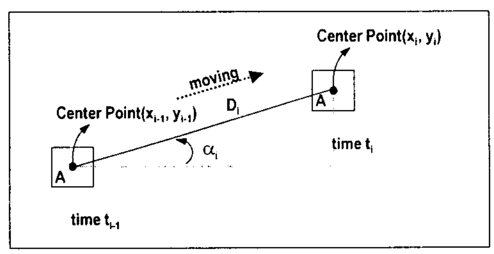


V

Figure 2. Example of moving object's trajectory

3. NEW SPATIO-TEMPORAL REPRESENTATION SCHEME

Both Li's and Shan's schemes concentrated on spatio-temporal relationships between moving objects, but they did not consider their moving distance during a given time interval. In order to support content-based retrieval on video data very well, it is necessary to consider the moving distance of objects in the video data as well as their spatio-temporal relationships. In order to approximate the position of an object, we use MBR (Minimum Bounding Rectangle) as shown in Figure 3. We define a moving object as one whose position is changed over a given time interval and (x_i, y_i) as the center point of object A in XY-coordinates. So, the trajectory of the moving object A is represented as $[(x_0, y_0, t_0), (x_1, y_1, t_1), \ldots, (x_n, y_n, t_n)]$ at time $t_0, t_1, \ldots, t_n$. We define I_i as the difference between a start frame and a finish frame in a set of consecutive video frames. Here, I_i means a time interval between time t_{i-1} and time t_i, $[t_{i-1}, t_i]$. First, the single moving object's trajectory is defined as follows.



Time Interval I_i=[t_{i-1}, t_i]

Figure 3. MBR representation of moving object

Definition 1: Let the motion (M_i) of a moving object A over time interval I_i be (α_i, D_i, I_i). Here, α_i is a direction being represented by the real angle is measured between the center points of moving object at time t_{i-1} and time t_i in counter-clockwise from the X-axis. Its domain is ranging from 0 to 360 degree and D_i is the moving distance of A over I_i which is represented as a relative distance with 0 to 100. For a given order list of time intervals $<I_1, I_2, \ldots, I_n>$, the trajectory of a single moving object A can be described by a list of motions $<M_1, M_2, \ldots, M_n>$, i.e.

$$<(\alpha_1, D_1, I_1), (\alpha_2, D_2, I_2), \ldots, (\alpha_n, D_n, I_n)>.$$

Example 1: Figure 4(a) shows a trajectory of object A. The trajectory for a single moving object A consists of a sequence of motions which is expressed by $<(0°, 15, I_1), (90°, 15, I_2), (40°, 23, I_3), (300°, 32, I_4), (0°, 15, I_5)>$.

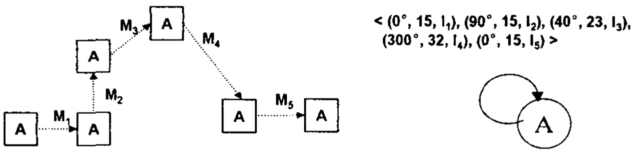


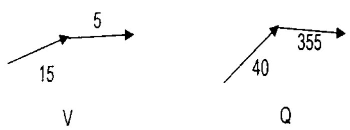
Figure 4. Trajectory and motion list of an object A

Because the moving distance of an object during a time interval plays an important role in calculating the similarity between a user query and moving objects more effectively, we propose a new similarity measure algorithm to consider the moving distance of objects as well as their spatial relationships, called SDST(Similarity measure based on moving Distances for Single object's Trajectories) algorithm. The SDST algorithms allows us to retrieve precise video results based on the moving distance of objects as well as to provide ranking for the retrieved video results to answer a user query.

Definition 2: For a single moving object's trajectory $V=\{V_1, V_2, \ldots, V_M\}$ in video databases and a query trajectory $Q=\{Q_1, Q_2, \ldots, Q_N\}$ ($M \geq N$), the difference between the angle of a video motion V_i and that of a query motion Q_i, $D_{ang}(V_i, Q_i)$, is defined as

$$\begin{aligned}&\text{If } |V_i - Q_i| > 180° \\ &\quad D_{ang}(V_i, Q_i) = (360° - |V_i - Q_i|) \qquad (1)\\ &\text{otherwise} \\ &\quad D_{ang}(V_i, Q_i) = |V_i - Q_i|\end{aligned}$$

Example 2: For a given video trajectory V and query trajectory Q, the difference between the angle of a video motion V_i and that of a query motion Q_i, $D_{ang}(V_i, Q_i)$, is as follows :



$$D_{ang}(V_1, Q_1) = |15° - 40°| = 25°$$
$$D_{ang}(V_2, Q_2) = (360° - |5° - 355°|) = 10°$$

Definition 3: Given a single video moving object's trajectory $V=\{V_1, V_2, \ldots, V_M\}$ and a query trajectory $Q=\{Q_1, Q_2, \ldots, Q_N\}$ ($N \geq 1$), the difference between the direction of a video motion V_i and that of a query motion Q_i, $SR_i(V_i, Q_i)$, is defined as

$$SR_i(V_i, Q_i) = \frac{\cos\left(D_{ang}(V_i, Q_i)\right) + 1}{2} \qquad (2)$$

Definition 4: Given a single video moving object's trajectory $V=\{V_1, V_2, \ldots, V_M\}$ and a query trajectory $Q=\{Q_1, Q_2, \ldots, Q_N\}$ ($N \geq 1$), the difference between the moving distance of a video motion V_i and that of a query motion Q_i, $SD_i(V_i, Q_i)$, is defined as

$$SD_i(V_i, Q_i) = 1 - \frac{\left|D^{V}_{R(i)} - D^{Q}_{R(i)}\right|}{Max\,(D^{V}_{R(i)} - D^{Q}_{R(i)})} \qquad (3)$$

Definition 5: Given a single video moving object's trajectory $V=\{V_1, V_2, \ldots, V_M\}$ and a query trajectory $Q=\{Q_1, Q_2, \ldots, Q_N\}$ ($N \geq 1$), the similarity between a video motion V_i and a query motion Q_i, SDST(V, Q), by using formula (2) and (3) is defined as follows. Here, ω_1 and ω_2 mean the weight of the direction and that of the distance, respectively.

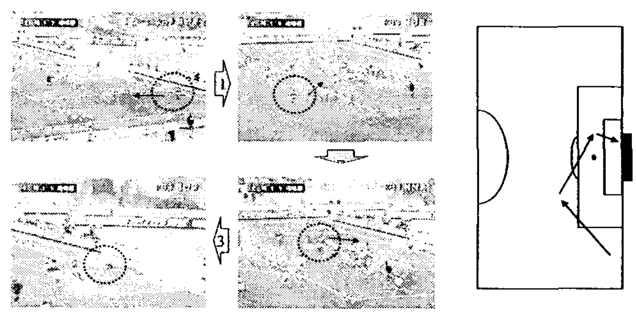
$$SDST(V,Q) = MAX \sum_{j=1}^{M-N+1} \left(\frac{\sum_{i=1}^{N} SR_{i+j}^{(1-w_i)} * SD_{i+j}^{(1-w_i)}}{N} \right)$$

4. Performance Analysis

In order to verify the usefulness of our spatio-temporal representation scheme, we do an performance experiment with the video data of soccer (football) games. Because users generally consider a soccer ball as a salient object in a soccer game video, we extract the trajectories of the soccer ball from the video data. In addition, we make use of the following data.

- For our experiment, we collect 360 video data from soccer game videotapes.
- The trajectory extracted from each video data has 2 to 15 motions.
- For a user query, we generate 40 query trajectories from the video database.
- Each query trajectory includes 2 to 3 motions.

Most of video data used in our experiment which are formatted as MPEG file (*. mpeg) include a shot of 'getting a goal'. We extract the trajectory of a soccer ball by manually tracing the soccer ball in soccer filed. Figure 5 shows an example to extract the trajectory of a soccer ball from the video data.

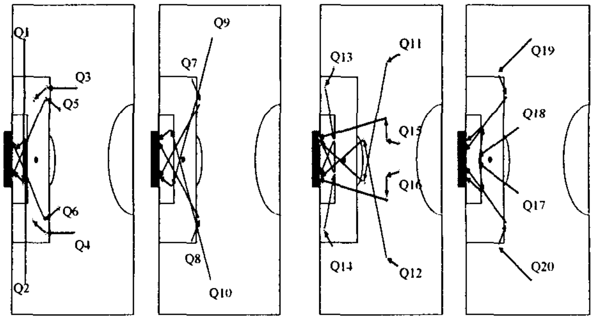


(a) Original Video Frame **(b) Trajectory extracted from (a)**

Figure 5. Example of trajectory

By considering possible 'getting a goal' trajectories from the video database, we make forty query trajectories consisting of twenty in 'the right field' and twenty in 'the left field' from the half line of the soccer field. We classify them into four groups based on the number of motions and the fields as follows. The query trajectories belonging to Type 1 and Type 3 are illustrated in Figure 6(a) and Figure 6(b), respectively.

- Type 1: ten queries with two motions in the right field (Q1 – Q10).
- Type 2: ten queries with two motions in the left field (Q11 – Q20).
- Type 3: ten queries with three motions in the right field (Q21 – Q30).
- Type 4: ten queries with three motions in the left field (Q31 – Q40).



(a) 10 queries belonging to Type 1 (b) queries belonging to Type 3

Figure 6. Each 10 queries belonging to Type 1 and Type 3

For our performance analysis, we implemented our spatio-temporal representation scheme as well as Li's and Shan's schemes under Windows PC with 128 MB memory by using Microsoft-Visual C++. We compare our scheme with the Li's and Shan's schemes in terms of retrieval effectiveness, that is, precision and recall measures [7]. Let RVR (**R**elevant **V**ideo data that are **R**etrieved) be the number of video data retrieved by a given query, RVD (**R**elevant **V**ideo data in **D**atabase) be the number of video data relevant to the query, and RVQ (**R**etrieved **V**ideo data to **Q**uery) be the number of relevant video data retrieved. To compute RVD, we make a test panel which finds relevant video data manually from the database. The test panel is composed of 10 graduate school students from our Computer Engineering department. The precision is defined as the proportion of retrieved video data being relevant and the recall is defined as the proportion of relevant video data being retrieved as follows.

$$\text{Precision} = \frac{RVR}{RVQ} \qquad \text{Recall} = \frac{RVR}{RVD}$$

For our performance comparison, we adopt the 11-point measure which is most widely used for measuring the precision and the recall. Table 2 shows the average precision and the average recall values of our scheme, Li's, and Shan's. Our scheme is shown to be superior to the Li's scheme in terms of both precision and recall. That is, our scheme holds about 20% higher precision and about 10% higher recall. Our scheme also achieves 17% higher precision that the Shan's scheme while it holds about the same recall. Figure 7 shows the recall-precision graph of our scheme, Li's, and John's with the 11-point measure.

Table 2. Comparison of retrieval effectiveness

	Retrieval Effectiveness	
	Average Precision	Average Recall
Li's scheme	0.23	0.42
Shan's scheme	0.26	0.46
Our scheme	**0.43**	**0.44**

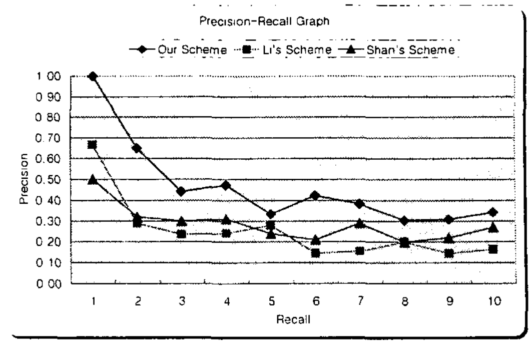


Figure 7. Recall-precision graph

5. CONCLUSIONS

For efficient content-based retrieval on video data, we proposed a new spatio-temporal representation scheme using moving objects' trajectories which considered the moving distance of an object during a given time interval as well as its spatial relations. In addition, we proposed a new similarity measure algorithm, called SDST, to calculate the similarity between a user query and moving objects more precisely and effectively. It allows us to retrieve precise video results based on the moving distance of objects as well as to provide ranking for the retrieved video results to answer a user query. For our performance analysis, we implement our spatio-temporal representation scheme and compare it with Li's and Shan's schemes in terms of retrieval effectiveness. We finally show from our experiment that our scheme achieves about 20% higher precision while holding about the same recall, compared with Li's and Shan's scheme.

6. REFERENCES

[1] P. Aigram, H.Zhang and D.Petkovi, "Content-Based Representation and Retrieval of Visual Media : A state-of-the-ART Review", Multimedia Tools and Applications, Vol. 3, pp. 179-202 (1996).

[2] S. S. Intille and A. F. Bobick, "Visual tracking using closed-worlds", In Proceedings of the International Conference on Computer Vision (1995).

[3] J. F. Allen, "Maintaining Knowledge about Temporal Intervals", Communication of the ACM, pp. 832-843 (1983).

[4] S. K. Chang, Q. Y. Shi and C. W. Yan, "Iconic Indexing by 2D Strings", IEEE Trans. Pattern Analysis, Machine Intelligence, Vol. 9, pp. 413-428 (1987).

[5] John Z. Li, M. Tamer Ozsu, Duane Szafron, "Modeling of Moving Objects in a Video Database", In Proceedings of the International Conference on Multimedia Computing and Systems, pp. 336-343 (1997).

[6] Man-Kwan Shan and Suh-Yin Lee, "Content-based Video Retrieval via Motion Trajectories", In Proceedings of the International Conference on SPIE, Vol. 3561, pp. 52-61 (1998).

[7] G. Salton, "A New Comparison between Conventional Indexing (MEDLARS) and Automatic Text Processing(SMART)," Journal of the American Society for Information Science, Vol. 23, No. 2, pp 75-84 (1972).