JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 22, 1519-1527 (2006)

Short Paper

Similarity-based Motion Track Management for Video Retrieval*

PEI-YI CHEN AND ARBEE L. P. CHEN
Department of Computer Science
National Chengchi University
Taipei, 116 Taiwan
E-mail: alpchen@cs.nccu.edu.tw

The motion track is an important feature to show the spatio-temporal relationship of a video object [2, 5, 7, 8]. In this paper, we propose a novel motion track representation to represent the motion track in the X-Y plane and the trend of velocity changes. Moreover, a new similarity measure for comparing two motion tracks based on the representation is proposed. Furthermore, the motion track segmentation method is proposed to handle a complicated motion behavior and the relevance feedback is used to improve the query results.

Keywords: video retrieval, motion track, segmentation, similarity measure, relevance feedback

1. MOTION TRACK REPRESENTATION

The motion track of a moving object can be represented by the motion trajectory and the velocities and both of them are modeled in polynomial.

1.1 Representation of Trajectories and Velocities

We use the first-, second-, or third-order polynomial of Y with variable X to represent the motion trajectory on the X-Y plane. In the MPEG-7 motion descriptor [3, 6], only the first- and second-order polynomials are considered. We also use the third-order polynomial because of its low computation time and high flexibility compared with the first-and second-order polynomials for modeling arbitrary curves of motion tracks. For more higher-order polynomials, the complexity will be larger, and the trajectories which need higher-order polynomial regression to model can be split into pieces of 3^{rd} or less polynomials. Therefore in this paper, we use $1^{st}/2^{nd}/3^{rd}$ polynomials as the trajectory representation.

Received June 23, 2004; revised Febuary 15 & June 07 & September 8, 2005; accepted October 5, 2005.
Communicated by Suh-Yin Lee.

* The preliminary version has been presented in the Proceedings of SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia, 2004, U.S.A.

We also record the velocities along the X-axis and Y-axis of a moving object, denoted V_x and V_y. V_x and V_y are the derivatives of the regressive curves of the x values and y values along time, respectively.

2. QUERY PROCESSING

The retrieval method in this paper includes: (i) segmentation of the motion tracks; (ii) similarity measure based on the proposed representation of motion tracks; and (iii) weight adjustment by relevant feedback.

2.1 Segmentation

For a motion track with a turning behavior, using one single regressive curve may not be able to capture the turning characteristics. However, it is an important feature of the motion tracks which users may be interested in. We define a *turning point* along the X-axis as follows. For a sequence of points along the X-axis x_i, $i = 1, 2, \ldots, m$, x_{turn} is a turning point when

(i) $x_1, x_2, \ldots, x_{turn-1} < x_{turn}$ and $x_{turn+1}, x_{turn+2}, \ldots, x_m < x_{turn}$, or
(ii) $x_1, x_2, \ldots, x_{turn-1} > x_{turn}$ and $x_{turn+1}, x_{turn+2}, \ldots, x_m > x_{turn}$.

The motion track with the turning behavior along the X-axis will be segmented on the turning points. Each segment can then be modeled by a regressive curve. Based on this method we can preserve the turning behaviors and precisely represent the motion track.

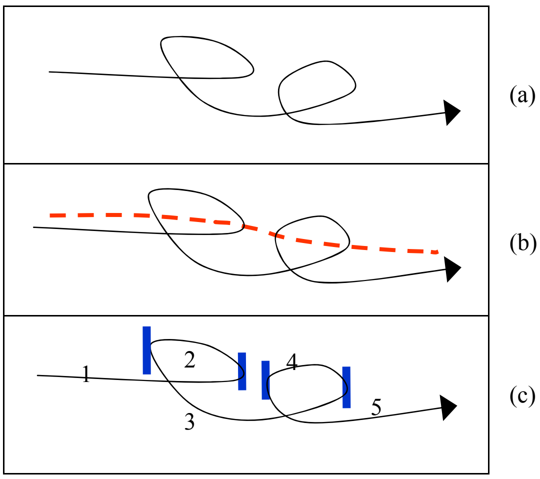


Fig. 1. (a) A motion track with turning behavior, (b) the corresponding regressive curve (the dotted line), and (c) the motion track segmented into five segments on four turning points.

2.2 Similarity Features of Motion Trajectories

After constructing the motion trajectory and the velocity polynomials, the coefficients and some other meaningful features based on the polynomials will be used to derive the similarity measure.

The motion track is modeled by a first-, second-, or third- order polynomial. For a third-order polynomial $y = ax^3 + bx^2 + cx + d$, if it has two peaks (a local maximum and a local minimum), the polynomial of derivative of y should satisfy the inequality as:

$$b^2 \geq 3ac. \tag{1}$$

This inequality is then used to check whether this third-order polynomial is degenerate. In the following, we define the SF for each type of the polynomial and explain how these SFs are derived and the advantages of using these values.

(i) Non-degenerate third-order polynomial

$SF = \left|\frac{y_2 - y_1}{x_2 - x_1}\right|$, where (x_1, y_1) and (x_2, y_2) are the two peaks of the polynomial.

(ii) Degenerate third-order polynomial

$SF = w_1 \times S_{slope} + w_2 \times S_{area}$, where S_{slope} is the slope of the tangent line on the inflection point, and S_{area} is the area between the tangent line on the inflection point $R(R_x, R_y)$ and the cubic curve in the range of $R_x \pm \varepsilon$, the value of ε is the range that to measure S_{area}; also, w_1 and w_2 are weights used for a better combination of the two values.

By observation, the inflection point (the point on which the second derivative is 0) and the *smoothness* of the curve are important features of the degenerate third-order polynomial. The two weights depend on the user's viewpoint for the similarity. If the range of the motion tracks is important, the w_2 can be larger; if the user takes care about the shape of the motion track, the w_1 is more important than w_2. For the variable ε, there exists a tradeoff between the precision and efficiency. The ε value should be condered together with the w_1 and w_2. Fig. 2 illustrates an example to show the slope of the tangent line S_{slope} and the area S_{area}. For the variable ε, there exists a tradeoff between the precision and efficiency. Fig. 2 illustrates an example to show the slope of the tangent line S_{slope} and the area S_{area}.

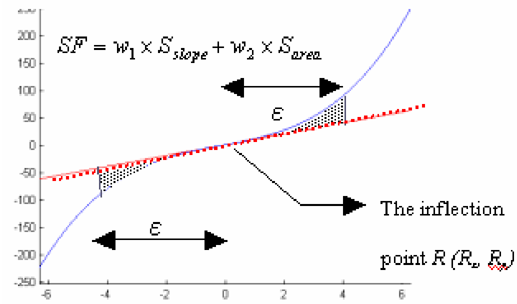


Fig. 2. The degenerate third-order polynomial with the tangent line passing through the inflection point.

(iii) Second-order polynomial
SF = the area between the tangent line of the minimum/maximum point $R(R_x, R_y)$ and the curve in the range of $R_x \pm \varepsilon$.

(iv) First-order line
SF = the slope of the line. For a line, it is simple to use the slope as the SF.

2.3 Velocity Similarity

We consider the velocity similarity in two aspects, i.e., the changes of the velocity and the average velocity. In this paper, we consider the velocity V_x and V_y as second-order polynomials. We define the similarity between the velocity trends, called *VelSim*, from 0 to 1 to show the similarity of trends.

Table 1. Six possible velocity trends.

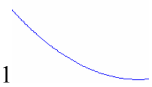
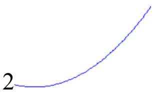
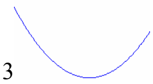
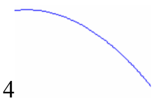
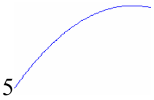
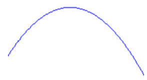
Decreasing	Increasing	Both
 1	 2	 3
 4	 5	 6

Table 2. The definition of *VelSim*.

Type	1	2	3	4	5	6
1	1	0	0.5	0.8	0	0.4
2	0	1	0.5	0	0.8	0.4
3	0.5	0.5	1	0.4	0.4	0
4	0.8	0	0.4	1	0	0.5
5	0	0.8	0.4	0	1	0.5
6	0.4	0.4	0	0.5	0.5	1

2.4 Similarity of Motion Tracks

We combine the above values for the similarity between a query segment Q_{SEG} and a motion track segment C_{SEG} in the database. For convenience, we use distance as the similarity measure where a larger distance implies a lower degree of similarity.

$$Dis(Q_{SEG}, C_{SEG}) = w_1 \times |SF_Q - SF_C| + w_2 \times [(1 - VelSim_x) + (1 - VelSim_y)] + w_3 \times |\overline{V}_Q - \overline{V}_C| \tag{2}$$

where SF_Q and SF_C are the similarity features of Q_{SEG} and C_{SEG}; $VelSim_x$ and $VelSim_y$ represent the similarity between the V_x and V_y of the Q_{SEG} and C_{SEG} respectively, and $\overline{V}_Q$

and $\overline{V}_C$ are the average velocity of Q_{SEG} and C_{SEG}; and w_1, w_2 and w_3 are the weights for the three components. If Q_{SEG} and C_{SEG} belong to different types, s, that is, SF_Q and SF_C are different types of regression curves, then the $Dis(Q_{SEG}, C_{SEG})$ is set to ∞.

2.5 Indexing

In order to facilitate efficient query processing, the segments in the database are simply indexed. Each motion track is segmented by the method described in section 2.1. The following steps show how to index the segments. The index structure is shown in Fig. 3.

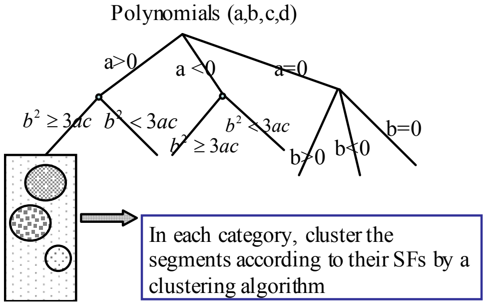
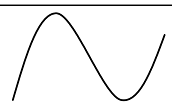
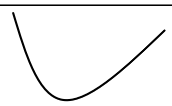
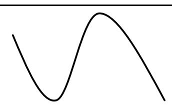
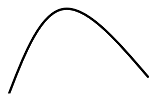
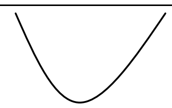
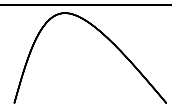


Fig. 3. The index structure.

Table 3. The relationship between the sign of *a*(3rd-order) or *b*(2nd-order) and the shape of the curve.

	3rd-order non-degenerate	3rd-order degenerate
$a > 0$		
$a < 0$		

	2nd-order
$b > 0$	
$b < 0$	

Assume the motion segment is represented as $y = ax^3 + bx^2 + cx + d$.

Step 1: All segments are divided into three categories: $a > 0$, $a < 0$, and $a = 0$. Of $a \neq 0$ then whether it is a third-order polynomial, and whether a is larger than zero is used to separate the polynomials into different types. Table 3 shows different types with their corresponding curves.

Step 2: In the $a > 0$ or $a < 0$ category, the segments are divided based on the inequality (1), that is, the segments are divided into degenerate and non-degenerate types.

Step 3: In each type, use a clustering algorithm to cluster the segments by the values of the SF.

The leaf node records the segments of motion tracks in the database. In each leaf node we record: (*Vid, MTid, start_time, end_time, offset, TrajSimilarityfeature, VelSimilarityfeature*) where *Vid* is the video identifier; *MTid* is the motion track identifier; *start_time* and *end_time* are the time stamps of this video clip containing the motion track *MTid*; offset is the serial identifier of the segment that the shows the order in the motion track *MTid*; *TrajSimilarityfeature* consists of (*a*, *SF*); and *VelSimilarityfeature* consists of (*vel_x_type, vel_y_type, average velocity*) to record the average velocity and the types of V_x and V_y.

Notice that it is not necessary to record the intervals of the curves as done in the time series data analysis, since the polynomials are constructed based on the property of the polynomials with different orders.

The disadvantage of this two-level KNN method is that the value of *k* should be larger, or we cannot find enough answers. However, since the size of the database is usually large, we can prune a lot of candidates in the first-level KNN. Furthermore, some approximations like skipping or inserting one segment can be applied on this method.

2.6 Weight Adjustment by Relevance Feedback

Since whether the motion tracks are similar or not is very subjective, we design a relevance feedback mechanism for users to adjust their answers.

We adopt the main idea of the standard deviation method in our approach. First, the *positive* and *negative* results are chosen by users and then the variances of the three features of positive and negative examples are analyzed respectively. For the positive results, the bigger the variance of some feature is, the smaller the feature weight should be. For the negative results, the smaller the variance of some feature is, the smaller the feature weight should be.

In this paper we focus on the w_2 and w_3 adjustment to help users to find results with similar velocity trends or speeds. The way to adjust the weights w_2 and w_3 for positive results in our video retrieval approach is illustrated as follows:

$$w_2 = \begin{cases} \dfrac{w_2}{Pos\_\operatorname{var} V_x \times Pos\_\operatorname{var} V_y \times R^2}, & \text{if } Pos\_\operatorname{var} V_x \neq 0 \text{ and } Pos\_\operatorname{var} V_y \neq 0 \\ \dfrac{w_2}{Pos\_\operatorname{var} V_x \times R} \times C, & \text{if } Pos\_\operatorname{var} V_x = 0 \\ \dfrac{w_2}{Pos\_\operatorname{var} V_y \times R} \times C, & \text{if } Pos\_\operatorname{var} V_y = 0 \end{cases} \tag{3}$$

$$w_3 = \frac{w_3}{Pos\_\operatorname{var} Vel}. \tag{4}$$

For all positive results, the variances of the V_x and V_y distances are calculated as $Pos\_\mathrm{var}V_x$ and $Pos\_\mathrm{var}V_y$, and the V_x and V_y distances are the *VelSim* distances between the query and the V_x and V_y, respectively. Since the range of the distance is from 0 to 1 and the variance is always less than 1, we re-range it to a constant *R* to prevent w_2 from always growing larger. If the $Pos\_\mathrm{var}V_x$ or $Pos\_\mathrm{var}V_y$ is equal to 0, it means it is a very important feature for the user. To prevent from divide-by-0 and enhance the effect of this feature, we multiply w_2 by a constant C.

In formula (4), the w_3 is re-weighted by dividing w_3 by the variance of the average velocity of the positive results.

3. EXPERIMENT RESULTS

We build a video retrieval system based on the motion track and analyze the performance.

Table 4 shows the effectiveness comparison of Lee's work [4] and ours. The precisions and the ranks of the most similar results are compared. The precision is measured by the number of relevant results out of the number of the retrieved results, and the most similar result is chosen by users. In our approach, the precision is about 85%.

Table 4. The precisions of our approach vs. Lee's work. The total number of motion tracks = 150.

Precision of our approach	Most similar rank	Precision of Lee's work	Most similar rank
0.844	1.067	0.505	1.733

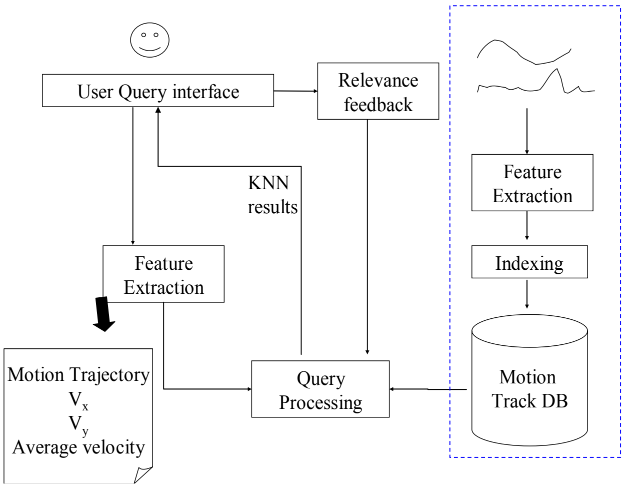


Fig. 4. The overview of our approach.

4. CONCLUSION

We propose a new descriptor to represent the motion track in the X-Y plane and the trend of velocities. The similarity measure based on the properties of the polynomial is defined for comparing two motion tracks. Moreover, segmentation method is also proposed and relevance feedback is used to improve the query results. By the experiment results we show that our approach is more effective than existing ones using polynomial representations.

REFERENCES

1. P. Y. Chen and A. L. P. Chen, "Video retrieval based on similarity of motion tracks of moving objects," in *Proceedings of SPIE Storage and Retrieval Methods and Applications for Multimedia*, Vol. 5307, 2004, pp. 550-558.
2. S. Dagtas, W. Al-Khatib, A. Ghafoor, and R. L. Kashyap, "Models for motion-based video indexing and retrieval," *IEEE Transactions on Image Processing*, Vol. 9, 2000, pp. 88-101.
3. S. Jeannin and A. Divakaran, "MPEG-7 visual motion descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, 2001, pp. 720-724.
4. K. Lee, W. You, and J. Kim, "Video retrieval based on the object's motion trajectory," in *Proceedings of SPIE Visual Communications and Image Processing*, Vol. 4067, 2000, pp. 114-124.
5. A. Yoshitaka, M. Yoshimitsu, M. Hirakawa, and T. Ichikawa, "V-QBE: video database retrieval by means of example motion of objects," in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, 1996, pp. 453-457.
6. J. Martinez, "Overview of the MPEG-7 standard," ISO/IEC JTC1/SC29/WG11 N4509, Pattaya, 2001.
7. M. Garcia and H. Nicolas, "Video object trajectory analysis," in *Proceedings of the IEEE International Conference on Image Processing*, Vol. 1, 2002, pp. 585-588.
8. Z. Aghbari, K. Kaneko, and A. Makinouchi, "Content-trajectory approach for searching video databases," *IEEE Transactions on Multimedia*, Vol. 5, 2003, pp. 516-531.

Pei-Yi Chen (陳珮宜) received the B.S. degree in 2000, and then received the M.S. degree in 2002, both in Computer Science in National Tsing Hua University.

Arbee L. P. Chen (陳良弼) received the B.S. degree in Computer Science from National Chiao Tung University, Taiwan in 1977, and the Ph.D. degree in Computer Engineering from the University of Southern California in 1984. He joined Department of Computer Science, National Tsing Hua University, Taiwan, as a Visiting Specialist in 1990 and then became a Professor in 1991. In August 2001 he took a leave from National Tsing Hua University to chair Department of Computer Science and Information Engineering at National Dong Hwa University, Taiwan. In August 2004, he moved to National Chengchi University to be a Chengchi University Chair Professor. He was a Member of Technical Staff at Bell Communications Research from 1987 to1990, an Adjunct Associate Professor in the Department of Electrical Engineering and Computer Science, Polytechnic University, New York, and a Research Scientist at Unisys, California, from 1985 to 1986. His current research interests include data stream management, multimedia databases, and data mining.

Dr. Chen organized 1995 IEEE Data Engineering Conference and 1999 Interna-

tional Conference on Database Systems for Advanced Applications in Taiwan, and served as a program or conference chair for many international conferences. He has been a visiting scholar at Kyoto University, Beijing Tsinghua University, Stanford University and King's College London. He delivered a speech at IEEE Shannon Lecture Series at Stanford University in 2005. He has been a recipient of the National Science Council Distinguished Research Award since 1996, and is listed in Who's Who in the Republic of China, Government Information Office.