Trajectory-Based Video Indexing and Retrieval Enabling Relevance Feedback

Thi Lan Le, Alain Boucher, Monique Thonnat

Trajectory-Based Video Indexing and Retrieval Enabling Relevance Feedback

Lan Le Thi [1,2], Alain Boucher [1,3], Monique Thonnat [2]
(1) International Research Center in Multimedia Information Communication and Applications,
Hanoi University of Technology, Viet Nam
(2) Projet ORION, INRIA, 2004 route des Lucioles, B.P. 93, 06902 Sophia Antipolis, France
(3) Equipe MSI, Institut de la Francophonie pour l'Informatique, ngo 42 pho Ta Quang Buu, Hanoi, Viet Nam
Thi-Lan.Le@mica.edu.vn

Abstract—**This paper is proposing an approach for retrieving videos based on object trajectories. First, a trajectory is translated into a sequence of symbols based on a symbolic representation, beyond the initial numeric representation, which does not suffer from scaling, translation or rotation. Then, in order to compare trajectories based on their symbolic representations, two similarity measures are proposed, inspired by works in bioinformatic. Moreover, based on these similarity measures, two relevance feedback strategies are given. Experimental results for two databases show that the proposed similarity measures gave results as good as other existing measures. Real advantages of these measures are the possibility for the partial matching and for relevance feedback.**

Keywords: **Video Indexing and Retrieval, Trajectory Matching, Relevance Feedback**

I. INTRODUCTION

Advances in computer technologies and the advent of the World Wide Web have made an explosion of multimedia data being generated, stored, and transmitted. For managing this amount of information, one needs developing efficient content-based retrieval approaches that enable users to search information directly via its content. Currently, the most common approach is to exploit low-level features (such as colors, textures, shape and so on). When working with videos, motion is also an important feature. When browsing a video, people are more interested in the actions of a car or an actor than in the background. Moving objects attract most of users' attention. Among the extracted features from object movement, trajectory is more and more used. In order to use the trajectory information in content-based video indexing and retrieval, one must have an efficient representation method allowing not only to index trajectories, but also to respond to the various kinds of queries and retrieval needs. For retrieval aspects, the matching strategy is also of importance.

Many works exist on trajectory representation, which can be classified into 3 levels of representation: numeric, symbolic and semantic. At the numeric representation level, one can use directly the raw data from a tracking module like in [1]. However, data from tracking contain much noise because they are obtained from algorithms tracking motion from frame to frame based on low-level features. Therefore, some methods of preprocessing such as trajectory smoothing [2] are proposed. However, citing [3] from the domain of temporal data mining, using directly the raw numeric values from trajectories can limit the efficiency of the algorithms, data structures, and some methods, or algorithms, are not (well) defined for numeric approaches but exist for symbolic approaches. Moreover, symbolic representation is closer to human perception (although not yet semantic). For trajectories, researchers have proposed many symbolic representations from one dimension [4], [5] to two dimensions [6]. A semantic representation is also of high interest for trajectories, translating values into concepts. Some preliminary works are proposed in [7].

Despite of much effort in trajectory representation, one does not have yet a successful trajectory-based video indexing and retrieval approach. One problem is that the user is not included in the retrieval loop when conceiving a representation scheme. In information retrieval, user-centered scheme are very important because a simple query can rarely fit the complete user needs. An approach enabling interaction with the user allows him/her not only to refine the query but also to judge the results given by the system.

From this previous observation, our work is aiming to add relevance feedback ability in some existing and efficient trajectory representations. Relevance feedback ability can be expressed by the possibility to change the measure distance or by a representation that can adapt better with the user needs.

Our work is concentrating on the symbolic level because of its robustness to noise, and its ability to link with the semantic level. Chen has proposed an efficient symbolic representation, beyond the numeric presentation, which does not suffer from scaling, translation or rotation [6]. In this paper, we are reusing this representation method and we are extending it with two new similarity measures inspired from the idea of sequence alignment in the bioinformatic domain [8].

The main contributions of our paper are the following:

- Propose two similarity measures for trajectory comparison at the symbolic level.
- Present two methods for relevance feedback in trajectory-based video indexing and retrieval based on the proposed similarity measures.

The rest of the paper is organized as follow. In Section 2, we are proposing a structure for a **T**rajectory-**B**ased **V**ideo **I**ndexing and **R**etrieval **E**nabling Relevance **F**eedback (TBVIREF) which includes trajectory representation at the numeric and symbolic level, sequence alignment and feedback. Some experimental results are shown in section 3. Section 4 is concluding this paper with some directions for future work.

II. Trajectory-Based Video Indexing and Retrieval Enabling Relevace Feedback (TBVIREF)

A. General description

We are proposing an architecture for TBVIREF (Fig. 1). In this architecture, object tracking is done by a preprocessing module (not shown here), and object trajectories are taken as input. In the real physical world, a trajectory is represented following 3 dimensions. But without a priori contextual information, trajectories are represented in 2D. Knowing the application and its context, it can be useful to represent a trajectory in 3D or to map the 2D trajectory into the monitored environment [9]. In this paper, we will consider only the general case in 2D, without a priori knowledge on the application.

For indexing, all object trajectories are processed through numeric and symbolic representation modules. The output is a symbolic representation of the global trajectory.

For retrieval, a given trajectory by the user is also processed through these two modules, and comparison is made with the trajectories in the database based on two available similarity measures (section II.D). Trajectories that are the most similar (given the chosen similarity measure) with the trajectory query are returned to users.

Moreover, unlike other approaches for trajectories, this approach includes the user in the retrieval loop. The user judges the returned results as relevant or irrelevant through an interface. Then, retrieval loops again and continues until the user is satisfied.

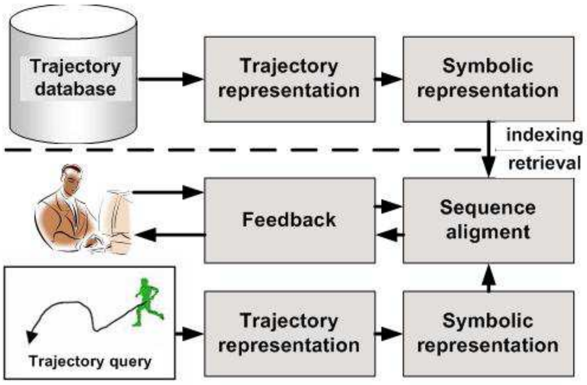


Figure 1. Architecture for Trajectory-Based Video indexing and Retrieval Enabling Relevance Feedback

B. Numeric trajectory representation

Working with raw data from the object trajectory is not always suitable because these data are sensible to noise and are affected by rotation, translation and scaling. In order to cope with this problem, we have chosen among the existing representation methods the one from [6], which also uses both direction and distance information of the movement.

Given a sequence, $T_A=[(x_{a,1},y_{a,1}),\ldots,(x_{a,n},y_{a,n})]$, n being the length of T_A, a sequence of (movement direction, movement distance ratio) pairs M_A is defined as a sequence of pairs $M_A=[(\theta_{a,1},\delta_{a,1}),\ldots,(\ \theta_{a,n-1},\delta_{a,n-1})]$. The movement direction is defined as:

$$\phi_{a,i} = \arctan\frac{y_{a,(i+1)} - y_{a,(i)}}{x_{a,(i+1)} - x_{a,(i)}}$$

$$\theta_{a,i} = \begin{cases} \phi_{\mathrm{a,i}} & \text{if } x_{a,(i+1)} - x_{a,(i)} \geq 0 \\ \phi_{a,i} + \Pi & \text{if } x_{a,(i+1)} - x_{a,(i)} < 0 \text{ and} \\ & \quad y_{a,(i+1)} - y_{a,(i)} > 0 \\ \phi_{a,i} - \Pi & \text{if } x_{a,(i+1)} - x_{a,(i)} < 0 \text{ and} \\ & \quad y_{a,(i+1)} - y_{a,(i)} \leq 0 \end{cases} \qquad (1)$$

and the movement distance ratio is defined as:

$$\delta_{a,i} = \begin{cases} \dfrac{\sqrt{(y_{a,(i+1)} - y_{a,i})^2 + (x_{a,(i+1)} - x_{a,i})^2}}{TD(T_A)} & \mathrm{TD(T_A)} \neq 0 \\ 0 & \mathrm{TD(T_A)=0} \end{cases} \qquad (2)$$

$$TD(T_A) = \sum_{1 \leq j \leq n-1} \sqrt{(y_{a,(j+1)} - y_{a,j})^2 + (x_{a,(j+1)} - x_{a,j})^2}$$

Raw trajectory data given to this module are transformed into sequence of pairs of movement direction and movement distance ratio. We can use directly this sequence to compare trajectories or we can use it as an intermediate information for the symbolic representation module.

C. Symbolic trajectory representation

Using the previous numeric representation for trajectories, a proposed symbolic representation from [6] is computed as follows:

Given ε_{dir} and ε_{dis}, two dimensional (movement direction, distance ratio) space is divided into $\left(\frac{2\pi}{\varepsilon_{dir}}\right)\times\left(\frac{1}{\varepsilon_{dis}}\right)$ subregions. Fig.2 gives an example of this quantization map, ε_{dir} and ε_{dis} are chosen $\pi/4$ and 0.125. Each subregion SB_i is represented by two (movement direction, distance ratio) pairs: $(\theta_{bl,i},\delta_{bl,i})$ and $(\theta_{ur,i},\delta_{ur,i})$, which are the bottom left and the upper right coordinates of SB_i. A distinct symbol A_i is assigned for subregion SB_i.

A pair of movement direction and movement distance ratio $(\theta_{a,i},\delta_{a,i})$ will be represented by a symbol A_i if $\theta_{bl,i} \leq \theta_{a,i} < \theta_{ur,i}$ and $\delta_{bl,I} \leq \delta_{a,i} < \delta_{ur,i}$. Once the two threshold values ε_{dir} and ε_{dis}

are given for the trajectory data, the size of movement pattern alphabet is fixed.

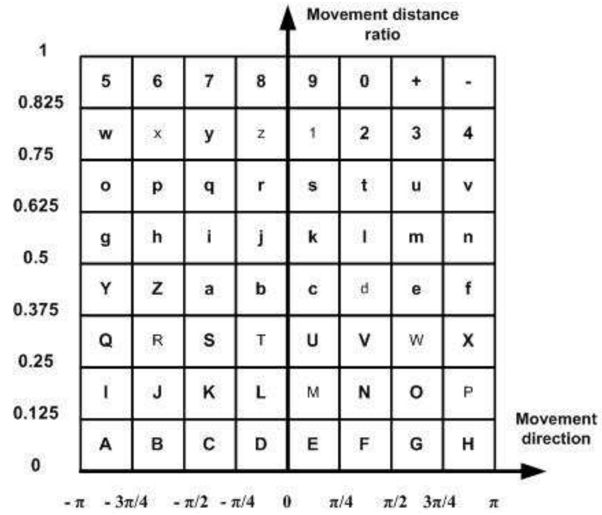


Figure 2. A quantization map for the symbolic representation (reprinted from [6])

D. Sequence alignment and matching

The sequence alignment consists in the construction of a dictionary and the trajectory matching. Dictionary construction transforms a sequence of n symbols $S=[s_1, \ldots,s_n]$ into a set of distinct ℓ length-words $\{w_j\}$, ℓ being the length of a word. This word set is called a dictionary.

Lets D_Q being the dictionary for a sequence query S_Q (D_Q consists of r distinct ℓ length-words $\{w_{iQ}\}$), D_T being the dictionary of a sequence S_T in the database (D_T consists of k distinct ℓ length-words $\{w_{jT}\}$). Because a word may appear several times in a sequence and in the dictionary, we store the position of each word and the number of their occurrences.

Before presenting the proposed similarity measures, we give here two notations: *Hit* and *Ap_hit* notations.

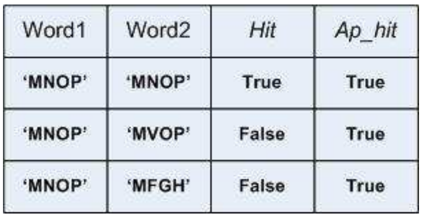
Definition: Hit

Given two sequences $S_Q=[s_{1Q}, \ldots,s_{nQ}]$ and $S_T=[s_{1T}, \ldots,s_{mT}]$ after the dictionary construction process, S_Q can be represented by r distinct ℓ length-words $\{w_{iQ}\}$, i from 1 to r and S_T can be represented by k distinct ℓ length-words $\{w_{jT}\}$, j from 1 to k. A *Hit* can be identified between S_Q and S_T if one find a word w_{iQ} in S_Q and a word w_{jT} in S_T satisfy $w_{iQ} = w_{jT}$, i from 1 to r and j from 1 to k.

Definition: Ap_hit

Given two sequences $S_Q=[s_{1Q}, \ldots,s_{nQ}]$ and $S_T=[s_{1T}, \ldots,s_{mT}]$ after dictionary construction process, S_Q can be represented by r distinct ℓ length-words $\{w_{iQ}\}$, i from 1 to r and S_T can be represented by k distinct ℓ length-words $\{w_{jT}\}$, j from 1 to k. A *Ap_hit* (approximate hit) can be identified between S_Q and S_T if one find a word w_{iQ} in S_Q and a word w_{jT} in S_T satisfying $w_{iQ} = w_{jT}$ or w_{iQ} being a neighbor of w_{jT}. Two words are neighbor if their orderly symbols are assigned for the neighbor regions in the quantization map.

One can realize that the *Ap_hit* notation can resolve the problem of quantization (two close numeric values being assigned different symbols because of the quantization). We give in Fig. 3 some examples of these notations. With two given words: word1 is 'MNOP' and word2 is alternatively 'MNOP', 'MVOP' and 'MFGH', using the same quantization map in Fig. 2, *Hit* and *Ap_hit* notations give different values.



Word1	Word2	*Hit*	*Ap_hit*
'MNOP'	'MNOP'	True	True
'MNOP'	'MVOP'	False	True
'MNOP'	'MFGH'	False	True

Figure 3. Some examples of *Hit* and *Ap_hit* notations with two given words. word1 is 'MNOP' and word2 is alternatively 'MNOP', 'MVOP' and 'MFGH', using the same quantization map as in Fig. 2, *Hit* and *Ap_hit* notations give different values.

Based on these *Hit* and *Ap_hit* notations, we propose two similarity measures:

- Scanning the sequence S_Q until detecting a *Hit* or a *Ap_hit* at i^{th} word between this sequence with a sequence S_Q in the database, making an extension alignment in [8] and computing a similarity measure $d(S_Q, S_T, i)$. This first similarity measure is computed as:

$$d_2(S_Q, S_T)=\max_{i=1\ldots r}(d(S_Q,S_T,i)) \quad (3)$$

- This similarity measure take into account the number of the *Hit* or the *Ap_hit* between the sequence query S_Q and a sequence S_Q in the database. A word hit vector *word_hit*, for j from 1 to k, each element *word_hit[i]* storing the number the *Hit* or the *Ap_hit* of word w_j between the S_Q and the S_T. This second similarity measure between S_T and S_Q is defined as:

$$d_2(S_Q, S_T)=\sum_{i=1}^{r} \text{word\_hit}\ [i] \quad (4)$$

From these definitions, one can see that partial matching can be used for both of the proposed similarity measures. By identifying a subsequence inside the trajectory query, the similarity measures (3) and (4) between this subtrajectory query and trajectories in the database can be computed. Furthermore, the proposed similarity measures allow the user to decide the importance of some parts in the trajectory query by giving their weight.

Because the proposed similarity measures are sensible to the starting point for comparing two trajectories, we detect the point in the sequences where the first *Hit* or *Ap_hit* is found and then compute the similarity from this point

E. Relevance Feedback

Relevance feedback is a well known technique in the information retrieval domain. It covers a range of techniques intended to improve a user query and facilitate retrieval of information relevant to the user needs. This technique has been firstly used for text retrieval, but now it becomes more and more important for image and video retrieval as well. Up to now, to our knowledge, there is not any work dedicated to the trajectory-based video indexing and retrieval enabling relevance feedback.

In our approach, associated with the two proposed similarity measures, we have developed two new methods for including feedback. Even if results are not yet available for these methods, we present them in this section as current development and to stress the possibility of including interactions in the presented similarity measures.

Let RT be the set of N the most similar trajectories
RT= [RT_1,…,RT_N] at current search result.

Relevance feedback 1: The user is judging the results of the current search as being relevant or irrelevant. From this feedback information, one can know which word hit is more significant than the others for this query. Let *index_hit* be the index of this word. For the next retrieval loop, in spite of using the similarity measure (3), a new similarity measure is defined as:

$$d_{1,new}(S_Q, S_T) = d(S_Q, S_T, index) \quad (5)$$

Relevance feedback 2: A given query S_Q consisting of r distinct ℓ length-words, a weight vector is specified where *weight[i,t]* is the weight for the words w_j at t, j from 1 to r. At t=0, this vector is initialized with the number of occurences of the word w_j in the trajectory query. From the feedback information, with each result sequence RT_j, this vector is updated as:

$$\begin{aligned} &weight[i,t+1]) = weight[i,t]) + word\_hit[i] \text{ if RT}_j \text{ is relevant} \\ &weight[i,t+1]) = weight[i,t]) + word\_hit[i] \text{ if RT}_j \text{ is irrelevant} \end{aligned} \quad (6)$$

For the next retrieval loop, in spite of using the similarity measure (4), a new similarity measure is defined as:

$$d_{2,new}(S_Q, S_T) = \sum_{i=1}^{r} (word\_hit[i] * weight[i,t+1]) \quad (7)$$

One can see that a word often appearing in the relevant examples means that this word is more important. Therefore, in the next retrieval loop, this word is chosen to compute the similarity measures (relevance feedback 1) or to add more weight on its hit (relevance feedback 2) and the opposite for irrelevant words. From this observation, the proposed relevance feedback methods can take into account the user needs.

III. EXPERIMENTS

A. Video and trajectory databases

As said in section II.A showing the general description of our architecture, we are supposing that object tracking is done by a preprocessing module, and object trajectories are taken as input for our work. Therefore, in this section, in order to test our approach, we are using two trajectory databases. We are comparing our proposed similarity measures with the existing Edit Distance on Movement Pattern (EDM) [6].

The ASL data set from UCI KDD[1] data archive consists in samples of signs from the Australian Sign Language. More than 95 signs were collected from 5 different writers. In total, this sign database comprises 6577 signs. Fig. 4 gives some samples of the word 'eat' written by 5 different writers.

The trajectory database of Yuan Ze University [2] containing 2500 trajectories which comes from 50 categories was generated. Fig. 5 shows all kinds of trajectory types used in this database. Each category includes 50 samples that were drawn in different sizes, moving directions, rotations, and translations.

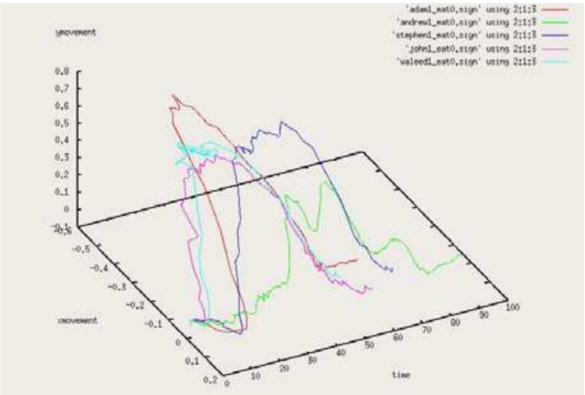


Figure 4. Some samples from the ASL database of the word 'eat' from 5 different writers

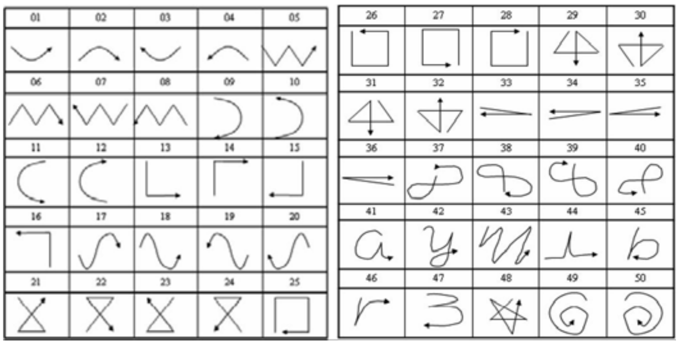


Figure 5. All kinds of trajectory types used in the Yuan Ze University database. Each category includes 50 samples.

[1] http:// kdd.ics.uci.edu/databases/auslan/

[2] http://mmplab.eed.yzu.edu.tw/ trajectory/trajectory.rar

B. *Experiments*

In these experiments, ε_{dir} and ε_{dis} are chosen as $\pi/4$ and 0.125 for the symbolic representation.

In order to evaluate and compare the results, we use a well-known measure in the information retrieval domain: recall/precision curve [10]. Recall defines the number of relevant documents retrieved as fraction of all relevant documents and precision defines the number of relevant documents as a fraction all the documents in retrieved by the system. The better method, the nearer it approaches to the ideal point where recall is 100% and precision is 100%.

Fig. 6 is giving the retrieval results for the ASL trajectory database. In this experiment, the chosen length ℓ for words is 5. The curve with violet circles presents the retrieval results using the EDM distance, the curve with blue rectangles presents retrieval results using the second proposed similarity measure (4). One can see that for this database the proposed similarity measures give results comparable with those for EDM. Moreover, with a small number of returned results (recall is between 0 and 10), the proposed measure gives a lightly better result. This characteristic can be useful because in the information retrieval domain, the user is usually interested in only some relevant results. However, the results for both measures on the ASL database are not very good because the ASL database is a difficult database. One word can be written in different ways by the same writer (different trajectories); moreover, the trajectories in this database are acquired by extracting a glove position during the writing phase, with much noise. The presented experimental results here are obtained without using any pre-processing on the data like noise filter.

Fig. 7 gives the retrieval results for the second proposed similarity measure (4) for the Yuan Ze University database with different word lengths. For this database, results are better for a short word length like 3 than for longer word lengths.

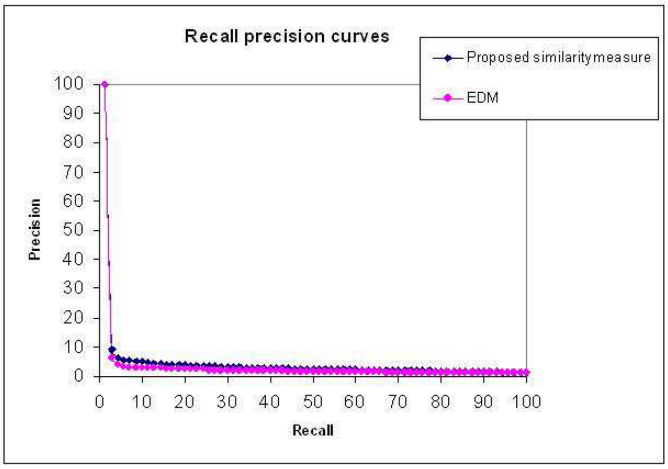


Figure 6. Recall and precision curves for the ASL trajectory database. The curve with violet circles presents the retrieval results using the EDM distance [6], while the curve with the blue rectangles presents the retrieval results using the second proposed similarity measure.

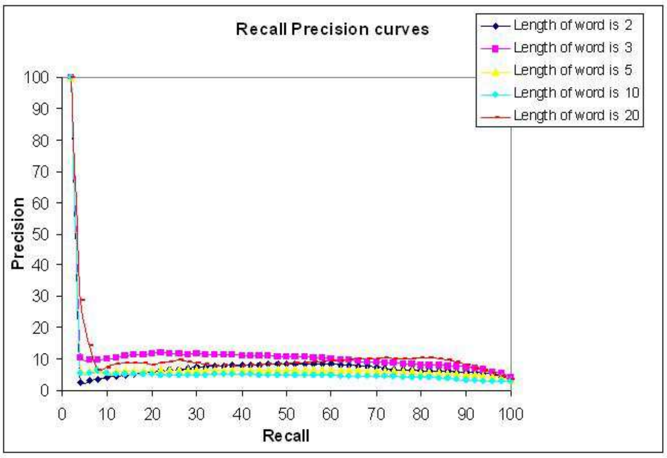


Figure 7. Recall and precision curves of the proposed similarity measures on the second trajectory database with different word lengths. The curve with blue diamonds presents the results with the used word length is 2, the curve with violet rectangles presents results with the used word length is 3, the curve with yellow triangles presents results with the used word length is 5, the curve with green circle presents results with the used word length is 10 and the curve with red dot presents results with the used word length is 20.

As one can see from above results, choosing the word length affects significantly the retrieval results, so it could be useful to develop some heuristics. Moreover, a fixed length is not always suitable for various types of trajectories.

From the obtained results, we can see that the proposed similarity measures can give interesting and good results for partial matching, as shown in Fig. 8. Using a simple trajectory query (Fig. 8a) for retrieving more complex trajectories which include the query trajectory (Fig. 8 b, c, d). This characteristic can be very important in trajectory-based video indexing and retrieval when users are interested in only one part of object trajectory, or for retrieving trajectories containing a given pattern.

Apart from the above advantages, the proposed similarity measures have some weakness. They suffer from the starting point from which we divide the words and the length of words. For the starting point, we are using the sequence alignment to detect the first *Hit* or *Ap_hit* between two trajectories and are computing distances between them from this point. The alignment method that we used needs to be improved in order to be more robust. For choosing the word length, it could be useful to develop some heuristics.

Two methods of relevance feedback are also proposed. From the theoretical analysis, these methods promise giving a retrieval improvement. The implementation and some experimentation should be done to prove their performance.

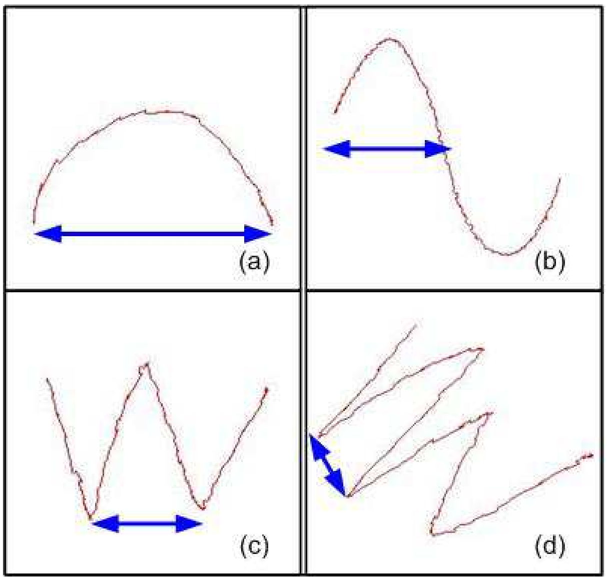


Figure 8. Partial matching using the proposed similarity measures. (a) simple trajectory query (b)(c)(d) complex retrieved trajectories which include the initial query.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, a Trajectory-Based Video Indexing and Retrieval Enabling Relevance Feedback approach has been proposed. In this approach, trajectory matching is effected at the symbolic level based on a numeric representation that does not suffer from noise, rotation, translation. Two similarity measures have been presented. From these measures, two ways for making the relevance feedback have been proposed. Some experimental results have been shown.

As we have discussed in the experiment section, the proposed measures enable doing some partial matching and relevance feedback. These are important techniques to improve the retrieval results. However, they are affected by the starting point from which we divide the words and the length of word. Further work is needed to strengthen the methods on this point.

In addition, in this paper, two methods for enabling relevance feedback in trajectory retrieval have been proposed. Further work will provide experiments to prove their effectiveness.

REFERENCES

[1] S. Dagtas, W. Al-Khatib, A. Ghafoor and R. L. Kashyap, “Models for motion-based video indexing and retrieval”, IEEE Transactions on Image Processing (2000), 370– 377.

[2] E. Sahouria, A. Zakhor, “A trajectory Based Video Indexing System for Street Surveillance”, IEEE Int. Conf. on Image Processing (ICIP).

[3] J. Lin, E. Keogh, S. Lonardi and B. Chiu, “A Symbolic Representation of Time Series, with Implications for Streaming Algorithms”, In Proc. of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13(2003).

[4] R. Agrawal, G. Psaila, E. L. Wimmers and M. Zaot, “Querying shapes of histories”, In Proc. Twenty-first Int. Conf. on Very Large Databases (VLDB '95) (1995), 502—514.

[5] P.Y. Chen and A. L. P. Chen, “ Video Retrieval Based on Motion Tracks of Moving Objects”, In SPIE Storage and Retrieval Methods and Applications for Multimedia, NewYork (2004), 15--16.

[6] L. Chen, M. T. O¨su and V. Oria, “Symbolic Representation and Retrieval of Moving Object Trajectories”, In MIR’04, NewYork (2004), 15–16.

[7] J. Z. Li, M. T. A Ozsu and D. Szafron, “Modeling of moving objects in a video databas”, In Proc. 4th Int. Conf. on Multimedia and Computing System, 336--343, 1997.

[8] S. F. Atlschul and W. Gish and W. Miller and E. W. Myers: Basic local alignment search tool. In Journal Mol. Biol., (1995), 403-410.

[9] C. Piciarelli, G. L. Foresti and L. Snidara, “Trajectory clustering and its applications for video surveillance”, Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2005,40- 45.

[10] H., Muller and V., Müler and D., McG.Squire, “Performance Evaluation in Content-Based Image Retrieval : Overview and Proposals”, Technical report vision, N. 99.05, December 1, 1999, Universite de Geneve.