# Video Query Using Temporal Signature and Similarity Matching

Shi-Wei Lo[1, 2] and Fang-Pang Lin[*1]

[1] *National Center for High-performance Computing, No. 7, R&D 6th Rd., Hsinchu Science Park, Hsinchu City, Taiwan, R.O.C.*

[2] *Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan, R.O.C.*

**Keywords:** Video Query, Video Representation, Similarity Measure.

**Abstract.** Large amount of video data is stored and distributed in wide variety of application. Due to the fast video material increases, manage and query of video become more and more important. In this paper, we address a temporal signature representation and similarity model to retrieval the similar video within database by video query. Experimental results on real date are presented. The experimental results show that the statistical approach permits accurate query of video clip, in particular, the performance of the approach was found extremely satisfactory with determine all similar video in database.

## 1. INTRODUCTION

Due to a large amount of video clips are stored in digital storages and database, the effective techniques for video query are currently important work. In the video query, one critical technique is the measuring the similarity of visual information for query clip and database. For satisfy the user's query by a video clip, a temporal signature representation and signature similarity model are proposed here.

The most important technique for the video query is efficient retrieval method based on the content of video clips. The techniques for content based video retrieval relied on textual annotations, or on key frame representations [1]. The textual annotation need automatic method of image annotation and using one word queries to retrieval from images and videos [2]. There still remains a difficult problem that has not yet been fully exploited. Such as textual query in a lot of similar content or object in video, common rule of textual annotation to make sure the key word has same criterion, how many words is sufficient for texting an image or video, etc. In the video representation of frame based, there are two categories for video retrieval techniques: frame-by-frame and key-frame based. The frame-by-frame based technique use the whole frames of video sequence to representing the features. Another technique is key-frame based that use the few but important key frames of shots to representing a video sequence. Due to the simplifying sequence by few key frames the computational cost of the video pre-processing and similarity measuring is reduced. The drawback of key frame based is that leave out the continuous of temporal features and relevance between an individual shots [3].

Dimitrova and Abdel-Mottaleb [1] propose a DC+M signatures to represent a sequence of video frames. They use DC coefficients and motion information to characterize the video clips as the signatures for measuring the similarity between two video clips. This method can only use in MPEG

[*] Corresponding author, email: fangpanglin@gmail.com

coded video and Motion JPEG which consist of the DC coefficients only. They also use the key frames for deriving signatures. Chen, Chin and Liao [3] proposed a shot content representation for shot matching using all frames within shot. They are using all frames within an individual shot to exploit the spatio-temporal information in video. Additionally, in their research which all the video clips need to segment into shots. The shots are achieved by their boundary detection algorithm from video source.

Several researchers already looked into the problem of video query in clips. Recently, the represented signatures of video sequences are used to measuring the similarity in video query processing. The video query steps is (1) represents the video clips by signatures then (2) measure the similarity of signatures (3) retrieve the video clips in the database with most similar signature to this query clip. However, there is lack of simple and directly approach to handle the represented signatures and the similarity measure.

In this paper, we present a temporal signature representation method to characterize the video clips and a similarity model to search the similar signatures. After these the similarity measure, the most similar video from database are retrieved. In our approach, we do not detect the shot or boundary of video sequence but using the all frames within video sequence to characteristic a temporal signature. The results of this experiment clearly showed that this approach achieved the good accuracy for query by video clip.

## 2. VIDEO QUERY SYSTEM

The purpose of this paper was to determine the temporal signature of video clips that used to establish the similarity between query video and the videos within video database. There are two steps for query by video clip: (i) the signature was produced by frame-based chi-square test. (ii) the results signature features of video which was used for evaluating the similarity by Hausdorff distance. Those two steps was illustrated to the following respectively.

### 2.1. Video Temporal Signature Representation

The chi-square test is the most accepted test for comparing two binned distributions and be used in frame-by-frame change detection [4]. To decision the temporal variation in adjacent frame of the sequence is characterized by a value. This value includes variable values, which indicate the degree of variation between consecutive frames based on histograms. Histograms are less sensitive to object motion and disparity noise between two adjacent frames. In this paper, the variable values were used to represent the temporal signature of video in time domain.

The chi-square test ($X^2$) used for the area between the distributions is given in Eq. (1) as proposed by [5].

$$X^2 = \sum (R_i - S_i)^2/(R_i + S_i) \quad (1)$$

Where $i$ is number of gray levels (bins), $R_i$ the number of events in bin $i$ for previous frame, $S_i$ the number of events in the same bin $i$ for the next frame. The chi-square test output the degree of variation of two processed frames. For chi-square value, a low score represents a lower variation than a high score. The results values were represent as temporal signature of input video sequence.

The video in query or in database ware all processed by chi-square test that the output value

exporting as the time domain variation. The time domain variation represents the temporal signature of video clip that used to determine similarity between query video and database.

### 2.2. Video Similarity Measure

The Hausdorff distance measures the distances between two sets of points. Thus, it can be used to measure the similarity between two patterns of points when they are superimposed on one other [5]. Additionally, the Hausdorff distance can be used in a wide range of applications, such as real scene recognition, tracking, engineering drawing understanding, and aerial image analysis [6]. Here the Hausdorff distance was used to measure the similarity of signatures which produced from chi-square test.

Given two point sets *A* and *B*, the Hausdorff distance between *A* and *B* is given in Eq. (2). The *A* and *B* can be temporal signature of query video clip and video database respectively.

$$H(A,B) = max(h(A,B),h(B,A)), \tag{2}$$

where $h(A,B) = max_{a \in A} min_{b \in B} \| a - b \|$, from *A* to *B*, and *h(B,A)* is from *B* to *A*. Commonly, the *h(A,B)* is not equal to *h(B,A)* but in proposed temporal signature, the point sets, can be arrange in time series model that restrict the distance from *A* to *B* as same as from *B* to *A*. Therefore, the calculation of *h(B,A)* can be avoided with this model.

The Hausdorff distance detects the similarity of the two input signatures; the Hausdorff distance will be computed by search in full length of signature and produced distance in each search step. Thus the minimum Hausdorff distance *min[H(A,B)]* represents the maximum similarity index of database that is most similar to query video clip. The measure of similarity of video clips was accomplished by using temporal signature as feature curves represented in time domain then search the minimum Hausdorff distance for each signature.


Fig. 1 Experimental video database contain 20 different clips. The scene of video is taking from Lifou, Melanesia, New Caledonia and all clips are from this location. (ReefVid: Free Reef Video Clip Database. http://www.reefvid.org/).

## 3. EXPERIMENTAL VIDEO DATABASE

The experimental video database consisted of the clips listed in Fig. 1. The experimental video clips were taken from ReefVid: Free Reef Video Clip Database. Each video include at the least one

shot or more than two shots. The video clips include many shot were able to reappear the complexity of actual situation of recording. In addition, the content of video clips is taken from around the Lifou Island and most of clips are actions among the coral reef that were complex with extensive camera motion such as pan/tile/zoom/tracking. These video also not have the editor's graphical effects such as fade in, fade out and dissolves.

The video clips have image size of 384x288 and coded at 25 fps of MPEG-4. Each clip has variance recording length from the shortest 90 frames (video-368) to 1952 frames (video-362). Query video clips were picking up the every clip in the video database. Then picked query video clip were perform search in entire video database by compute the similarity of signatures.

TABLE I. SIMILARITY FACTOR FOR INDIVIDUAL VIDEO QUERIES.

| Similarity | Video-357 | Video-358 | Video-359 | Video-360 | Video-361 | Video-362 | Video-363 | Video-364 | Video-365 | Video-366 | Video-367 | Video-368 | Video-369 | Video-370 | Video-371 | Video-372 | Video-373 | Video-374 | Video-375 | Video-376 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Video-357 | 0.0 | 19.5 | 20.6 | 38.9 | 7.3 | 29.7 | 34.6 | 11.1 | 37.5 | 38.4 | 38.6 | 4.7 | 8.1 | 8.0 | 11.3 | 37.2 | 60.1 | 7.9 | 38.6 | 35.0 |
| Video-358 | 19.5 | 0.0 | 11.6 | 185.8 | 7.2 | 196.1 | 16.0 | 10.1 | 14.2 | 11.2 | 22.1 | 4.8 | 8.6 | 9.0 | 10.6 | 111.4 | 63.4 | 8.9 | 47.0 | 91.7 |
| Video-359 | 20.6 | 11.6 | 0.0 | 1.6 | 0.2 | 4.3 | 3.0 | 3.8 | 9.0 | 1.7 | 0.7 | 3.6 | 9.7 | 0.3 | 2.7 | 2.0 | 80.7 | 0.3 | 0.7 | 1.3 |
| Video-360 | 38.9 | 185.8 | 1.6 | 0.0 | 0.2 | 250.9 | 132.8 | 3.5 | 122.3 | 1.6 | 123.1 | 3.1 | 9.0 | 0.3 | 2.1 | 249.0 | 122.7 | 0.3 | 134.5 | 231.6 |
| Video-361 | 7.3 | 7.2 | 0.2 | 0.2 | 0.0 | 0.3 | 1.6 | 0.8 | 1.5 | 1.0 | 0.4 | 3.5 | 9.1 | 0.2 | 0.8 | 0.3 | 0.4 | 0.2 | 0.3 | 0.8 |
| Video-362 | 29.7 | 196.1 | 4.3 | 250.9 | 0.3 | 0.0 | 47.5 | 3.6 | 9.9 | 3.4 | 56.7 | 3.1 | 8.4 | 0.4 | 2.5 | 202.9 | 32.6 | 0.3 | 94.4 | 201.6 |
| Video-363 | 34.6 | 16.0 | 3.0 | 132.8 | 1.6 | 47.5 | 0.0 | 3.4 | 7.8 | 3.1 | 13.2 | 2.8 | 7.5 | 1.8 | 2.3 | 91.9 | 78.9 | 1.8 | 13.2 | 90.1 |
| Video-364 | 11.1 | 10.1 | 3.8 | 3.5 | 0.8 | 3.6 | 3.4 | 0.0 | 3.5 | 3.3 | 3.7 | 3.3 | 8.8 | 0.9 | 3.8 | 3.1 | 3.6 | 0.9 | 3.6 | 3.1 |
| Video-365 | 37.5 | 14.2 | 9.0 | 122.3 | 1.5 | 9.9 | 7.8 | 3.5 | 0.0 | 8.0 | 7.4 | 3.2 | 7.5 | 1.8 | 2.7 | 7.0 | 78.3 | 1.8 | 8.4 | 7.1 |
| Video-366 | 38.4 | 11.2 | 1.7 | 1.6 | 1.0 | 3.4 | 3.1 | 3.3 | 8.0 | 0.0 | 1.8 | 3.2 | 8.5 | 1.4 | 1.9 | 1.6 | 79.1 | 1.4 | 1.7 | 1.5 |
| Video-367 | 38.6 | 22.1 | 0.7 | 123.1 | 0.4 | 56.7 | 13.2 | 3.7 | 7.4 | 1.8 | 0.0 | 3.4 | 8.9 | 0.5 | 2.4 | 105.0 | 80.3 | 0.4 | 95.2 | 90.4 |
| Video-368 | 4.7 | 4.8 | 3.6 | 3.1 | 3.5 | 3.1 | 2.8 | 3.3 | 3.2 | 3.2 | 3.4 | 0.0 | 6.1 | 3.8 | 3.2 | 2.7 | 3.4 | 3.7 | 3.4 | 2.7 |
| Video-369 | 8.1 | 8.6 | 9.7 | 9.0 | 9.1 | 8.4 | 7.5 | 8.8 | 7.5 | 8.5 | 8.9 | 6.1 | 0.0 | 10.0 | 8.3 | 7.7 | 9.2 | 9.9 | 9.2 | 8.2 |
| Video-370 | 8.0 | 9.0 | 0.3 | 0.3 | 0.2 | 0.4 | 1.8 | 0.9 | 1.8 | 1.4 | 0.5 | 3.8 | 10.0 | 0.0 | 1.7 | 0.3 | 0.6 | 0.3 | 0.4 | 1.1 |
| Video-371 | 11.3 | 10.6 | 2.7 | 2.1 | 0.8 | 2.5 | 2.3 | 3.8 | 2.7 | 1.9 | 2.4 | 3.2 | 8.3 | 1.7 | 0.0 | 1.9 | 2.5 | 1.7 | 2.3 | 1.9 |
| Video-372 | 37.2 | 111.4 | 2.0 | 249.0 | 0.3 | 202.9 | 91.9 | 3.1 | 7.0 | 1.6 | 105.0 | 2.7 | 7.7 | 0.3 | 1.9 | 0.0 | 85.5 | 1.3 | 104.5 | 152.2 |
| Video-373 | 60.1 | 63.4 | 80.7 | 122.7 | 0.4 | 32.6 | 78.9 | 3.6 | 78.3 | 79.1 | 80.3 | 3.4 | 9.2 | 0.6 | 2.5 | 85.5 | 0.0 | 0.6 | 80.1 | 22.6 |
| Video-374 | 7.9 | 8.9 | 0.3 | 0.3 | 0.2 | 0.3 | 1.8 | 0.9 | 1.8 | 1.4 | 0.4 | 3.7 | 9.9 | 0.3 | 1.7 | 1.3 | 0.6 | 0.0 | 0.3 | 0.9 |
| Video-375 | 38.6 | 47.0 | 0.7 | 134.5 | 0.3 | 94.4 | 13.2 | 3.6 | 8.4 | 1.7 | 95.2 | 3.4 | 9.2 | 0.4 | 2.3 | 104.5 | 80.1 | 0.3 | 0.0 | 102.6 |
| Video-376 | 35.0 | 91.7 | 1.3 | 231.6 | 0.8 | 201.6 | 90.1 | 3.1 | 7.1 | 1.5 | 90.4 | 2.7 | 8.2 | 1.1 | 1.9 | 152.2 | 22.6 | 0.9 | 102.6 | 0.0 |

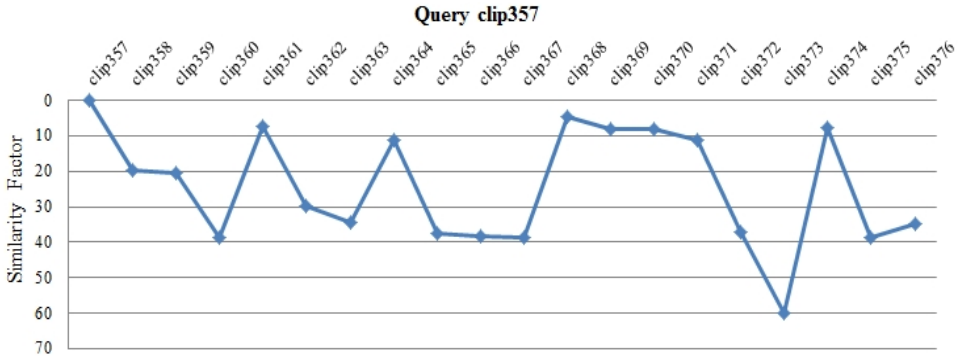


Fig. 2 Similarity factor of query video-357. The maximum similarity for query video to each video database has been plot. The lower value of similarity factor is indicating more similar to query video clip.

## 3. RESULTS

The proposed video query approach has been validated by experiments with a variety of video contents. The results of this experiment were reported in table 1. Table 1 shows the each search result by query video listed in top row. In the top row of Table 1 the query video is start from 357 to 376 respectively.

The value in the table 1 is the similarity factor between query and database that lower value indicates the more similar to the query video. If the similarity factor is close to the zero which presents the query video and relative video of database are perfect match.

Fig. 2 shows the similarity factor of query video-357 to the all video in the database. It is without doubt that query video-357 have perfect match to itself and its similarity factor is located in zero. Fig. 2 is plotted by extract the partial data from Table 1. Each query video could plot similar figure as the same way. The searching of similarity factor which used Hausdorff distance on signatures of video database was evaluated. Note that higher peak (close to zero) in results presented the most similar video.

## 3. CONCLUSIONS

This paper presents a new temporal signature representation of video clips which used the temporal variation from frame base chi-square test. The similarity measure was performed by Hausdorff distance for query video and database. The experimental results show that each query video can be retrieve accurately. We have described in this paper an original and efficient approach to the video query. It involves the chi-square test and the characterization of temporal features. We have evaluated the experimental video database in this paper. These results show that this approach can provide remarkable accuracy for retrieving similar video clips form video database. Furthermore, the estimation of the temporal signature involved in the proposed technique can also be exploited for further application such as video copyright enforcement, content-based video clustering and so on.


## Acknowledgments

The authors would like to acknowledge the National Center for High-performance Computing in providing huge computing resources to facilitate this research. This research has been funded in part by National Science Council of Taiwan, Grant Number: NSC 100-2923-I-492-001.


## References


[1] N. Dimitrova, M. Abdel-Mottaleb, Content-based video retrieval by example video clip. Storage and Retrieval for Image and Video Databases V. Vol. 3022 (1997), p. 59-70.

[2] S.L. Feng, R. Manmatha, V. Lavrenko, Multiple Bernoulli relevance models for image and video annotation. Proceedings of the 2004 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2. (2004), p. 1002-1009

[3] L.-H. Chen, K.-H. Chin, H.-Y.M. Liao, On the query of video database, in: Proceedings of the 5th WSEAS international conference on Multimedia, internet & video technologies, World Scientific and Engineering Academy and Society (WSEAS), Corfu, Greece, 2005, p. 213-218

[4] I.K. Sethi, N. Patel, A Statistical Approach to Scene Change Detection. P Soc Photo-Opt Ins. Vol. 2410 (1995), p.329-338

[5] W.H. Press, Numerical recipes in C : the art of scientific computing, 2nd ed., Cambridge University Press, Cambridge ; New York, 1992.

[6] Y. Xilin, I.C. Octavia, Line-Based Recognition Using A Multidimensional Hausdorff Distance. IEEE transactions on pattern analysis and machine intelligence. Vol.21 (1999), p. 01-916