DOI:10.5392/IJoC.2010.6.3.005

Content similarity matching for video sequence identification

Sang Hyun Kim
School of Electrical Engineering, College of Science and Engineering
Kyungpook National University, Gajang-Dong, Sangju
Kyungpook 742-711, Korea

ABSTRACT

To manage large database system with video, effective video indexing and retrieval are required. A large number of video retrieval algorithms have been presented for frame-wise user query or video content query, whereas a few video identification algorithms have been proposed for video sequence query. In this paper, we propose an effective video identification algorithm for video sequence query that employs the Cauchy function of histograms between successive frames and the modified Hausdorff distance. To effectively match the video sequences with a low computational load, we make use of the key frames extracted by the cumulative Cauchy function and compare the set of key frames using the modified Hausdorff distance. Experimental results with several color video sequences show that the proposed algorithm for video identification yields remarkably higher performance than conventional algorithms such as Euclidean metric, and directed divergence methods.

***Keywords**: Content Similarity, Video Identification, Modified Hausdorff Distance, and Cauchy Function.*

1. INTRODUCTION

Advances in digital media technologies lead to various techniques for indexing, retrieval, and manipulation of multimedia data such as image, video, audio, text, and speech. Especially, the standardization of digital video has accelerated the rapid growth of digital video databases and their efficient management has been one of the important issues. To efficiently manage and utilize digital video, various video indexing and retrieval algorithms have been proposed. A large number of video indexing and retrieval methods have focused on frame-wise query or indexing, whereas a relatively few algorithms have been presented for video sequence or shot identification. In this paper, we propose an efficient algorithm to index the video sequences and to identify the video sequences for video sequence query.

For video indexing such as shot boundary detection, most algorithms in the compressed domain may yield more miss and false shot boundaries than those in the uncompressed domain [1]-[4]. A shot represents a physically temporal interval by record and stop operations of a camera. If the video indexing algorithm detects a lot of false or miss shot boundaries, the accuracy can be reduced, where the accuracy is defined in terms of the numbers of correct and incorrect (false and miss) detections. In this paper, to improve the accuracy and performance of video indexing and sequence identification, we introduce the Cauchy function as a similarity measure between histograms of consecutive frames, which yields a high performance compared with conventional measures.

The key frames extracted from segmented video shots can be used not only for video shot clustering but also for video sequence matching or browsing, where the key frame is defined by the frame that is significantly different from the previous frames. Several key frame extraction algorithms have been proposed, in which similar methods used for shot boundary detection were employed with proper similarity measures. The key frame extraction method using the set theory employing the semi-Hausdorff distance and key frame selection using skin-color and face detection have been also proposed [5]. In this paper, we propose the efficient algorithm to extract key frames using the cumulative Cauchy function measure and compare its performance with that of conventional algorithms.

Content similarity matching for video sequence identification can be performed by evaluating the similarity between data sets of key frames [6]. In this paper, to improve the matching accuracy with the set of extracted key frames we employ the Cauchy function and the modified Hausdorff distance. Experimental results with several color video sequences show that the proposed method yields the high matching performance and accuracy with a low computational load compared with conventional algorithms.

The rest of the paper is organized as follows. Content similarity measures for video identification are presented in Section 2. The proposed similarity matching using key frames and sequence identification algorithm is described in Section 3 and experimental results are shown in Section 5. Finally conclusions are given in Section 5.

* *Corresponding author. E-mail : shk@knu.ac.kr*
Manuscript received Feb. 24, 2010 ; accepted Jul. 14, 2010

2. CONTENT SIMILARITY MEASURES FOR SEQUENCE IDENTIFICAION

We first review the conventional distance measures and compare the algorithmic procedures. The commonly used video identification from video sequences utilizes histogram comparisons, because histograms show less sensitivity to frame changes within a shot and extraction of histograms is computationally efficient compared with the motion based methods. Most common algorithms using histogram comparison [7] include Euclidean metric and directed divergence.

Euclidean metric: The Euclidean metric for histogram is defined by $\sqrt{\sum_j (H_{t+1}(j) - H_t(j))^2}$.

The Euclidean metric is the ordinary distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. By using this formula as distance, Euclidean space becomes a metric space. In Equation, $H_t(j)$ represents the histogram of *t*th frame, and *j* shows *j*th bin.

Directed divergence: The divergence measure is defined by the sum of directed divergences. The directed divergences of histograms are expressed as

$$\sum_j H_{t+1}(j)\log\frac{H_{t+1}(j)}{H_t(j)} + \sum_j H_t(j)\log\frac{H_t(j)}{H_{t+1}(j)}.$$

The proposed algorithm employs the Cauchy function for video identification, in which the Cauchy function is used as a similarity measure [8].

With the given histograms, the similarity can be represented by

$$F(p_j, q_j) = \sum_{j=1}^{N} f(x_j) \tag{1}$$

where p_j and q_j denote the previous and current histogram, respectively, and x_j represents the *j*th bin of the histogram difference, i.e., $x_j = p_j - q_j$, f signifies the Cauchy function employed. The Cauchy function represented by

$$f(x_j) = \log\left(1 + \left(\frac{x_j}{a}\right)^2\right) \tag{2}$$

is employed, where *a* is a parameter that determines the shape of a function and is determined by experimental tuning.

3. PROPOSED CONTENT SIMILARITY MATCHING

3.1. Content Similarity based on the Key Frames Using the Cauchy Function

To retrieve video sequences, we first extract the key frames using the cumulative Cauchy function and evaluate the similarity between video sequences by employing the modified Hausdorff distance between sets of key frames.

In the proposed algorithm, we use the cumulative Cauchy function

$$C(p,q) = \sum_t^{t+k} \left\{ \sum \log\left(1 + \left(\frac{p_j - q_j}{a}\right)^2\right) \right\} \tag{3}$$

to efficiently extract key frames, where *k* denotes the number of accumulated frames that can be varied depending on the criteria for key frame extraction.

The key frames are detected if the cumulative Cauchy function value $C(p_j, q_j)$ between the current frame and the previous key frame is larger than the given threshold. The key frames extracted within video shots can be used not only for representing contents in video shots but for efficiently matching the video sequence with a very low computational load. The cumulative Cauchy function can also be used as a good measure to extract key frames.

3.2. Sequence identification Using the Modified Hausdorff Distance

For matching between video sequences, we employ the modified Hausdorff distance measure [9]. In this paper, to efficiently evaluate the similarity between sets of key frames, we use the modified Hausdorff distance *D*(*S*,*R*) given by

$$D(S,R) = \max[\min_{r \in R}\{d(s_1, r)\}, \ \dots \ , \min_{r \in R}\{d(s_n, r)\}] \tag{4}$$

where $S = \{s_1, \ \dots \ , s_n\}$ represents the set of key frames for the query sequence and $R = \{r_1, \ \dots \ , r_m\}$ signifies the set of key frames for matching sequences, with *n* and *m* denoting the total numbers of elements in sets *S* and *R*, respectively. To improve the matching accuracy, we employ the Cauchy function in Eq. (2) to compute the distance function *d*(*s*,*r*).

The process of sequence identification is shown in Fig. 1. In this figure, the concept for content similarity computation is represented, where key frames are first extracted for the query sequence and the sequence to be matched. The content similarity between the two sequences is computed using proposed modified Hausdorff distance. The computation of content similarity between sequences can be performed by key features with remarkably reduced complexity. If the query shot contains *Q* frames and the matched shot has *M* frames, similarity computation between two shots has been computed $Q \times M$ times. But in proposed method using the modified Hausdorff distance with key frames the computational complexity can be reduced by $K\_q \times K\_m$, where K_q represents the number of key frames for query shot and K_m shows the number of key frames to be matched video sequence. Therefore the reduction of key frames is directed related to computational complexity.

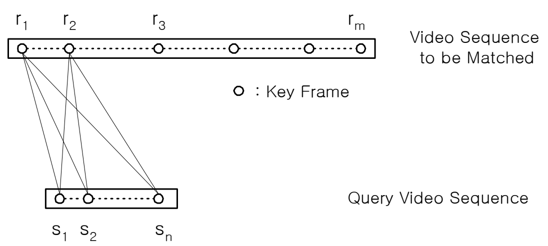

Fig. 1 Process of video sequence identification

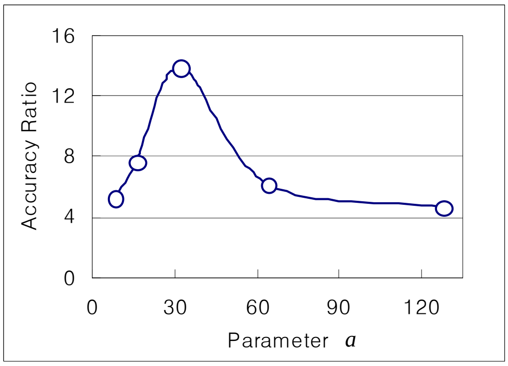

Fig. 2. Performance comparison of video sequence identification for varying parameter values of the Cauchy function.

In case of video contents, key feature can be color, motion, and texture of key frames, whereas audio contents can be employed band energy ratio and signal bandwidth [10].

The normalized similarity metric represents the dissimilarity between the two sequences. The normalized values for 'Matching shots' shows low whereas values for 'Non-matching shots' represents high. The accuracy ratio between 'Matching shots' and 'Non-matching shots' shows the dissimilarity where the algorithm with a high accuracy ratio can match sequences accurately. Simulation results of video sequence identification are shown in Section 4.

3.3. Sequence similarity for color video

For color video sequence identification, the extended Cauchy function is employed for color histograms. The Cauchy function between color histograms is defined by

$$F_c(p_j, q_j) = \sum f_Y(x_j) + \sum f_U(x_j) + \sum f_V(x_j) \qquad (5)$$

where p_j and q_j represent the histograms of previous and current frames, respectively, and x_j signifies the difference between p_j and q_j.

The subscripts *Y*, *U*, and *V* denote each color component of the *YUV* format video. In the case of color video sequences, we have combined the respective histograms for each component. In Eq. (5), the Cauchy function $f(x_j)$ is given in Eq. (2).

The video sequence identification using the modified Hausdorff distance for color video sequences can be applied using Eq. (4). By using the extended Cauchy function, the matching accuracy can be improved.

4. SIMULATION RESULTS AND DISCUSSIONS

To show the effectiveness of the proposed algorithm, we simulate the video sequence identification using color test sequences: animation video sequence consisting of nine shots within 330 frames and real video sequence consisting of 230 shots within 6,170 frames containing large motions and dynamic scene changes with illumination and brightness variation.

To extract the key frames we use two criteria. If both the cumulative Cauchy function value in Eq. (3) and the Cauchy function value in Eq. (2) between the previous key frame and the current frame are larger than threshold values, the candidate frame is extracted as a key frame. Even though the accumulated Cauchy function value is larger than the threshold value, the accumulated Cauchy function can be gradually increased because the Cauchy function value between the previous key frame and the current frame may have the value smaller than the threshold. To extract a key frame both conditions must be satisfied. Once the key frame is extracted, the cumulative Cauchy function is reset to zero. If the thresholds to extract key frames are large, the number of key frames and the computational load can be reduced.

In the proposed method, to determine the parameter a of the Cauchy function in Eq. (2) we first have evaluated the performance for varying parameter values. Fig. 2 shows the ratio for several parameter values of a for the 'Animation' sequence, where the ratio represents the accuracy of video sequence identification.

The matching frames or shots can be determined by thresholding, and the ratio between matching shots and non-matching shots is related to the dissimilarity performance of matching methods. As shown in Fig. 2, the ratio between 'Matching shots' and 'Non-matching shots' is large at $a = 32$, thus parameter a is set to 32 in experiments for test sequences.

To show the performance of video sequence identification methods we have simulated three methods with color video sequences. In experiments, we apply the cumulative Cauchy function value to the Euclidean metric and directed divergence methods as well as to the proposed method. Table 1 shows identification results of the color sequence using the modified Hausdorff distance. In Table 1, query for animation video consists of 14 frames and query for real video consists of 22 frames.

Table 1. Performance comparison of video sequence identification using the modified Hausdorff distance (a) Animation video sequence (b) Real video sequence

(a)

Methods	Matching shots (*A*)	Non-matching shots (*B*)	Accuracy Ratio (*B*/*A*)
Euclidean Metric	0.077	0.561	7.286
Directed Divergence	0.072	0.606	8.417
Proposed Method	0.044	0.610	13.864

(b)

Methods	Matching shots (*A*)	Non-matching shots (*B*)	Accuracy Ratio (*B*/*A*)
Euclidean Metric	0.027	0.187	6.926
Directed Divergence	0.014	0.125	8.929
Proposed Method	0.027	0.369	13.667

'Euclidean metric', 'Directed Divergence', and 'Proposed Method' signify the Euclidean metric method, the directed divergence method, and the proposed method with Cauchy function using the modified Hausdorff distance, respectively.

In Table 1, 'Matching shots' ('Non-matching shots') represents the modified Hausdorff distances between sets of query key frames and the color video sequence to be compared containing 'Matching shots' ('Non-matching shots'). Specially, in proposed method the accuracy ratio between 'Matching shots' and 'Non-matching shots' is large compared with the conventional methods. The proposed algorithm with a large accuracy ratio can match sequences efficiently. Table 1 show that the proposed method using Cauchy function with modified Hausdorff distance can remarkably improve the matching accuracy with a low computational load for sequence identification for color video, compared with Euclidean metric and directed divergence using the modified Hausdorff distance.

In experimental results, the accuracy ratio can have different values for various test sequences, because the ratio is large when the query sequence has large dissimilar contents compared with that of the matched sequence, otherwise is small. In case of similar video content for query video, the modified Hausdorff distance between sets of query key frames and those of compared key frames represents small value. But in case of different video contents the modified Hausdorff distance shows large values. Therefore the improved accuracy ratio in proposed method can reduce the miss detection as well as the false detection in video sequence identification.

In digital content management [11], the proposed method can be applied to video sequence identification efficiently finding any sequence of long database video that shares similar content to query sequence.

5. CONCLUSIONS

This paper proposes the efficient method using the Cauchy function and the modified Hausdorff distance for video sequence identification. The proposed method gives a high accuracy and efficiency, compared with conventional methods such as Euclidean metric and directed divergence methods, with a low computational load. The Cauchy function for color video sequences also improves the accuracy. Experimental results show that the proposed algorithm can efficiently match the video sequences and accurately perform the sequence identification using key frames with the remarkably high accuracy ratio, compared with conventional methods. Further research will focus on the extension of the algorithm for various video sequences containing complex scenes.

ACKNOWLEDGEMENT

This work was supported by Kyungpook National University Research Grant, 2010.

REFERENCES

[1] H. Lu, B. C. Ooi, H. T. Shen, and X. Xue, "Hierarchical indexing structure for efficient similarity search in video retrieval," *IEEE Trans. Knowledge and Data Engineering*, vol. 18, no. 11, Nov. 2006, pp. 1544-1559.

[2] M. Worring and G. Schreiber, "Semantic image and video indexing in broad domains," *IEEE Trans. Multimedia*, vol. 9, no. 5, Aug. 2007, pp. 909-911.

[3] C. Snoek and M. Worring, "Multimedia Event-based video indexing using time intervals," *IEEE Trans. Multimedia*, vol. 7, no. 4, Aug. 2005, pp. 638-647.

[4] V. T. Chasanis, A. C. Likas, and N. P. Galatsanos, "Scene detection in video using shot clustering and sequence alignment," *IEEE Trans. Multimedia*, vol. 11, no. 1, Jan. 2009, pp. 89-100.

[5] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Face-based digital signatures for video retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 4, Apr. 2008, pp. 549-533.

[6] D. P. Mukherjee, S. Kumar, and S. Saha, "Key frame estimation in video using randomness measure of feature point pattern," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 5, May 2007, pp. 612-620.

[7] J. Xu, T. Yamasaki, and K. Aizawa, "Temporal segmentation of 3-D video by histogram-based feature vectors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 19, no. 6, June 2009, pp. 870-881.

[8] N. Sebe, M. S. Lew, and D. P. Huijsmans, "Toward improved ranking metrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-22, Oct. 2000, pp. 1132-1143.

[9] Z.-Q. Zhou and B. Wang, "A modified Hausdorff distance using edge gradient for robust object matching," in *Proc. Int. Conf. Image Analysis and Signal Processing*, Apr. 2009, pp. 250-254.

[10] S. Chu, S. Narayanan, and C. J. Kuo, "Environmental sound recognition with time-frequency audio features," *IEEE Trans. Audio, Speech, and Language Processing*, Aug. 2009, pp. 1142-1158.

[11] B. Lui, D. Chiu, H. Hu, and Y. Zhuang, "Ontology based content management for digital television services," in *Proc. IEEE Int. Conf. e-Business Engineering*, Oct. 2009, pp. 565-570.

Sang Hyun Kim

He received the B.S. and M.S. degrees in electronic and control engineering from Hankuk University of Foreign Studies, in 1997 and 1999, respectively, and the Ph.D. degree in electronic engineering from Sogang University, in 2003. In 2003 and 2004, he worked on the Digital Media Research Laboratory in LG Electronics Inc., as a Senior Research Engineer. In 2004 and 2005, he also worked on the Computing Laboratory at Digital Research Center in Samsung Advanced Institute of Technology, as a Senior Research Member. In 2005 and 2008, he had been with the department of electronic and electrical engineering at Sangju National University as an assistant professor. Since 2008, he has been with the school of electrical engineering at Kyungpook National University as an assistant professor. His current research interests are video indexing, video coding, and computer vision.