An Efficient Video Retrieval Algorithm Using Color and Edge Features

Sang-Hyun Kim*

Abstract

To manipulate large video databases, effective video indexing and retrieval are required. A large number of video indexing and retrieval algorithms have been presented for frame-wise user query or video content query whereas a relatively few video sequence matching algorithms have been proposed for video sequence query. In this paper, we propose an efficient algorithm to extract key frames using color histograms and to match the video sequences using edge features. To effectively match video sequences with low computational load, we make use of the key frames extracted by the cumulative measure and the distance between key frames, and compare two sets of key frames using the modified Hausdorff distance. Experimental results with several real sequences show that the proposed video retrieval algorithm using color and edge features yields the higher accuracy and performance than conventional methods such as histogram difference, Euclidean metric, Battachaya distance, and directed divergence methods.

Keywords : Edge matching, Hausdorff distance, Video retrieval, Video sequence matching

I. Introduction

To efficiently manage and utilize digital media, various video indexing and retrieval algorithms have been proposed. A large number of video indexing and retrieval methods have been proposed, focusing on frame-wise query or indexing, whereas a relatively few algorithms have been presented for video sequence matching or video shot matching. In this paper, we propose the efficient algorithm to match the video sequences for video sequence query.

If the video indexing algorithm shows a lot of false or miss shot boundaries, the accuracy can be reduced, where the accuracy is defined using the numbers of false and miss detections [1-4]. In this paper, to improve the accuracy of video sequence matching, we propose the efficient key frame extraction algorithm using the color histograms of consecutive frames and the sequence matching algorithm using the modified Hausdorff distance with edge features, which yields higher performance than conventional methods.

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 [5]. 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 set theory employing the semi-Hausdorff distance [6] and the key frame selection algorithm using skin-color and face detection [7] have been also proposed. In this paper, we propose the efficient algorithm to extract key frames using the cumulative measure and the distance between key frames, and compare its performance with that of conventional algorithms.

Video sequence matching using key frames can be performed by evaluating the similarity between data sets of key frames. In this paper, to improve the matching efficiency with the sets of extracted key frames we employ color and edge features. Experimental results with several video sequences show that the proposed methods give better matching accuracy than conventional algorithms. The accuracy ratio of the average normalized value of the non-matching shots to that of the matching shot is employed as the performance measure of matching methods.

The rest of the paper is structured as follows. The distance measures for video indexing are briefly discussed in Section II. The proposed algorithm for video sequence matching is presented in Section III and experimental results are shown in Section IV. Finally conclusions are given in Section V.

*상주대학교 전자전기공학부
논문 번호 : 2006-1-14　　접수 일자 : 2006. 1. 12
심사 일자 : 2006. 1. 12　　수정 완료 : 2006. 1 . 26

II. Distance Measure for Video Indexing

The commonly used video indexing methods utilize 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 [8]. Most common algorithms using histogram comparison include histogram difference [1], Euclidean metric [9], Battachaya distance [10], and directed divergence [11]. The color model used in this paper is YUV model.

1. Histogram Difference

The histogram difference is defined by

$$\sum_j H_{t+1}(j) - H_t(j)\,| \tag{1}$$

where $H_t(j)$ signifies the histogram in the jth bin, $0 \leq j \leq 255$, with the subscrip t denoting the t th frame. Bin signifies the gray level range of the histogram representation.

2. Euclidean Metric

The Euclidean metric for histogram is defined by

$$\sqrt{\sum_j (H_{t+1}(j) - H_t(j))^2}. \tag{2}$$

3. Battachaya Distance

The Battachaya distance with respect to histogram is used to estimate the distance between histogram features, defined by

$$-\,ln\,(\sum_j \sqrt{H_{t+1}(j)H_t(j)}\,) \tag{3}$$

where j represents the bin index of the histogram.

4. Directed Divergence

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

$$\sum_j H_{t+1}(j) log\frac{H_{t+1}(j)}{H_t(j)} + \sum_j H_t(j) log\frac{H_t(j)}{H_{t+1}(j)}. \tag{4}$$

III. Proposed Video Sequence Matching

To match video sequences, we first extract key frames using the cumulative measure and the distance between key frames, and evaluate the similarity between two video sequences by employing the modified Hausdorff distance between two sets of key frames: one extracted from the query sequence and the other from the video sequence to be matched. Fig. 1 shows the block diagram of the proposed algorithm, where the sets of key frames are first extracted for the query sequence and the sequence to be matched. The similarity between two sets of key frames is computed using the modified Hausdorff distance. The proposed video sequence matching algorithm consists of three steps: key frame extraction, key frame matching, and video sequence matching.

1. Key Frame Extraction Using the Cumulative Measure and the Distance between Key Frames

In the proposed algorithm, we use the cumulative measure based on the histogram difference to extract

$$C = \sum_t^{t+k} (\sum_j H_{t+1}(j) - H_t(j)\,|) \tag{5}$$

candidate key frames efficiently, where k denotes the total number of accumulated frames that can be varied depending on the criteria for key frame extraction [11]. Note that application of the cumulative concept over k frames to (1) yields the cumulative measure (5).

The key frames are detected if the cumulative value C between the current frame and the previous key frame is larger than the given threshold, and the histogram difference (1) between the previous key frame and the current frame is larger than 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 sequences varying threshold with a very low computational load [12].

2. Key Frame Matcing Using Edge Features

To match video sequences efficiently, we perform edge matching. To extract edge features we employ the Marr-Hildreth edge detector [13].

To efficiently match edges of two key frames, we propose a novel approach. The edge matching procedures in Y (luminance) component are summarized as follows.

1) Extract the query edge image $E_q(i,j)$ from the query key frame and the matched edge image $E_m(i,j)$ from the matched key frame.

2) Obtain the 'common edge image $E_c(i,j)$' from $E_q(i,j)$ and $E_m(i,j)$ using 'AND' operation.

3) Calculate the edge matching rate (EMR) defined by

$$EMR = \frac{NEP in\, E_C(i.j)}{Min\,\{NEP in\, E_q(i,j),\, NEP in\, E_m(i.j)\}} \tag{6}$$

where NEP represents the number of edge pixels.

The EMR is used to calculate the similarity between key frames. Note that to reduce a computational load edge matching is performed only on key frames rather than on all the frames. Edge matching is applied to video sequence matching efficiently and the experimental results are shown in Section IV.

3. Video Sequence Matching Using the Modified Hausdorff Distance

For matching between video sequences, we employ the modified Hausdorff distance measure [14]. In this paper, to efficiently evaluate the similarity between two sets of key frames, we use the modified Hausdorff distance $D_{SR}(k)$ given by

$$D_{SR}(k) = max[min\{d(s_1,r_k),d(s_1,r_{k+1})\},...,min\{d(s_n,r_k),d(s_n,r_{k+1})\}] \quad (7)$$

where $S=\{s_1,...,s_n\}$ represents the set of key frames for the query sequence and $R=\{r_1,...,r_m\}$ signifies the set of key frames for the sequence to be matched, with n and m denoting the total numbers of elements in sets S and R, respectively [11]. In (7), $D_{SR}(k)$ represents the modified Hausdorff distance between the kth and (k+1)th key frames of the sequence to be matched, with $d(s,r)$ signifying the

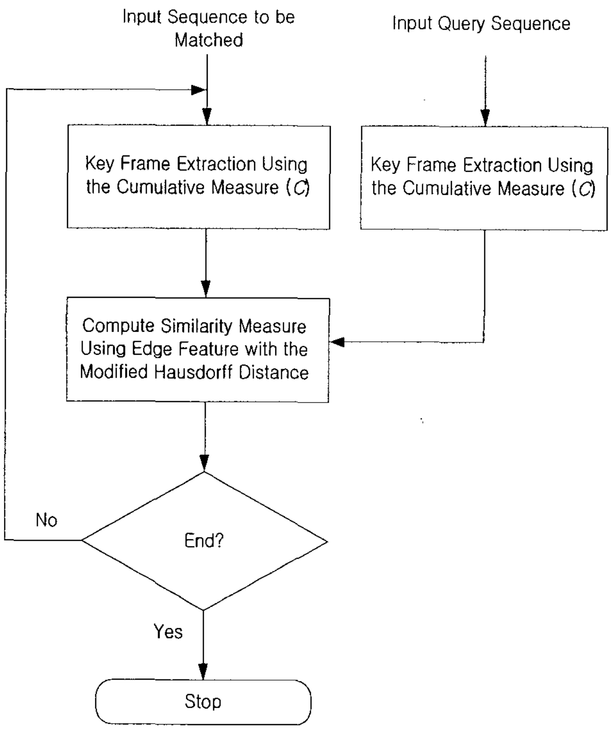


Fig. 1. Flowchart of the proposed video sequence matching algorithm.

EMR in (6) obtained from edge matching between two key frames.

The normalized modified Hausdorff distance $NMHD(i)$ can defined by

$$NMHD(i) = \frac{D_{SR}(KFN(i))}{max(D_{SR}(k))} \quad (8)$$

where $KFN(i)$ signifies the key frame number for the ith frame.

Fig. 1 shows the flowchart of the proposed algorithm, where key frames are extracted for the query sequence and the sequence to be matched and the similarity between key frames is computed. The normalized similarity metric (8) represents the dissimilarity between the two sequences: the normalized values for 'Matching shots' are small whereas those for 'Non-matching shots' are large. The ratio of the average normalized values for 'Non-matching shots' to that for 'Matching shots' represents the separation capability. It is noted that the algorithm with a large ratio can match sequences accurately. Simulation results of video sequence matching are presented in Section IV.

Fig. 2. Key frames within the 'News Article' video sequence.

IV. Simulation Results and Discussions

1. Key Frame Extraction

To show the effectiveness of the proposed algorithm, we simulate color video sequence matching using two test sequences: 'Animation' sequence consisting of nine shots within 330 frames and the 'News Article' sequence consisting of 57 shots within 12,920 frames containing large motions and dynamic scene changes as shown in Fig. 2.

To extract the key frames we use two criteria. If both the cumulative value in (5) and the histogram difference value in (1) 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 value is larger than the threshold value, the accumulated value can be gradually increased because the histogram difference 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 (1) and (5) must be satisfied [11]. Once the key frame is extracted, the cumulative value 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.

2. Video Sequence Matching

To show the performance of video sequence matching, five methods are simulated with color video sequences. In experiments for key frame extraction, we apply the cumulative concept to all of form similarity measures: histogram difference (1), Euclidean metric (2), Battachaya distance (3), and directed divergence (4). Table 1(a) shows matching results of the color 'Animation' sequence using the modified Hausdorff distance. In experiments for video sequence matching, we assumed that the video sequence to be matched includes the query sequence and the query sequence has similar frame length to same shot within the video sequence to be matched. In Table 1(a), the query sequence, the same as shot 2 in the 'Animation' sequence, consists of frames from 49 to 83, and 'Histogram difference', 'Euclidean metric', 'Battachaya distance', and 'directed divergence' signify the histogram difference method [1], the Euclidean metric method [9], the Battachaya distance method [10], and the directed divergence method [11], respectively. Note that the modified Hausdorff distance measure is applied to all methods.

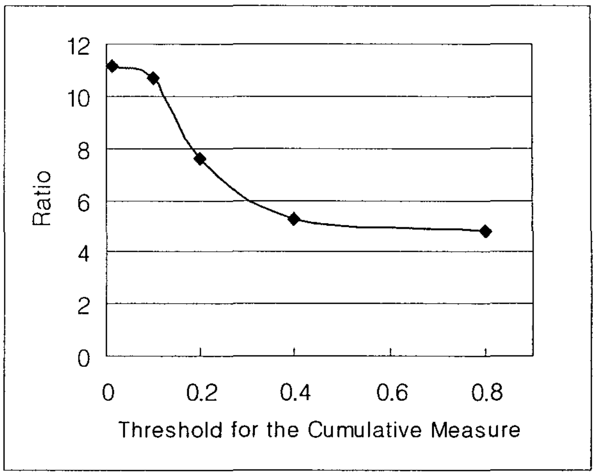
Table 1. Performance comparison of video sequence matching using the modified Hausdorff distance. (a) 'Animation' sequence (330 frames). (b) 'News Article' sequence (12,920 frames).

(a)

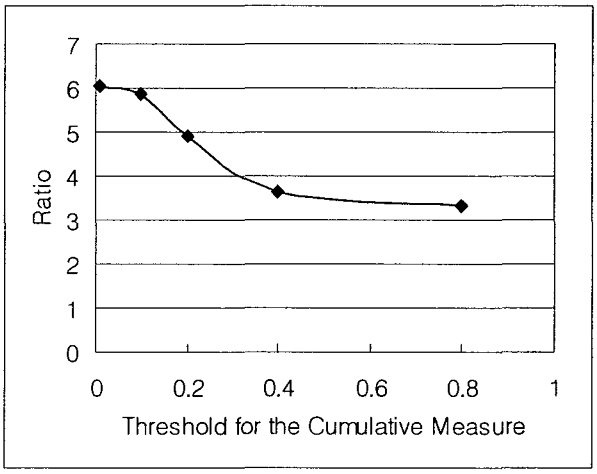
Methods	Matching shots (A)	Non-matching shots (B)	Ratio (B/A)
Histogram Difference	0.123	0.634	5.154
Euclidean Metric	0.077	0.561	7.286
Battacharya Distance	0.076	0.562	7.395
Directed Divergence	0.072	0.606	8.417
Proposed Method	0.044	0.610	13.864

(b)

Methods	Matching shots (A)	Non-matching shots (B)	Ratio (B/A)
Histogram Difference	0.128	0.456	3.563
Euclidean Metric	0.090	0.413	4.589
Battacharya Distance	0.089	0.413	4.640
Directed Divergence	0.076	0.329	4.329
Proposed Method	0.180	0.883	5.861



(a)



(b)

Fig. 3. Ratio (B/A) of the proposed video sequence matching algorithm as a function of the cumulative measure. (a) 'Animation' sequence. (b) 'News Article' sequence.

In Table 1, 'Matching shots' ('Non-matching shots') represents the average of normalized modified Hausdorff distances (8) between the set of query key frames and the set of color video sequence to be compared over the interval that contains (does not contain) 'Matching shots'. Table 1(b) shows matching results using the modified Hausdorff distance for the 'News article' sequence, in which the query sequence is composed of frames from 3708 to 3745, extracted from shot 16 between 3705 and 3747 in the 'News article' sequence. As shown in Table 1(b), the normalized values of 'Matching shots' are small whereas those of 'Non-matching shots' are large.

As shown in Table 1, in the proposed method using edge and color features the ratio between 'Matching shots' and 'Non-matching shots' is larger than the conventional methods using only color histograms. That is, the algorithm using edge features can reduce the number of false matchings, whereas the conventional video sequence matching methods may yield a lot of false matchings. The conventional methods show wide variations for 'Non-matching shots'. In contrast, the proposed method employing edge features shows small fluctuations for 'Non-matching shots'. The proposed video sequence matching also shows large accuracy ratio for both query sequence extracted from 'Animation' and 'News Article' sequences compared with that of the conventional methods.

Fig. 3 shows the ratio (B/A) of the proposed video sequence matching algorithm as a function of the threshold of cumulative value with 'Animation' sequence and 'News Article' sequence. As shown in Fig. 3, the ratio decreases as the threshold increases. That is, to reduce the computational load, the number of key frames can be reduced by increasing the threshold for the cumulative value, however yielding low performance. Therefore, in experiments, the threshold is set to 0.1, which satisfies both the performance and the computational load.

Table 1 shows that the proposed method using color and edge features can improve the accuracy for color video sequence matching, compared with conventional measures such as the histogram difference, Euclidean metric, Battachaya distance, and directed divergence.

In MPEG-7 standardization, any specific video sequence matching method is not described. The proposed method can be applied to MPEG-7 standard by using the MPEG-7 descriptors [9].

V. Conclusions

This paper proposes the efficient video sequence matching method using color and edge features with the modified Hausdorff distance. It gives a higher accuracy and efficiency than conventional methods such as the histogram difference, Euclidean metric, Battachaya distance, and directed divergence methods, with a similar computational load. The combination of color histograms and edge features improves the accuracy of video sequence matching. Experimental results with real video sequences show that the proposed algorithm can successfully extract key frames and match video sequences efficiently, showing a higher accuracy than the conventional methods. Further research will focus on the extension of the algorithm for various video sequences containing complex scenes.

References

[1] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Trans. Circuits Syst. Video Technol.*, vol. CSVT-10, pp. 533-544, Feb. 2000.

[2] A. Hanjalic, "Shot-boundary detection: Unraveled and resolved?," *IEEE Trans. Circuits Syst. Video Technol.*, vol. CSVT-12, no. 2, pp. 90-105, Feb. 2002.

[3] V. Kobla, D. Doermann, and K. I. Lin, "Archiving, indexing, and retrieval of video in compressed domain," in *Proc. SPIE Conf. Multimedia Storage and Archiving Systems*, Boston, MA, Nov. 1996, vol. 2916, pp. 78-89.

[4] J. Meng, Y. Juan, and S.-F. Chang, "Scene change detection in a MPEG compressed video sequence," in *Proc. IS&T/SPIE Symposium Digital Video Compression: Algorithms and Technologies*, San Jose, CA, Feb. 1995, vol. 2419, pp. 14-25.

[5] M. M. Yeung and B. Liu, "Efficient matching and clustering of video shots," in *Proc. IEEE Int. Conf. Image Processing*, Washington, D.C., Oct. 1995, vol. 1, pp. 338-341.

[6] H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. CSVT-9, pp. 1269-1279, Dec. 1999.

[7] F. Dufaux, "Key frame selection to represent a video," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, Canada, Sept. 2000, vol. 2, pp. 275-278.

[8] A. Akutsu, Y. Tonomura, H. Hashimoto, and Y. Ohba, "Video indexing using motion vectors," in *Proc. SPIE Conf. Visual Communications and Image Processing*, Boston, MA, Nov. 1992, vol. 1818, pp. 1522-1530.

[9] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. CSVT-11, pp. 703-715, June 2001.

[10] S. Liapis, E. Sifakis, and G. Tziritas, "Color and/or texture segmentation using deterministic relaxation and fast marching algorithms," in *Proc. Int. Conf. Pattern Recognition*, Barcelona, Spain, Sept. 2000, vol. 3, pp. 617-620.

[11] S. H. Kim and R.-H. Park, "An efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. CSVT-12, pp. 592-596, July 2002.

[12] M. R. Naphade, M. M. Yeung, and B.-L. Yeo, "A novel scheme for fast and efficient video sequence matching using compact signatures," in *Proc. IS&T/SPIE Conf. Storage and Retrieval for Media Databases 2000*, San Jose, CA, Jan. 2000, vol. 3972, pp. 564-572.

[13] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Society of London B*, 1980, vol. 207, pp. 187-217.

[14] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff

distance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-15, pp. 850-863, Sept. 1993.

Sang-Hyun Kim
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. Since 2005, he has been with the department of electronic and electrical engineering at Sangju National University as a full time lecturer. His current research interests are video retrieval, video coding, and computer vision.