http://dx.doi.org/10.5392/IJoC.2016.12.1.001

An Efficient Video Retrieval Algorithm Using Key Frame Matching for Video Content Management

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

ABSTRACT

To manipulate large video contents, 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 that extracts key frames using color histograms and matches the video sequences using edge features. To effectively match video sequences with a 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 real sequence show that the proposed video sequence matching algorithm using edge features yields the higher accuracy and performance than conventional methods such as histogram difference, Euclidean metric, Battachaya distance, and directed divergence methods.

Key Words: *key frame matching, modified Hausdorff distance, video indexing, video retrieval, video content management*

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]. 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 a 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 [2]. Many 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 and the key frame selection algorithm using skin-color and face detection have been also proposed. In this paper, we propose the efficient key frame extraction algorithm that employs 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. As the performance measure of matching methods, we introduce the accuracy ratio of the average normalized value of the non-matching shots to that of the matching shot.

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.

II. DISTANCE MEASURES 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

* *Corresponding author, Email: shk@knu.ac.kr*
Manuscript received Oct. 08, 2015; revised Dec. 15, 2015; accepted Jan. 15, 2016

motion based methods. Most common algorithms using histogram comparison include histogram difference, Euclidean metric, Battachaya distance, and directed divergence [3].

A. Histogram Difference

The histogram difference is defined by

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

where $H_t(j)$ signifies the histogram in the j th bin, $0 \le j \le 255$, with the subscript t denoting the t th frame and bin signifying the gray level range of the histogram representation.

B. Euclidean Metric

The Euclidean metric between histograms is defined by

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

C. Battachaya Distance

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

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

where j represents the bin index of the histogram.

D. Directed Divergence

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

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

III. PROPOSED VIDEO RETRIEVAL ALGORITHM

To match video sequences, we first extract key frames using the cumulative measure and the distance between key frames, then 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. 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.

A. 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

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

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

The key frames are detected when two criteria are satisfied: 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 the 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 with a very low computational load [4].

B. Key Frame Matching Using Edge Features

To match video sequences efficiently, we perform edge matching. To extract edge features we employ the Marr-Hildreth edge detector. Edges in the current frame $I(i, j)$ satisfy

$$\nabla^2 G * I(i, j) = 0 \quad (6)$$

where $*$ denotes the convolution operator. The Gaussian function G is defined by

$$G(r) = \exp\left(\frac{-r^2}{2\sigma^2}\right) \quad (7)$$

where $r = \sqrt{i^2 + j^2}$ and σ represents a Gaussian spread parameter, standard deviation. $\nabla^2 G$ signifies the Laplacian of Gaussian [5]

$$\nabla^2 G(r) = \frac{r^2 - 2\sigma^2}{\sigma^4} \exp\left(\frac{-r^2}{2\sigma^2}\right). \quad (8)$$

The edge density and thickness can be controlled by σ.

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

$$\text{EMR} = \frac{\text{NEP in } E_c(i, j)}{\text{Min}\{\text{NEP in } E_q(i, j), \text{NEP in } E_m(i, j)\}} \quad (9)$$

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.

C. Video Sequence Matching Using the Modified Hausdorff Distance

For matching between video sequences, we employ the modified Hausdorff distance measure. 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

$$\max[\min\{d(s_1,r_k),d(s_1,r_{k+1})\},\ \ldots\ ,\min\{d(s_n,r_k),d(s_n,r_{k+1})\}] \quad (10)$$

where $R=\{r_1,\ \ldots\ ,r_K\}$ signifies the set of key frames for the sequence to be matched and $S=\{s_1,\ \ldots\ ,s_N\}$ represents the set of key frames for the query sequence, with K and N denoting the total numbers of elements in sets S and R, respectively. $D_{SR}(k)$ represents the modified Hausdorff distance between the kth and (k+1)th key frames of the sequence to be matched, with $d(s,r)$ ($s\in S$ and $r\in R$) signifying the EMR in (9) obtained from edge matching between two key frames [6].

Let indices t and k specify the time index and the key frame index of the video sequence to be matched, respectively, $1\le t\le T$ and $1\le k\le K<T$ with T and K representing the total number of the test frames and the total number of its key frames, respectively. Let f_{kt} denote the function that maps the key frame index k to the time index t. f_{kt}^{-1} represents the inverse function of f_{kt}. The normalized modified Hausdorff distance $NMHD(t)$ can be defined by

$$NMHD(t)=\frac{D_{SR}(k)}{\max_k D_{SR}(k)} \quad \text{for} \quad f_{kt}^{-1}(k)\le t<f_{kt}^{-1}(k+1). \quad (11)$$

Note that $NMHD(t)$ is constant for the time index t that corresponds to the key frame index interval between k anf k+1.

The normalized similarity metric (11) 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 *NMHD* 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.

IV. SIMULATION RESULTS AND DISCUSSIONS

A. Key Frame Extraction

To show the effectiveness of the proposed algorithm, we simulate video sequence matching using color test sequence: 'Animation' sequence consisting of nine shots within 330 frames with K=9 and T=330 containing large motions and dynamic scene changes.

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 frame is not regarded as a key frame since the accumulated value can be gradually increased with the histogram difference value between the previous key frame and the current showing the value smaller than the threshold. To extract a key frame, both conditions with respect to (1) and (5) must be satisfied. 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.

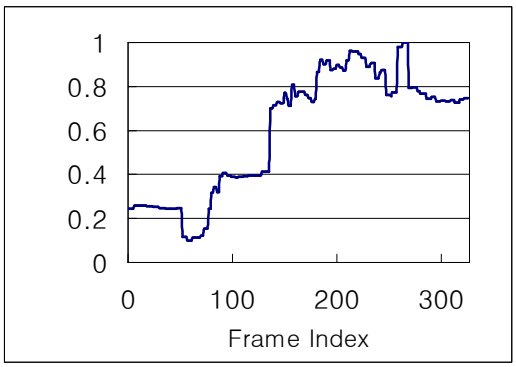
B. Video Sequence Matching

To show the performance of video retrieval algorithm 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. Table I and Fig. 1 show matching results of the color 'Animation' sequence using the modified Hausdorff distance.

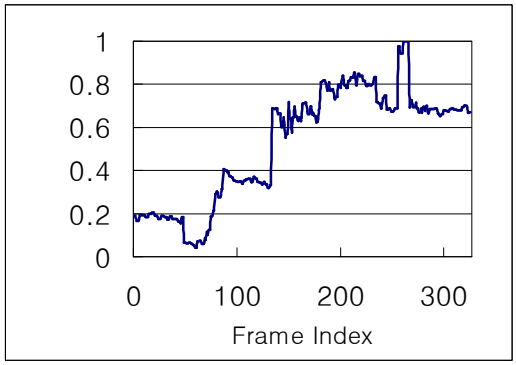
In experiments for video sequence matching, we assumed that the video sequence to be matched includes the query sequence and the query sequence has similar frame length to same shot within the video sequence to be matched.

Table I. Performance comparison of video sequence matching using the modified Hausdorff distance.

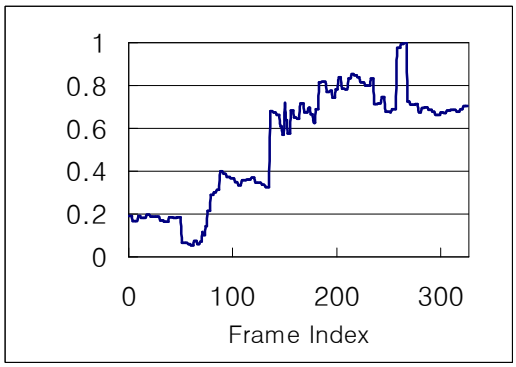
Methods	Matching shots (A)	Non-matching shots (B)	Ratio (B/A)
Histogram Difference	0.123	0.634	5.154
Euclidean Metric	0.077	0.561	7.286
Battacharya Distance	0.076	0.562	7.395
Directed Divergence	0.072	0.606	8.417
Proposed Method	0.073	0.782	10.712



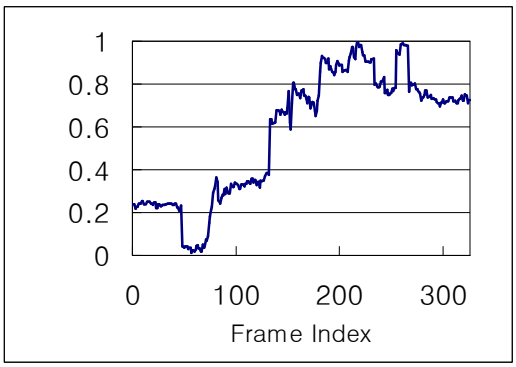
(a)



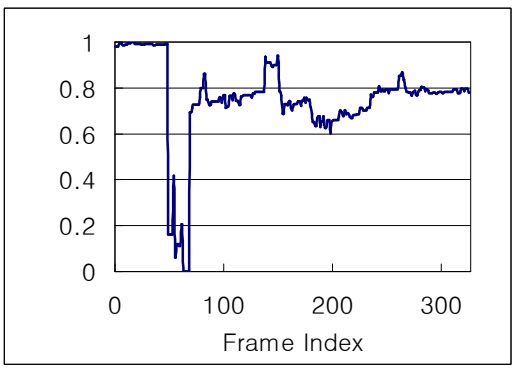
(b)



(c)



(d)



(e)

Fig. 1. Performance comparison of video sequence matching as a function of the frame number.
(a) Histogram Difference, (b) Euclidean Metric, (c) Battachaya Distance, (d) Directed Divergence, (e) Proposed Method

In Table I and Fig. 1, 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, the Euclidean metric method, the Battachaya distance method, and the directed divergence method, respectively. Note that the modified Hausdorff distance measure is applied to all methods.

Fig. 1 shows the normalized modified Hausdorff distance value between the set of query key frames and that of the video sequence to be compared as a function of the frame number. The normalized value is obtained by (11), resulting in the normalized value between 0 and 1. In Table I, 'Matching shots' ('Non-matching shots') represents the average of normalized modified Hausdorff distances (11) 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'.

As shown in Table I and Fig. 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 matching, whereas the conventional video sequence matching methods may yield a lot of false matching. 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' (see Fig. 1).

The proposed video sequence matching also shows large accuracy ratio for both query sequence extracted from sequences compared with that of the conventional methods.

Fig. 2 shows the ratio (B/A) of the proposed video sequence matching algorithm as a function of the threshold of cumulative value. As shown in Fig. 2, 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.

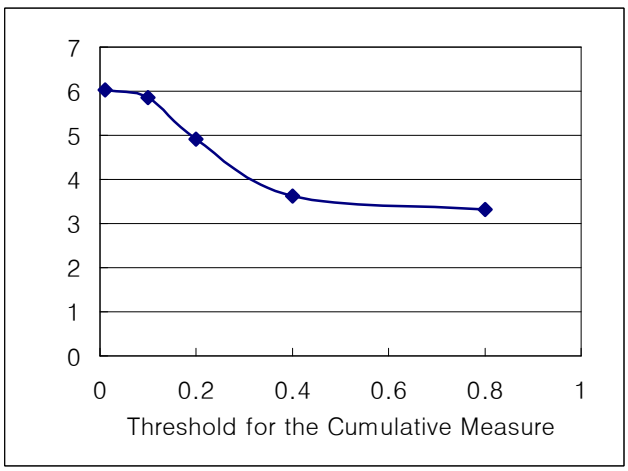


Fig. 2. Ratio (B/A) of the proposed video sequence matching algorithm as a function of the cumulative measure.

Therefore, in experiments, the threshold is set to 0.1, which satisfies both the performance and the computational

load. Table I 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 for video content management and automatic monitoring system efficiently with low computational complexity [7]-[10].

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. 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] X. Wen, L. Shao, W. Fang, and Y. Xue, "Efficient feature selection and classification for vehicle detection," IEEE Trans. Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 508-517, Mar. 2015.

[2] G. Luis, D. Tuia, G. Moser, and C. Gustau, "Multimodal classification of remote sensing images: A review and future directions," Proc. of IEEE, vol. 103, no. 9, pp. 1560-1584, Sep. 2015.

[3] Z. A. Jaffery and A. K. Dubey, "Architecture of noninvasive real time visual monitoring system for dial type measuring instrument," IEEE Sensors Journal, vol. 13, no. 4, pp. 1236-1244, Apr. 2013.

[4] Y. Yang, Z. Zha, Y. Gao, X. Zhu, and T. Chua, "Exploiting web Images for semantic video indexing via robust sample-specific loss," IEEE Trans. Multimedia, vol. 16, no. 6, pp. 1677-1689, Aug. 2014.

[5] 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, pp. 89-100, Jan. 2009.

[6] J. Geng, Z. Miao, and X.-P. Zhang, "Efficient heuristic methods for multimodal fusion and concept fusion in video concept detection," IEEE Trans. Multimedia, vol. 17, no. 4, pp. 498-511, Apr. 2015.

[7] H Yan, K. Paynabar, and H. Shi, "Image-based process monitoring using low-rank tensor decomposition," IEEE Trans. Automation Science and Engineering, vol. 12, no. 1, pp. 216-227, Jan. 2015.

[8] Y. Yin, Y. Yu, and R. Zimmermann, "On generating content-oriented geo features for sensor-rich outdoor video search," IEEE Trans. Multimedia, vol. 17, no. 10, pp. 1760-1772, Oct. 2015.

[9] S. Youm, Y. Jeon, S. Park, and W. Zhu, "RFID-based automatic scoring system for physical fitness testing," IEEE Systems, vol. 9, no. 2, pp.326-334, Jun. 2015.

[10] S. Ferdousi, F. Dikbiyik, M. F. Habib, M.Tomatone, and B. Mukherjee, "Disaster-aware datacenter placement and dynamic content management in cloud networks," IEEE Optical Communication and Networking, vol. 7, no.7, pp. 681-694, Jul. 2015.

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. Since 2005, he has been with the school of convergence & fusion system engineering at Kyungpook National University as an associate professor. His current research interests are video indexing, video coding, and computer vision.