Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 4, pp. 891 ~898

An Efficient Video Retrieval Algorithm Using Luminance Projection

Sang Hyun Kim[1)]

Abstract

An effective video indexing is required to manipulate large video databases. Most algorithms for video indexing have been commonly used histograms, edges, or motion features. In this paper, we propose an efficient algorithm using the luminance projection for video retrieval. To effectively index the video sequences and to reduce the computational complexity, we use the key frames extracted by the cumulative measure, and compare the set of key frames using the modified Hausdorff distance. Experimental results show that the proposed video indexing and video retrieval algorithm yields the higher accuracy and performance than the conventional algorithm.

Keywords : Hausdorff distance, key frame extraction, luminance projection, video indexing, video retrieval

1. Introduction

To efficiently manage and utilize digital media, various video indexing and retrieval algorithms have been proposed. Most video indexing and retrieval methods have focused on the frame-wise query or indexing, whereas there have been a relatively few algorithms for video sequence matching or video shot matching. In this paper, we propose the efficient algorithms to index the video sequences and to match the sequences for video sequence query.

Most algorithms used for video indexing may show the low accuracy in the compressed domain [1], which leads to false or miss segmentation. In this paper, to improve the accuracy and performance of video indexing and segmentation, we propose the efficient method using the luminance projection, which yields a higher performance than the conventional method.

---

1) Samsung Advanced Institute of Technology
E-mail: shkim2002@hotmail.com

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. The key frame is defined as the frame that is significantly different from the previous frames [2]. The key frames can be extracted by employing similar methods used in shot boundary detection with proper similarity measures, and several algorithms have been proposed. The key frame extraction method using set theory employing the semi-Hausdorff distance [3] and key frame selection using skin-color and face detection [4] have been also proposed. In this paper, we propose the efficient algorithm to extract key frames using the cumulative measure and compare its performance with that of the conventional algorithm.

Video sequence matching using key frames extracted from each shot can be performed by evaluating the similarity between each data set of key frames. In this paper, to improve the matching efficiency we propose the method using the modified Hausdorff distance to match the set of extracted key frames. Experimental results show that the proposed methods show the higher matching performance and accuracy than the conventional algorithm.

2. PROPOSED ALGORITHMS FOR VIDEO SEQUENCE MATCHING

The proposed algorithm employs luminance projection for video indexing. Luminance projection can be performed along two directions: horizontal and vertical directions. Luminance projection for each direction can be represented by

$$LPH(i) = \sum_{j=0}^{J-1} f(i,j) \tag{1}$$

$$LPV(j) = \sum_{i=0}^{I-1} f(i,j) \tag{2}$$

where $f(i,j)$ represents the intensity at (i, j), and LPH and LPV signify the luminance projections along horizontal and vertical directions, respectively. I and J represent vertical and horizontal size, respectively. The similarity distance between two projection values can be obtained by

$$DH(t) = \sum_{i} |LPH_{t-1}(i) - LPH_t(i)| \tag{3}$$

$$DV(t) = \sum_{j} |LPV_{t-1}(j) - LPV_t(j)| \tag{4}$$

where DH(t) and DV(t) represent distances for horizontal and vertical projections, respectively, and t and t-1 signify the current and previous frames, respectively.

For horizontal and vertical distances, the similarity distance can be determined by

$$Dis = \max(DH(t), DV(t)) \tag{5}$$

where Dis denotes the similarity distance between two frames.

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

2.1. Key Frame Extraction Using the Cumulative Measure

In our algorithms, we use the cumulative measure

$$C = \sum_{t=0}^{k-1} \{\max(DH(t), DV(t))\} \tag{6}$$

to extract key frames efficiently, where k denotes the number of accumulated frames. 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. The extracted key frames within video shots can be used not only for representing contents in video shots but for matching the video sequence efficiently with very low computational complexity [5]. The cumulative measure can also be used to extract key frames efficiently.

2.2. Video Sequence Matching Using the Modified Hausdorff Distance

For matching between video sequences, we employ the modified Hausdorff distance measure. Given two finite point sets $A = \{a_1, \cdots, a_u\}$ and $B = \{b_1, \cdots, b_v\}$ the Hausdorff distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{7}$$

where u and v represent the total numbers of elements of sets A and B, respectively, and $h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||$, with $|| \cdot ||$ denoting the norm on the points of A and B [6].

In this paper, to efficiently evaluate the similarity between sets of key frames, we use the modified Hausdorff distance $D(S,R)$ between sets of key frames

$$D(S, R) = \max[\min_{r \in R} \{d(s_1, r)\}, \cdots, \min_{r \in R} \{d(s_n, r)\}] \tag{8}$$

where $S = \{s_{1,} \cdots, s_n\}$ represents the set of key frames for the query sequence and $R = \{r_1, \cdots, r_m\}$ signifies the set of key frames for matching sequences, with n and mdenoting the total numbers of elements in sets S and R, respectively. In the proposed algorithm, we employ the difference measure for the luminance projection values as a distance function. Simulation results of video sequence matching are shown in Section 3.2.

3. SIMULATION RESULTS

3.1. Key Frame Extraction

To extract the key frames we use two criteria. If the cumulative value in Eq. (6) and the luminance projection value between the previous key frame and the current frame are larger than threshold values, the candidate frame can be extracted as a key frame. The process of key frame extraction is shown in Fig. 1. In Fig. 1, even though the accumulated value is larger than the threshold value, the accumulated value gradually increases because the luminance projection value between the previous key frame and the current frame is smaller than the threshold. Therefore, both conditions must be satisfied to accurately detect a key frame. If the key frame is extracted, the accumulated valueis reset to zero.

3.2. Video Sequence Matching

To show the effectiveness of the proposed algorithm, we simulate the color video sequence matching for the real color video sequences: 'News Article' consisting of 12,920 frames and 'Music Video' consisting of 6,170 frames containing large motions and dynamic scene changes as shown in Fig. 2. Table 1 and Fig. 3 show matching results for the histogram comparison method [1] and the luminance projection method using the modified Hausdorff distance for the 'News Article' video sequence. In Table 1 and Fig. 3, the difference measure has been employed as a histogram comparison method. Fig. 3 shows the normalized matching distance between the query key frame and the video sequence to be compared, as a function of the frame index. In experiments of video sequence matching, we apply the cumulative measure equally to the histogram method and the luminance projection method. The normalized value is small with matched shots, whereas it is large with different shots. The luminance projection method shows the largedifference between matching shot and dissimilar shot. In Table 1, 'Matching shot' represents theaverage value of the distance between the query key frame and the video sequence to be compared when the key frame is matched, whereas 'Dissimilar shot' represents the one when the key frame is not matched.

In Table 1, the ratio represents the accuracy of video sequence matching. The matching frames or shots can be determined by thresholding, and the ratios between matching shots and dissimilar shots are related to the dissimilarity performance of matching methods. Table 1 shows that the proposed method can improve the matching accuracy with the low computational complexity for video sequence matching, compared with the histogram comparison method.

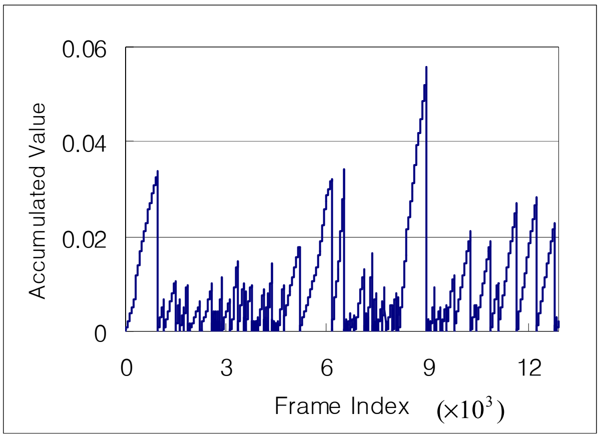


Fig. 1.Accumulated value as a function of the frame index in key frame extraction.

(a) News Article. (b) Music Video.

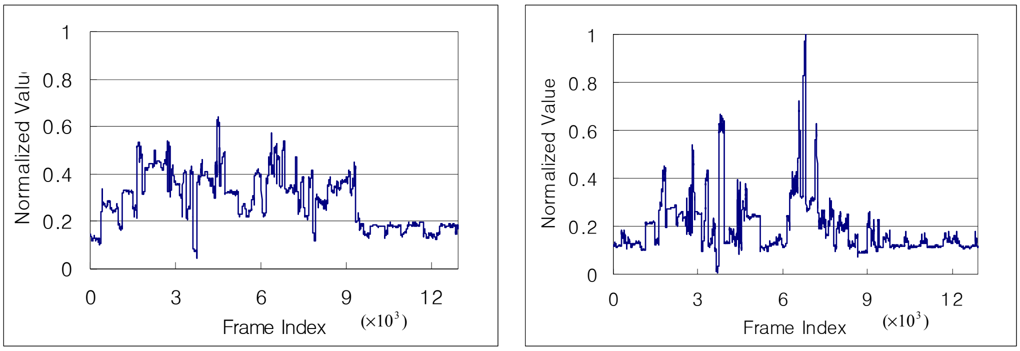
Fig. 2. Key frames within the real color video sequences.

Table 1. Performance comparison of matching methods for the real video sequence (News Article).

Method	Average value		Ratio (B/A)
	Matching shot (A)	Dissimilar shot (B)	
Histogram Method	0.082	0.292	3.561
Luminance Projection	0.043	0.195	4.535

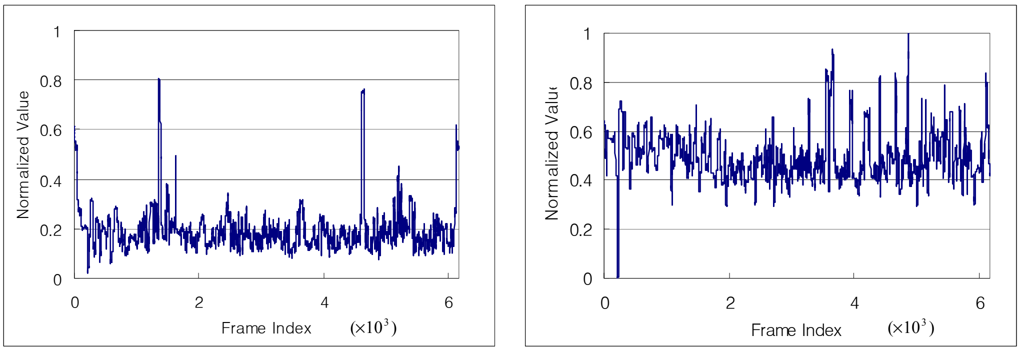
Table 2. Performance comparison of matching methods for the real video sequence (Music Video).

Method	Average value		Ratio (B/A)
	Matching shot (A)	Dissimilar shot (B)	
Histogram Method	0.036	0.196	5.444
Luminance Projection	0.004	0.498	124.5



(a) Histogram method. (b) Luminance projection method.

Fig. 3. Performance comparison of real video sequence matching as a function of the frame index (News Article).



(a) Histogram method. (b) Luminance projection method.

Fig. 4. Performance comparison of real video sequence matching as a function of the frame index (Music Video).

Table 2and Fig. 4 show matching results using the histogram comparison method and the luminance projection method for the 'Music Video' sequence. In Table 2, the proposed method using luminance projection also shows a largermatching accuracy than the histogram comparison method.

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 color descriptors [7].

4. CONCLUSIONS

This paper proposes the efficient video retrieval method using the luminance projection with the modified Hausdorff distance. The proposed method gives the higher accuracy and efficiency than the conventional method such as the histogram comparison method, with the similar computational complexity. Experimental results show that the proposed algorithm can extract key frames and match video sequences efficiently, and shows the higher accuracy than the conventional method. Further research will focus on the semantic video sequence indexing and on the verification with various video sequences containing complex video scenes.

References

1. U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Trans. Circuits and Systems for Video Technology*, vol. CSVT-10, no. 1, pp. 533-544, Feb. 2000.
2. 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.
3. H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, vol. CSVT-9, no. 8, pp. 1269-1279, Dec. 1999.
4. 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.
5. 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. SPIE Conf. Storage and Retrieval for Media Databases 2000*, San Jose, CA, Jan. 2000, vol. 3972, pp. 564-572.
6. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-15, no. 9, pp. 850-863, Sep. 1993.
7. B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color

and texture descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. CSVT-11, no. 6, pp. 703-715, June 2001.

[received date : Jul. 2004, accepted date : Sep. 2004]