효율적인 비디오 유사도 측정을 위한 휘도 투영모델
Luminance Projection Model for Efficient Video Similarity Measure

김상현
Sang-Hyun Kim

요 약

비디오 데이터들의 효율적 색인과 검색을 위해서는 비디오 시퀀스의 유사도 측정방법이 매우 중요한 요소이다. 본 논문은 비디오 시퀀스에 대한 효율적인 유사도 측정을 위해 휘도 성분 투사법을 제안한다. 기존의 알고리즘들이 히스토그램, 윤곽선, 움직임등과 같은 특성을 사용한 반면 본 논문에서 제안한 알고리즘은 휘도 성분을 투사하는 방법을 사용하여 비디오 유사도 특성을 효율적으로 나타낼 수 있다. 비디오 데이터의 효율적인 색인과 계산량 감소를 위해 누적된 유사도에 의해 추출된 키프레임들을 이용하여 비디오 시퀀스의 유사도를 구하고 수정된 하우스도르프 거리를 사용하여 키프레임 묶음들의 유사도를 측정하였다. 실험결과 제안한 휘도투시법을 사용한 비디오 색인 기법이 유사도 특성에서 기존의 특성을 사용한 방법에 비해 확연한 정확도 및 성능 차이를 보였다.

Abstract

The video similarity measure is very important factor to index and to retrieve for video data. In this paper, we propose the luminance projection model to measure the video similarity efficiently. Most algorithms for video indexing have been commonly used histograms, edges, or motion features, whereas in this paper, the proposed algorithm is employed an efficient measure using the luminance projection. To index effectively the video sequences and to decrease the computational complexity, we calculate video similarity using 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 luminance projection model yields the remarkable accuracy and performance than the conventional algorithm.

Keywords : luminance projection, video similarity measure, video indexing, key frame extraction, modified Hausdorff distance, video sequence matching

I. Introduction

To manage efficiently 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 video similarity measure using luminance projection algorithm 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.

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,

*School of Electrical Engineering, College of Science and Engineering Kyungpook National University
접수 일자 : 2008. 9. 29   수정 완료 : 2009. 4. 28
게재확정일자 : 2009. 4. 29

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.

II. Proposed luminance projection algorithm for video similarity

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

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

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

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

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

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

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

$$Dis = \max(DH(t), DV(t)) \quad (5)$$

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

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

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 $A=\{a_1,\ldots,a_u\}$ and $B=\{b_1,\ldots,b_v\}$ the Hausdorff distance is defined as

$$H(A,B)=\max(h(A,B),h(B,A)) \quad (7)$$

where u and v represent the total numbers of elements of sets A and B, respectively, and $h(A,B)=\max\min \| a-b \|$ with $\| \bullet \|$ denoting the norm on the points of A and B [6].

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

$$D(S,R) = \max\left[\min_{r\in R} d(s_1,r),\ldots,\min_{r\in R} d(s_n,r)\right] \quad (8)$$

where $S=\{s_1,\ldots,s_n\}$ represents the set of key frames for the query sequence and $R=\{r_1,\ldots,r_m\}$ 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. The proposed algorithm uses the difference measure for the luminance projection model as a distance function. Simulation results of video sequence matching are shown in Section 3.2.

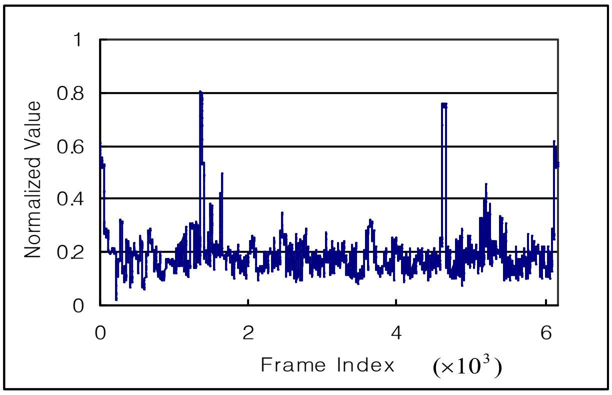
III. Simulation results

3.1. Key frame extraction

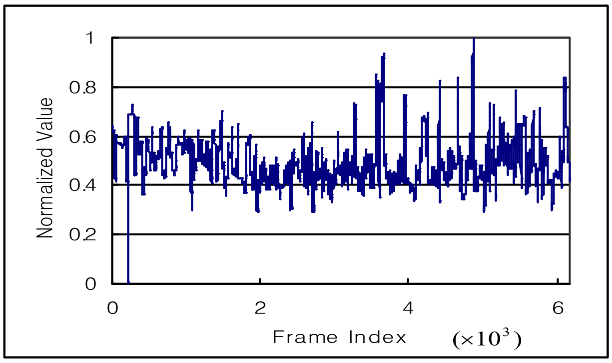
To extract the key frames the conventional methods use the histogram difference between frames or cumulative value of histogram difference. In this paper, to extract the key frames efficiently, we use two criteria. If the cumulative value in Eq. (6) and the luminance projection value between the previous key frame and the current frame are larger than threshold values, the candidate frame can be extracted as a key frame. 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 value is reset to zero. If the threshold for accumulated value is low, many key frames can be extracted.

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 sequence: 'Music Video' consisting of 6,170 frames containing large motions and dynamic scene changes. Table 1 and Fig. 1 show matching results for the histogram comparison method [1] and the luminance projection method using the modified Hausdorff distance for the 'Music Video' video sequence. In Table 1 and Fig. 1, the difference measure has been employed as a histogram comparison method. Fig. 1 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 Fig. 1(a), the normalized value represents the normalized similarity distance using the histogram difference, and in Fig. 1(b), the normalized value represents the normalized similarity distance using luminance projection in Eq. (5). In experiments of video sequence matching, we apply the cumulative measure 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 large difference between matching shot and dissimilar shot.



(a)



(b)

Fig. 1. Performance comparison of real video sequence matching as a function of the frame index (Music Video). (a) Histogram method, (b) Luminance projection method.

Table 1. Performance comparison of matching methods for the real video sequence (Music Video).

Methods	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

In Table 1, 'Matching shot' represents the average 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 remarkably with the low computational complexity for video sequence matching, compared with the conventional histogram comparison method. As shown in Table 1 and Fig. 1, the proposed luminance projection model yields the remarkable accuracy with large ratio between matching shot and dissimilar shot, and the difference characteristics for matching frames compared with the conventional histogram method.

In compressed domain, the proposed method also can be used to improve the matching accuracy with the DC images in video stream.

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].

IV. Conclusions

This paper proposes the efficient video sequence matching method using the luminance projection model 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 remarkable high 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] J.-R. Ding and J.-F. Yang, "Adaptive group-of-pictures and scene change detection methods based on existing H.264 advanced video coding information," *IET Image Processing*, vol. 2, pp. 85-94, Apr. 2008.

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

[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, vol. 2, pp. 275-278, Sept. 2000.

[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, vol. 3972, pp. 564-572, Jan. 2000.

[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] F. Falchi, C. Gennaro, and P. Savino, "Efficient video stream filtering," *IEEE Multimedia,* vol. 15, pp. 52-62, 2008.

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. 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.