

# Efficient Video Retrieval Scheme with Luminance Projection Model

Sang Hyun Kim<sup>\*</sup>

<sup>1</sup>School of Convergence & Fusion System Engineering, Kyungpook National University

## 휘도투시모형을 적용한 효율적인 비디오 검색기법

김상현<sup>\*</sup>

<sup>1</sup>경북대학교 융복합시스템공학부

**Abstract** A number of video indexing and retrieval algorithms have been proposed to manage large video databases efficiently. The video similarity measure is one of most important technical factor for video content management system. In this paper, we propose the luminance characteristics 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 similarity measure using the luminance projection. To index the video sequences effectively and to reduce 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 improved accuracy and performance than the conventional algorithm such as the histogram comparison method, with the low computational complexity.

**요약** 대용량 비디오 데이터베이스들을 효율적으로 관리하기 위해 많은 비디오 색인 및 검색 알고리즘들이 제안되고 있다. 비디오 콘텐츠 관리 시스템에서 비디오 유사도 측정방법은 가장 중요한 기술적 요소 중 하나이다. 본 논문에서는 비디오 유사도를 효율적으로 측정하기 위해 휘도특성 모델을 제안한다. 비디오 색인에 관한 대부분의 알고리즘들이 공통적으로 히스토그램, 윤곽선, 움직임 특성을 사용한 반면 본 논문에서 제안한 알고리즘은 휘도투시를 사용한 효율적인 유사도 측정법을 적용하였다. 비디오 시퀀스의 효율적인 색인과 계산량 감소를 위해 누적된 유사도에 의해 추출된 키프레임 들을 이용한 비디오 유사도를 계산하고 수정된 하우스도르프 거리를 사용하여 키프레임 묶음을 비교하였다. 실험결과 제안한 휘도투시 모델이 적은 계산량으로 기존의 히스토그램 비교법을 사용한 알고리즘에 비해 현저히 향상된 정확도 및 성능을 보였다.

**Keywords** : luminance projection, video similarity measure, video indexing, key frame extraction, modified Hausdorff distance, video retrieval

## 1. 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

This research was supported by Kyungpook National University Research Fund, 2012.

<sup>\*</sup>Corresponding Author : Sang Hyun Kim(Kyungpook National University)

Tel: +82-10-2702-4138 email: shk@knu.ac.kr

Received October 16, 2015

Revised (1st December 2, 2015, 2nd December 3, 2015)

Accepted December 4, 2015

Published December 31, 2015

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, and several algorithms have been proposed. The key frame extraction method using set theory employing the semi-Hausdorff distance and key frame selection using skin-color and face detection have been also proposed [3]. 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 [4]. 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 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

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

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

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  $A=\{a_1, \dots, a_u\}$  and  $B=\{b_1, \dots, b_v\}$  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  $\|\cdot\|$  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

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

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. The distance function can represent the similarity efficiently between key frame sets. 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.

## 3. Simulation results

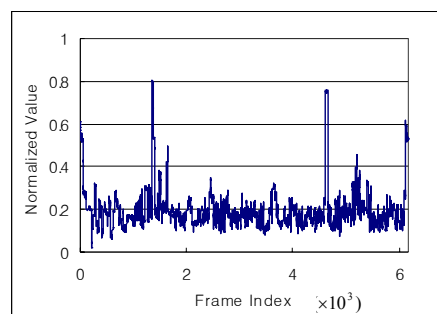
### 3.1 Key frame extraction

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 value 0.2, the candidate frame can be extracted as a key frame. Even though the

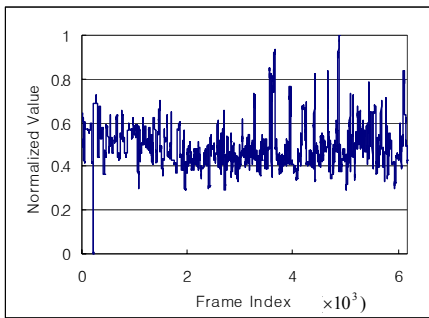
accumulated value is larger than the threshold value, the accumulated value gradually increases because the luminance projection value between the previous key frame and the current frame can be smaller than the threshold. Therefore, both conditions must be satisfied to accurately detect a key frame. 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, frame index at the horizontal axis represents sequential order of frames. 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.



(a)



(b)

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

Table 1. Performance comparison of matching methods for the real video sequence.

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

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.

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. In case of histogram method, the accuracy performance is very sensitive by threshold values, whereas the proposed luminance projection method is robust for threshold values due to the large accuracy ratio.

Table 1 shows that the proposed method can improve the matching accuracy ratio remarkably with the low computational complexity for video sequence matching, compared with the conventional histogram comparison

method. The proposed matching algorithm with Hausdorff distance can reduce computational complexity efficiently with large matching accuracy. As shown in Table 1 and Fig. 1, the proposed luminance projection model yields the remarkable accuracy with large accuracy 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 and video content management efficiently [7-8].

#### 4. 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 low 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] 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, 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.

---

**Sang Hyun Kim**

[Regular Member]



- Feb. 2003 : Department of Electronic Engineering, Sogang University, PhD.
- Mar. 2003 ~ May 2004 : Digital Media Research Laboratory, LG Electronics Inc., Senior Research Engineer.
- Jun. 2004 ~ Feb. 2005 : Computing Laboratory, Samsung Advanced Institute of Technology, Senior Research Member.
- Mar. 2005 ~ current : School of Convergence & Fusion System Engineering, Kyungpook National University, Associate Professor.

<Research Interests>

Image Coding, Video Indexing, Computer Vision