Video Content Indexing using Kullback-Leibler Distance

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

ABSTRACT

In huge video databases, the effective video content indexing method is required. While manual indexing is the most effective approach to this goal, it is slow and expensive. Thus automatic indexing is desirable and recently various indexing tools for video databases have been developed. For efficient video content indexing, the similarity measure is an important factor. This paper presents new similarity measures between frames and proposes a new algorithm to index video content using Kullback-Leibler distance defined between two histograms. Experimental results show that the proposed algorithm using Kullback-Leibler distance gives remarkable high accuracy ratios compared with several conventional algorithms to index video content.

Keywords: *Video Content Indexing, Content Management, Kullback-Leibler Distance and Cumulative Measure.*

1. INTRODUCTION

Increase in digital video databases has made it necessary to index and segment automatically video data. So it is required to automatically extract the scene changes and key frames from large video databases for fast and efficient video content indexing and retrievals.

To index video content, video segmentation is a first step, in which video is divided into a number of shots. A shot represents a physically temporal interval by record and stop operations of a camera. The boundaries between video shots and the process of segmenting video are called scene changes and scene change detection, respectively.

Main research on scene change detection can be grouped into two categories. The first category uses uncompressed video frames whereas the second one directly detects scene changes in compressed video form [1], [2]. Recently, the research on the compressed domain is more active because most video data are stored in compressed forms such as joint photographic experts group (JPEG) and motion picture experts group (MPEG). The key point of video indexing is to define an effective similarity measure between frames, and the absolute difference methods such as frame difference or histogram difference methods have been commonly used. In this paper, we propose the novel video content indexing in which a new similarity measure is introduced.

A Kullback-Leibler distance between two histograms extracted from uncompressed or compressed video content is proposed. Simulation results show that the proposed video content indexing can improve the accuracy performance such as a ratio of the peak detection value to the value obtained from frames without scene changes.

The key frames extracted from segmented video shots can be used not only for video shot clustering but also for video content retrieval or browsing. The key frames can be extracted by employing similar methods used in video content indexing with proper similarity measures. The key frame is defined by the frame which is significantly different from the previous frames [3]. The key frame extraction method using set theory employing the semi-Hausdorff distance [4] and key frame selection using skin-color and face detection [5] have been proposed. In this paper, we propose the efficient algorithm to extract key frames using the cumulative directed divergence measure and compare its performance with that of conventional algorithms.

Video content indexing 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 accuracy we propose the modified Hausdorff distance using the directed divergence function to match the set of extracted key frames. Experimental results show that the proposed method shows the high matching performance and accuracy compared with conventional algorithms.

* *Corresponding author. E-mail : shk@knu.ac.kr*
Manuscript received Aug.28, 2009 ; accepted Oct.01, 2009

2. ALGORITHMS FOR VIDEO CONTENT INDEXING

2.1. Content Indexing Measures

The content indexing from video uses frame difference methods based on pixel-by-pixel comparison or on the histogram of image values over the entire frame or a set of covering subregions. Most algorithms rely on histogram comparisons, because the characteristics of histograms show less sensitivity to the frame changes within a shot and also computationally efficient to extract the histogram for each frame. Several conventional algorithms have been presented [6].

1) *Pixel-by-pixel comparison.* The absolute intensity difference between corresponding pixels of successive frames is computed. A camera break is assumed whenever the percentage of pixels whose gray level difference is larger than a given value is greater than a prespecified threshold. The same method can be applied to color images, e.g., by taking the average RGB differences.

2) *Histogram comparison.* The area between the intensity distribution of two successive frames is computed: $\sum| H_{t+1}(j) - H_t(j) |$, where $H_t(j)$ signifies a histogram in the j th bin, with the subscript t denoting the t th frame. A camera break is assumed whenever the number of area is greater than a given threshold. Another commonly used formula is the so-called histogram intersection.

3) *Kolmogorov-Smirnov test.* It is based on calculating the cumulative distributions $C_1(x)$ and $C_2(x)$ of the pixel luminance in two successive frames and on measuring the maximum absolute difference between them. If the distributions are approximated by histograms, a defective estimate of the distance is obtained by $D = \max | C_1(j) - C_2(j) |$, where j denotes the j th bin.

4) *Net comparison algorithm.* A set of nonoverlapping image regions is considered and the selected regions in successive frames are compared by computing the difference of the average luminance values. Whenever the number of regions showing a gray level difference larger than a given threshold exceeds a predefined level, a camera break is assumed.

2.2. Proposed Algorithm

To index video content, we use the Kullback-Leibler distance [7], where the Kullback-Leibler distance derived from directed divergences between two pdfs p and q is given by

$$K(p,q) = \int p(x)\log\frac{p(x)}{q(x)}dx + \int(1-p(x))\log\frac{(1-p(x))}{(1-q(x))}dx \qquad (1)$$

The Kullback–Leibler distance originates from statistics where it is used to quantify differences between two probability distributions or densities. The video content indexing using histogram is comparing two probability distributions between frames. Therefore the Kullback–Leibler distance can be used for video indexing measure with high performance. In information theory, it is also known as the divergence, discrimination or relative entropy. In general, the Kullback–Leibler distance can be used to quantify differences in shape of strictly positive sequences of which the sum equals one. It also can be used in speech synthesis to quantify the differences between spectral envelopes [8]. In this paper, we employ the the Kullback–Leibler distance as a high performance measure to index video content efficiently.

Let p and q be probability density functions (pdfs), then the directed divergences of p and q are the two integrals in the functional $F(p,q)$ expressed as

$$F(p,q) = A\int q(x)\log\frac{q(x)}{p(x)}dx + B\int p(x)\log\frac{p(x)}{q(x)}dx \qquad (2)$$

in which the sum, with $A = B = 1$, signifies the divergence. In (1), if $A = 0$ and $B = 1$, then $F(p,q)$ represents the cross entropy. The terms such as expected weights of evidence, cross entropy, and discrimination information are also regarded as the directed divergence. Now we show that in computing the functional $F(p,q) = \int f(p(x),q(x))dx$, we may take the function $f(u,v) = Av\log(v/u) + Bu\log(u/v)$, for $u>0$ and $v>0$. That is, the directed divergence is derived from the equation satisfying the following properties.

Theorem 1. Suppose F satisfies the additivity and finiteness axioms, then F also satisfies the linear-invariance axioms.

Theorem 2. Suppose F satisfies the linear-invariance axioms, then f has the form

$$f(u,v) = g(u/v)p + D(u-v)\log v \qquad (3)$$

for some functions g and constant D.

Theorem 3. Suppose f has the form of (2) and F satisfies the positivity axiom or semiboundedness axiom, then $D = 0$; f has the form

$$f(u,v) = g(u/v)u \qquad (4)$$

for some functions g.

By *Theorems* 1, 2, and 3, we can prove *Theorem* 4.

Theorem 4. Suppose F satisfies the additivity, positivity, and finiteness axioms, then F has the form (1) for some constants $A, B \geq 0$, not both equal to zero.

To reduce the indexing complexity, we extract the key frames using the cumulative measure and evaluate the similarity between video content by employing the Kullback-Leibler distance. In our algorithm, we use the cumulative measure $C(p,q)$ defined by

$$C(p,q) = W \cdot \int_t^{t+k}\{K(p,q)\}dt \qquad (5)$$

to extract key frames efficiently, where W represents the constant and k denotes the number of accumulated frames.

The key frames are detected if the Kullback-Leibler distance 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 indexing the video content efficiently with a very low computational load. The cumulative measures can also be used as a measure to extract key frames.

To perform the content indexing for color video, the extended Kullback-Leibler distance is employed for color histograms.

3. SIMULATION RESULTS

3.1. Simulation Results for Uncompressed Data

We first simulate video sequences in the uncompressed domain. The results for uncompressed video content are shown in table 1. It is noted that the Kullback-Leibler distance yields the very small value for frames without scene changes. Table 1 shows the performance comparison of video indexing for frames without luminance changes, such as histogram intersection, histogram difference, and Kullback-Leibler distance.

Table 1. Performance comparison of video content indexing (without luminance changes).

Methods	Average value		Accuracy Ratio (*A*/*B*)
	with shot (*A*)	without shot (*B*)	
Histogram Intersection	0.211	0.073	2.89
Histogram Difference	0.581	0.025	23.24
Kullback-Leibler Distance	0.542	0.001	542.00

In table 1, 'with shot' represents the average value for frames with same shot whereas 'without shot' signifies the average value for frames with different shot. It is noted that the accuracy ratio for Kullback-Leibler distance gives remarkably high ratios.

Table 2 shows the performance comparison when the test frames contain the luminance changes, in which 'with shot (*BC*)' denotes the average value for frames with scene changes as well as brightness changes. It is noted that the Kullback-Leibler distance methods give higher accuracy ratios than other conventional methods. Therefore, they can be good measures to detect not only the scene changes but also illumination changes.

Table 2. Performance comparison of video content indexing (with luminance changes).

Methods	Average value			Accuracy Ratio (*C*/*E*)
	With shot (*C*)	With shot (*BC*)	Without shot (*E*)	
Histogram Intersection	0.248	0.330	0.073	3.39
Histogram Difference	0.478	0.436	0.025	19.12
Kullback-Leibler Distance	0.390	0.330	0.001	390.00

(where *BC* represents Brightness Changes)

3.2. Simulation Results for Compressed Data

We also simulate the video content indexing in the compressed domain using DC images [9], in which DC sequences consist of a set of DC coefficients.

Let $I(i,j,t)$ be the intensity of the original image at position (i,j) in the tth frame. The DC image $\bar{I}(u,v,t)$, $0 \le u \le U-1$, $0 \le v \le V-1$, is obtained by

$$\bar{I}(u,v,t) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I((u-1)M+m,(v-1)N+n,t) \quad (6)$$

where $I((u-1)M+m,(v-1)N+n,t)$ represents the intensity value at position (m,n) of the (u,v)th subblock in the t th input frame, and $\bar{I}(u,v,t)$ denotes that at position (u,v) in the $U \times V$ DC image of the t th frame. Note that $U \times V$ DC image $\bar{I}(u,v,t)$ is obtained by averaging the $MU \times NV$ image $I(i,j,t)$ and decimating by a factor of M (N) along the horizontal (vertical) direction, where $M \times N$ denotes the subblock size. The DC image can represent the coarse intensity and be used to extract the global frame characteristics.

Table 3 shows the result for DC sequences and it is noted that the accuracy ratio for the Kullback-Leibler distance also gives higher values than other conventional methods in compressed DC sequences.

In the case of direct indexing without decoding DC sequences for compressed MPEG sequences and JPEG databases, the proposed method using Kullback-Leibler distance can show a remarkable performance with high accuracy ratio.

Table 3. Performance comparison of video content indexing for DC sequences.

Methods	Average value			Accuracy Ratio (*M*/*N*)
	With shot (*M*)	With shot (*BC*)	Without shot (*N*)	
Histogram Intersection	0.808	0.717	0.619	1.30
Histogram Difference	0.534	0.506	0.147	3.63
Kullback-Leibler Distance	0.380	0.369	0.027	14.07

(where *BC* represents Brightness Changes)

4. CONCLUSIONS

To index video content efficiently, this paper proposes the new algorithm using the Kullback-Leibler distance defined between two histograms. The similarity measure using the Kullback-Leibler distance gives high accuracy and efficiency, compared with conventional methods such as an absolute difference method, with the similar computational complexity. Experimental results show that the proposed algorithm is fast and effective in indexing video content. Further research will

focus on the semantic video content indexing and the verifications with various video contents.

ACKNOWLEDGEMENT

This work was supported by the Kyungpook National University Research Grant, 2009.

REFERENCES

[1] M. Worring and G. Schreiber, "Semantic image and video indexing in broad domains," *IEEE Trans. Multimedia, vol. 9, no. 5, Aug.* 2007, pp. 909-911.

[2] C. Snoek and M. Worring, "Multimedia Event-based video indexing using time intervals," *IEEE Trans. Multimedia, vol. 7, no. 4, Aug.* 2005, pp. 638-647.

[3] D. P. Mukherjee, S. Kumar, and S. Saha, "Key frame estimation in video using randomness measure of feature point pattern," *IEEE Trans. Circuits and Systems for Video Technology, vol. 17, no. 5,* May 2007, pp. 612-620.

[4] 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,* Dec. 1999, pp. 1269-1279.

[5] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Face-based digital signatures for video retrieval," *IEEE Trans. Circuits and Systems for Video Technology, vol. 18, no. 4,* Apr. 2008, pp. 549-533.

[6] 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,* Jan. 2009, pp. 89-100.

[7] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Information Theory, vol. IT-26, no. 1,* Jan. 1980, pp. 26-37.

[8] E. Klabbers and R. Veldhuis, "Reducing audible spectral discontinuities," *IEEE Trans. Speech Audio Processing, vol. 9,* Jan. 2001, pp. 39–51.

[9] H. Lu, B. C. Ooi, H. T. Shen, and X. Xue, "Hierarchical indexing structure for efficient similarity search in video retrieval," *IEEE Trans. Knowledge and Data Engineering, vol. 18, no. 11,* Nov. 2006, pp. 1544-1559.

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. 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.