An Efficient Algorithm for Video Sequence Matching Using the Modified Hausdorff Distance and the Directed Divergence

Sang Hyun Kim, *Student Member, IEEE*, and Rae-Hong Park, *Senior Member, IEEE*

***Abstract*—To manipulate large video database, effective video indexing and retrieval are required. A large number of video retrieval algorithms have been presented for framewise user query or video content query, whereas a few video-sequence matching algorithms have been investigated. In this paper, we propose an efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence of histograms between successive frames. To effectively match the video sequences with a low computational load, we use the key frames extracted by the cumulative directed divergence and compare the set of key frames using the modified Hausdorff distance. Experimental results with color video sequences show that the proposed algorithms for video sequence matching yield better performance than conventional algorithms such as histogram difference, histogram intersection, and Chi-square test methods.**

***Index Terms*—Directed divergence, Hausdorff distance, key frame extraction, video indexing, video sequence matching.**

I. Introduction

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

For efficient video indexing and retrieval, 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 [1]–[5]. In this paper, to improve the accuracy and performance of video indexing and segmentation, we employ the directed divergence between histograms of consecutive fames, which yields better performance than conventional measures. To reduce the computational complexity, we match the key frames of the query sequence and the sequence to be matched.

Manuscript received March 2, 2001; revised October 22, 2001. This work was supported by the Brain Korea 21 Project. This paper was recommended by Associate Editor J. R. Ohm.

The authors are with the Department of Electronic Engineering, Sogang University, Seoul 100-611, Korea (e-mail: rhpark@sogang.ac.kr).

Publisher Item Identifier 10.1109/TCSVT.2002.800512.

The key frames 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 [6]. Several 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 the set theory employing the semi-Hausdorff distance [7] and key frame selection using skin-color and face detection [8] have been also proposed.

In video sequence matching, the key frames are usually extracted from the shot or scene unit sequence after shot or scene boundary detection, where it is difficult to match the video sequence directly. Among direct sequence-matching algorithms, the computational complexity to extract the key frames and to match the key frames within the video sequence is large. In this paper, we propose the new matching algorithm using the modified Hausdorff distance to reduce the computational complexity and the novel approach to computation of the similarity between key frames to improve the accuracy performance. The proposed algorithm extracts key frames efficiently using the cumulative directed divergence measure. We compare its performance with that of conventional algorithms. In the proposed method, the key frames are extracted from the video sequence directly, thus the additional computational overhead is not required compared with shot-based key frame extraction methods.

Video sequence matching can be performed by evaluating the similarity between data sets of key frames. In this paper, to improve the matching efficiency with the set of extracted key frames we introduce the modified Hausdorff distance and the directed divergence. Experimental results with color video sequences show that the proposed method yields better performance and accuracy than conventional algorithms.

The rest of the paper is organized as follows. Proposed key frame extraction and video sequence matching algorithms are described in Section II and experimental results are shown in Section III. Finally, conclusions are given in Section IV.

II. Algorithms for Video Sequence Matching

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

A. Key Frame Extraction Using the Cumulative Directed Divergence

The divergence measure is defined by the sum of directed divergences [9]. 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

$$F(p,q) = A\int q(x)\log\frac{q(x)}{p(x)}dx + B\int p(x)\log\frac{p(x)}{q(x)}dx \quad (1)$$

where the sum, with $A = B = 1$, signifies the divergence [10]. If $A = 0$ and $B = 1$, then $F(p,q)$ represents the cross entropy [11]. The terms such as expected weights of evidence, cross entropy and discrimination information are also regarded as the directed divergence. The functional $F(p,q)$ satisfies the additivity, positivity and finiteness axioms [10].

In the proposed algorithm, we use the cumulative directed divergence measure defined by

$$C(p,q) = W\cdot\int_t^{t+k}\left\{\int q(x)\log\frac{q(x)}{p(x)}dx + \int p(x)\log\frac{p(x)}{q(x)}dx\right\}dt \quad (2)$$

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 directed divergence value $C(p,q)$ 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 also for matching video sequences with a very low computational load [12].

The proposed video sequence matching using key frames can reduce the computational complexity by up to 30%, in which the computational complexity depends on the threshold value of the cumulative measure.

B. Video Sequence Matching Using the Modified Hausdorff Distance

For matching between video sequences, we employ the modified Hausdorff distance measure. Let $A = \{a_1, \ldots, a_i\}$ and $B = \{b_1, \ldots, b_j\}$ be two finite point sets, where i and j represent the total numbers of elements of sets A and B, respectively. Then the Hausdorff distance is defined as

$$H(A,B) = \max(h(A,B), h(B,A)) \quad (3)$$

where the directed measure $h(A,B)$ is defined by $\max_{a\in A}\min_{b\in B}\|a - b\|$, with $\|\cdot\|$ denoting the norm on the points of A and B [13].

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

$$D(S,R) = \max\left[\min_{r\in R}\{d(s_1,r)\}, \min_{r\in R}\{d(s_2,r)\}, \ldots, \min_{r\in R}\{d(s_n,r)\}\right] \quad (4)$$

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 [14]. To improve the accuracy, we employ the directed divergence in (1) as the distance function $d(s,r)$. Simulation results of video sequence matching are shown in Section III.

C. Color Video Sequence Matching

To perform color video matching, the extended directed divergence is employed for color histograms. The directed divergence $F_c(p,q)$ based on color histograms is defined as

$$\begin{aligned}F_c(p,q) =& \alpha\left(\sum_x q_Y(x)\log\frac{q_Y(x)}{p_Y(x)} + \sum_x p_Y(x)\log\frac{p_Y(x)}{q_Y(x)}\right)\\ &+ \beta\left(\sum_x q_U(x)\log\frac{q_U(x)}{p_U(x)} + \sum_x p_U(x)\log\frac{p_U(x)}{q_U(x)}\right)\\ &+ \gamma\left(\sum_x q_V(x)\log\frac{q_V(x)}{p_V(x)} + \sum_x p_V(x)\log\frac{p_V(x)}{q_V(x)}\right)\end{aligned} \quad (5)$$

where p and q represent the histograms of previous and current frames, respectively, and the subscripts Y, U and V denote each color component in YUV color format. In experiments, the weighting factors α, β and γ are set to 1. By using the extended divergence measure for color histograms, the accuracy can be improved.

III. Simulation Results and Discussions

To show the effectiveness of the proposed algorithm, we simulate the color video sequence matching for the animation sequence consisting of nine shots within 330 frames (Fig. 1) and the real color video sequence consisting of 12 920 frames (Fig. 2) containing large motions and dynamic scene changes.

To extract the key frames, we use two criteria. If both the cumulative directed divergence value in (2) and the directed divergence 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. 3. In Fig. 3, even though the accumulated directed divergence value is larger than the threshold value, the accumulated directed divergence gradually increases because the directed divergence value between the previous key frame and the current frame is smaller than the threshold. Therefore, both conditions must be satisfied to detect a key frame. If the key

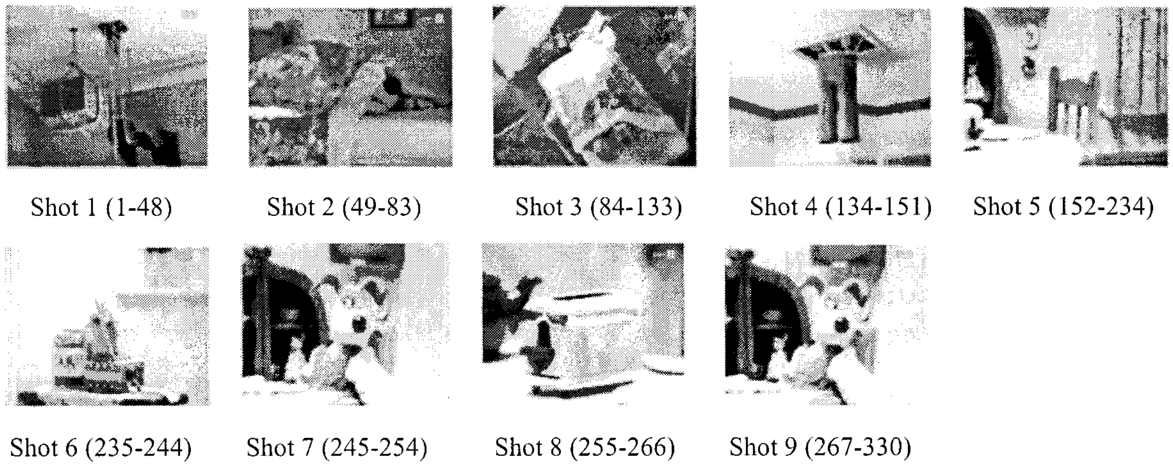

Fig. 1. Shot boundaries with start and end frames of each shot in the animation color video sequence.

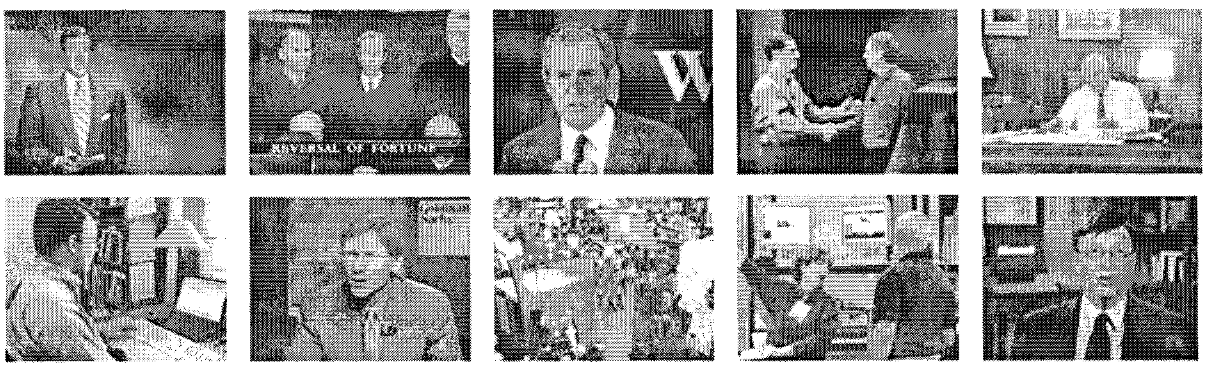

Fig. 2. Key frames within the real color video sequence (News article).

frame is extracted, the accumulated directed divergence is reset to zero.

To show the performance of video sequence matching methods, we have simulated four methods with color video sequences. Table I and Fig. 4 show matching results of the animation color video sequence using the modified Hausdorff distance. In Fig. 4, the normalized modified Hausdorff distance value between the set of query key frames and the video sequence to be compared is shown as a function of the frame number. In Table I, "Matching shot" ("Dissimilar shot") represents the modified Hausdorff distances between sets of query key frames and the color video sequence to be compared containing matching shot (nonmatching shot). In Fig. 4 and Table I, "modified histogram difference (intersection)" signifies the histogram difference (intersection) method using the modified Hausdorff distance and "Chi-square test" represents the Chi-square method using the modified Hausdorff distance. Table II and Fig. 5 show matching results using the modified Hausdorff distance for the real color video sequence (News article).

In Tables I and II, 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. Tables I and II show that the proposed

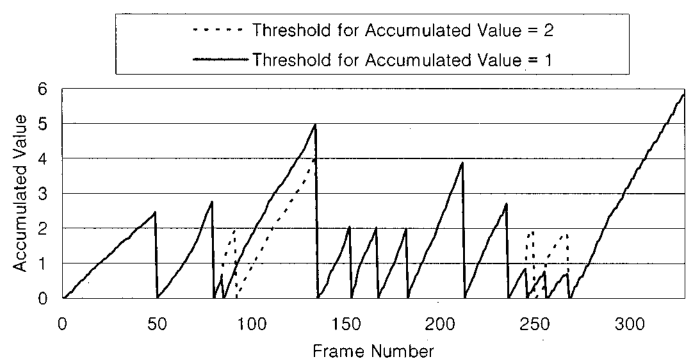

Fig. 3. Accumulated value as a function of the frame number in key frame extraction.

method using the modified Hausdorff distance and the directed divergence can remarkably improve the accuracy with the low computational complexity for color video sequence matching, compared with histogram difference, histogram intersection and Chi-square test methods using the modified Hausdorff distance.

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 [15].

TABLE I
PERFORMANCE COMPARISON OF ANIMATION COLOR VIDEO SEQUENCE MATCHING USING THE MODIFIED HAUSDORFF DISTANCE

Methods	Matching shot (A)	Dissimilar shot (B)	Ratio (B/A)
Modified Histogram Difference	0.036	0.185	5.139
Modified Histogram Intersection	0.809	0.872	1.078
Chi-Square Test	0.023	0.111	4.826
Proposed Method	0.072	0.606	8.417

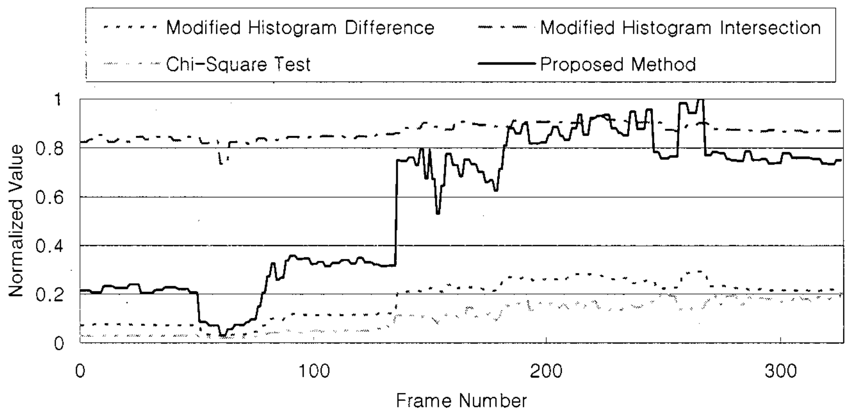

Fig. 4. Performance comparison of animation color video sequence matching as a function of the frame number.

TABLE II
PERFORMANCE COMPARISON OF REAL COLOR VIDEO SEQUENCE MATCHING USING THE MODIFIED HAUSDORFF DISTANCE

Methods	Matching shot (A)	Dissimilar shot (B)	Ratio (B/A)
Modified Histogram Difference	0.082	0.292	3.561
Modified Histogram Intersection	0.644	0.763	1.185
Chi-Square Test	0.048	0.070	1.458
Proposed Method	0.076	0.329	4.329

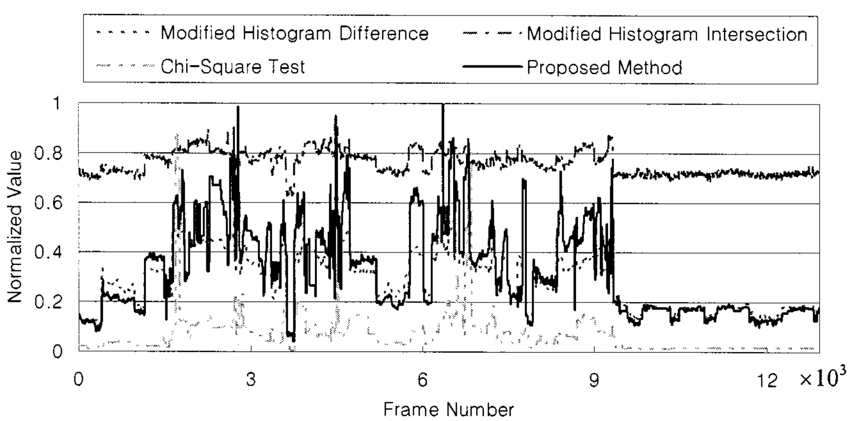

Fig. 5. Performance comparison of the real color video sequence matching as a function of the frame number.

IV. CONCLUSIONS

This paper proposes the efficient method using the modified Hausdorff distance and the directed divergence for video sequence matching. The proposed method gives better results than conventional methods such as histogram difference, histogram intersection and Chi-square test methods. Experimental results show that the proposed algorithms can efficiently match the video sequences with a low computational load and extract key frames effectively. Further research will focus on the semantic video sequence indexing and matching and verification with various video sequences containing complex scenes.

REFERENCES

[1] B.-L. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 533–544, Dec. 1995.

[2] N. V. Patel and I. K. Sethi, "Video shot detection and characterization for video databases," *Pattern Recognit.*, vol. 30, pp. 583–592, Apr. 1997.

[3] R. Brunelli, O. Mich, and C. M. Modena, "A survey on the automatic indexing of video data," *J. Vis. Commun. Image Repres.*, vol. 10, no. 2, pp. 78–112, June 1999.

[4] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 533–544, Feb. 2000.

[5] J. Meng, Y. Juan, and S.-F. Chang, "Scene change detection in a MPEG compressed video sequence," in *Proc. IS&T/SPIE Symp. Digital Video Compression: Algorithms and Technologies*, vol. 2419, San Jose, CA, Feb. 1995, pp. 14–25.

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

[7] H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 1269–1279, Dec. 1999.

[8] F. Dufaux, "Key frame selection to represent a video," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Vancouver, Canada, Sept. 2000, pp. 275–278.

[9] S. H. Kim and R.-H. Park, "A novel approach to scene change detection using a cross entropy," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Vancouver, Canada, Sept. 2000, pp. 937–940.

[10] R. W. Johnson, "Axiomatic characterization of the directed divergences and their linear combinations," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 709–716, Nov. 1979.

[11] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26–37, Jan. 1980.

[12] 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. IS&T/SPIE Conf. Storage and Retrieval for Media Databases 2000*, vol. 3972, San Jose, CA, Jan. 2000, pp. 564–572.

[13] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 850–863, Sept. 1993.

[14] S. H. Kim and R.-H. Park, "An efficient algorithm for video sequence matching using the Hausdorff distance and the directed divergence," in *Proc. SPIE Conf. Visual Communications and Image Processing 2001*, vol. 4310, San Jose, CA, Jan. 2001, pp. 754–761.

[15] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 703–715, June 2001.

Sang Hyun Kim (S'99) received the B.S. and M.S. degrees in electronic and control engineering from Hankuk University of Foreign Studies, Korea, in 1997 and 1999, respectively. He is currently working toward the Ph.D. degree in the Department of Electronic Engineering at the Sogang University, Seoul, Korea.

His current research interests are video indexing, video coding, and computer vision.

Rae-Hong Park (S'76–M'84–SM'99) was born in Seoul, Korea, in 1954. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively.

He joined the faculty of the Department of Electronic Engineering at Sogang University, Seoul, Korea, in 1984, where he is currently a Professor. In 1990, he spent a sabbatical year at the Computer Vision Laboratory of the Center for Automation Research, University of Maryland, College Park, as a Visiting Associate Professor. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* in 1995-1996. His current research interests are video communication, computer vision, and pattern recognition.

Dr. Park received a Post-Doctoral Fellowship from the Korea Science and Engineering Foundation (KOSEF) in 1990. He received the Academic Award in 1987 from KITE and the Haedong Paper Award in 2000 from the Institute of Electronics Engineers of Korea (IEEK). He received the First Sogang Academic Award in 1997 and the Professor Achievement Excellence Award in 1999, from Sogang University.