IEICE TRANS. INF. & SYST., VOL.E99–D, NO.6 JUNE 2016

LETTER

Key Frame Extraction Based on Chaos Theory and Color Information for Video Summarization

Jaeyong JU[†a)], **Taeyup SONG**[††], **Bonhwa KU**[†], ***Nonmembers***, ***and*** **Hanseok KO**[†], ***Member***

SUMMARY Key frame based video summarization has emerged as an important task for efficient video data management. This paper proposes a novel technique for key frame extraction based on chaos theory and color information. By applying chaos theory, a large content change between frames becomes more chaos-like and results in a more complex fractal trajectory in phase space. By exploiting the fractality measured in the phase space between frames, it is possible to evaluate inter-frame content changes invariant to effects of fades and illumination change. In addition to this measure, the color histogram-based measure is also used to complement the chaos-based measure which is sensitive to changes of camera /object motion. By comparing the last key frame with the current frame based on the proposed frame difference measure combining these two complementary measures, the key frames are robustly selected even under presence of video fades, changes of illumination, and camera/object motion. The experimental results demonstrate its effectiveness with significant improvement over the conventional method.

key words: *key frame extraction, video summarization, chaos theory, content change*

1. Introduction

Recent massive growth of video based big data has led to the need for effective video summarization techniques for such video data management tasks as indexing, browsing and retrieval. One of the mechanisms for summarizing videos is to extract the key frames, which represent the most important content of the video [1].

One of the popular approaches for key frame extraction is the frame difference based method, which examines significant content changes between frames based on some criteria. Various visual features have been used for evaluating frame difference such as color histograms and frame correlations [2]. DeMenthon *et al.* [3] extracted the key frames by finding discontinuities on a trajectory curve, which represent a video sequence. These methods are intuitive and simple in nature and suitable for online applications. Another approach is the clustering based method. This approach clusters the frames based on visual similarity and then selects one key frame from each cluster. For evaluating similarity, color histogram features are commonly used in these methods [4]–[6]. These methods usually show better performance than the frame difference based methods, but most of them do not consider temporal information of key frames and are computationally expensive [1]. Also, the most commonly used color histogram features are invariant to changes in camera/object motion, but sensitive to fades and illumination change. Thus, it is a challenging task to develop a key frame extraction method robust to both illumination change/fades and changes of camera/object motion.

While Chaos theory has been an active field of research for many applications including image-based motion detection [11], [12], no prior work has been known to exploit it to video summarization. The previous research by Farmer [11] analyzed the effects of motion and illumination in phase space. This research demonstrated that the phase space trajectories due to motion and varying illumination result in chaos-like behavior and non-chaotic, respectively. Motivated from [11], Farmer's work is exploited in this paper to cope with the challenging problems described above in key-frame extraction. The main difference is that while Farmer focused on the object motion change using the chaos-based measure, this paper focuses on the content change while suppressing sensitiveness to the motion change of the chaos-based measure by combining it with the color histogram-based measure.

In summary, we propose an efficient and robust technique for key frame extraction based on inter-frame difference. To evaluate the inter-frame content changes between frames, a novel frame difference measure combining two complementary measures, i.e., the chaos-based measure and the color histogram-based measure, is proposed. Using this measure as a criterion for key frame selection, the key frames are robustly selected even under presence of video fades, changes of illumination, and camera/object motion.

2. Proposed Key-Frame Extraction

2.1 Proposed Framework for Key Frame Extraction

The proposed method for key frame extraction consists of three steps. In the pre-processing step, the original video is pre-sampled by a certain sampling rate to reduce the amount of data to be processed. In the main step, the frames are sequentially compared with the last key frame based on the proposed dissimilarity measure and new key frames are chosen only if there is a significant inter-frame difference (i.e., the dissimilarity measure between frames is higher than a certain threshold). When calculating the dissimilarity measure, the frames are divided into non-overlapping sections

Manuscript received November 30, 2015.
Manuscript revised January 27, 2016.
Manuscript publicized February 23, 2016.

[†]The authors are with the School of Electrical Engineering, Korea University, Seoul, 136–713, Korea.
[††]The author is with the Department of Biomicrosystem Engineering, Korea Univeristy, Seoul, 136–713, Korea.
a) E-mail: jyju@ispl.korea.ac.kr

DOI: 10.1587/transinf.2015EDL8247

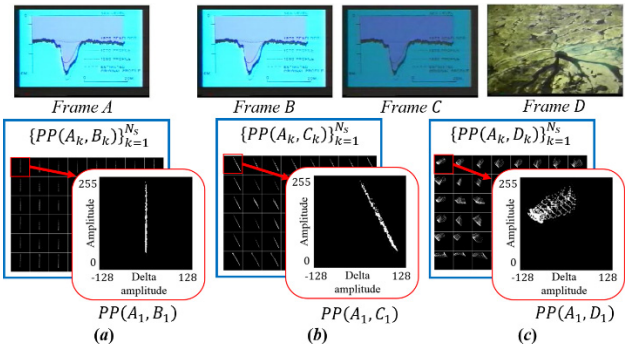

Fig. 1 Phase space plots between the corresponding sections of each frame pair: (a) similar frames (A, B), (b) faded frames (A, C), (c) different frames (A, D)

and the corresponding sections of the current frame are then compared with the last key frame. In addition, meaningless frames such as monochromatic (e.g., totally black or white) frames are excluded by examining whether it has a very low average standard deviation of all sections of the frame. In the post-processing step, after extracting the key frames, those key frames that are very similar to each other are further removed based on the color histogram-based similarity.

2.2 Proposed Frame Difference Measure

By exploiting chaos theory, we construct the frame difference measure, focusing on two elements: (i) sensitivity to detection of a significant content change between frames, and (ii) insensitivity to the effects of fades and illumination changes.

Various researchers have modelled illumination changes as multiplicative linear effects [8]. Video fades are obtained by incorporating photographic effect usually through editing. Fade-out is a gradual transition of a scene by diminishing overall brightness and fade-in is a reverse transition of fade-out. In both cases, fades also can be mathematically modelled as luminance scaling operations [9].

On the contrary, in chaos theory, it has been shown that non-linear dynamical systems which are driven between images with an underlying multiplicative process often exhibit chaotic behaviour in the phase plot. Here, the phase plot is a geometric representation of the trajectories of a dynamical system in the phase plane (i.e., a two-dimensional phase space). In the case of building the phase plot between two images, it is very similar to the joint histogram with the following change: rather than using the amplitude from the first image and the amplitude from the second image directly in the joint histogram, the phase plot uses the amplitude from the first image (y-axis) and the relative delta-amplitude between each pixel in the two images (x-axis) [11]. Peitgen [7] states that chaotic behaviour of dynamical systems can be detected in the phase plot. Figure 1 shows examples of phase plots between the corresponding sections of each frame pair where $PP(A_k, B_k)$ is the phase plot between the corresponding k-th section of the frames A and B and N_s is the number of sections. While Fig. 1 (a) and (b) exhibit non-chaotic behaviour such as lines even under fades, Fig. 1 (c) exhibits complex chaotic-like behaviour due to the significant content change. Thus, the significant content change between frames exhibits a chaos-like behaviour which results in a high fractal measure of the phase space trajectory.

The Hausdorff dimension provides a theoretical estimate of the fractal measure of the phase space trajectory. However, since it is quite difficult to calculate, we use the Box Counting method which is the most popular approximations to the Hausdorff dimension. The Box Counting dimension of the phase space S, $\dim_{BC}(S)$, is defined as [7]:

$$\dim_{BC}(S) = \lim_{\delta \to 0}(\log \mathrm{N}_\delta(S)/ - \log \delta) \tag{1}$$

where $N_\delta(A)$ is the number of boxes of size δ that cover the space trajectory over the phase space S. This fractal dimension then is used to establish a dissimilarity measure between frames. The following properties of the fractal dimension, $\dim_F$, are particularly useful for evaluating inter-frame differences:

$$1 \leq \dim_F(PP(A,B)) \leq 2 \tag{2}$$

$$\dim_F(PP(A,A)) \leq \dim_F(PP(A,B)) \tag{3}$$

Based on these properties, the process of key frame extraction will be to select new key frames if there is a significant inter-frame difference between the last key frame and the current frame, i.e., a high fractal dimension value close to 2. Finally, the chaos-based dissimilarity measure between two frames A and B in the range of 0-1 is defined as follows:

$$D_c(A,B) = \frac{1}{N_s}\sum_{k=1}^{N_s} \dim_{BC}(PP(A_k,B_k)) - 1 \tag{4}$$

where N_s is the number of sections and $PP(A_k, B_k)$ is the phase plot between the corresponding k-th section of the frames A and B.

In addition to this measure, the color histogram-based measure are also used to complement the chaos-based measure sensitive to changes of camera/ object motion. However, since the chaos-based measure is sensitive to changes of camera/object motion, those changes between similar content frames may result in a relatively high frame difference value if using this measure alone. On the other hand, if only the histogram-based measure is used, a frame difference value between different content frames with similar color distribution may be relatively low and that of similar frames under fades/illumination changes may be relatively high. Thus, we combine these two complementary measures to construct the unified and robust frame difference measure rectifying their shortcomings. To compute the color histograms, we adopt the HSV color space which provides relative closer representation of human perception. The bins of each component are set to 16 bins for H, and 8 bins for S and V components, respectively. The HSV histograms are then normalized in the range of 0–1 and the color histogram feature vector (32-dim) is constructed by concatenating these

three histograms. To calculate the histogram-based dissimilarity measure between two frames, the color histograms are computed for each section of the frames. Then, the Bhattacharyya coefficient [10] is computed between the histograms of corresponding sections of the frames. Finally, the color histogram-based dissimilarity measure in the range of 0-1 is defined as follows:

$$D_h(A,B) = 1 - \frac{1}{N_s}\sum_{k=1}^{N_s}\Omega(h_{A_k}, h_{B_k}) \tag{5}$$

where $\Omega(\cdot,\cdot)$ is the Bhattacharyya coefficient value between two histograms, and h_{A_k} and h_{B_k} are the normalized color histogram features of the corresponding k-th sections of the frames A and B, respectively.

To combine the chaos-based and the histogram-based dissimilarity measures for the frame difference measure, first, these two complementary measures are re-normalized using a sigmoid function in the range of 0-1, respectively, as follows:

$$\widetilde{D}_c(A,B) = \frac{\phi_c(D_c(A,B)) - \phi_c(0)}{\phi_c(1) - \phi_c(0)},$$
$$\text{where } \phi_c(x) = \frac{1}{1+\exp(-a_c(x-\tau_c))} \tag{6}$$
$$\widetilde{D}_h(A,B) = \frac{\phi_h(D_h(A,B)) - \phi_h(0)}{\phi_h(1) - \phi_h(0)},$$
$$\text{where } \phi_h(x) = \frac{1}{1+\exp(-a_h(x-\tau_h))} \tag{7}$$

In each sigmoid function $\phi_c(x)$ and $\phi_h(x)$, a_c and a_h are the positive slope parameters and τ_c and τ_h are the predefined x-intercepts to consider a significant inter-frame change. The reason for re-normalizing each dissimilarity measure using a sigmoid function is to stretch the important range of dissimilarity values between significant and non-significant change while suppressing the less important range of dissimilarity values. This way, selecting key frame by thresholding on the frame difference measure can be more effective. If $D_c(A,B)$ is greater than τ_c, it is considered that there is a significant inter-frame change between the frames A and B and the value of $\widetilde{D}_c(A,B)$ is close to 1, and 0 vice versa. In common with the above case, if $D_h(A,B)$ is greater than τ_h, it is considered that there is a significant inter-frame difference and the value of $\widetilde{D}_h(A,B)$ is close to 1, and 0 vice versa.

Finally, the proposed total frame difference measure between the frames A and B is defined by:

$$\widetilde{D}_T(A,B) = w_c\widetilde{D}_c(A,B) + w_h\widetilde{D}_h(A,B) \tag{8}$$

where w_c and w_h are the weights for each dissimilarity term $\widetilde{D}_c(A,B)$ and $\widetilde{D}_h(A,B)$, respectively, and $w_c+w_h = 1$. A high value of $\widetilde{D}_T(A,B)$ close to 1 indicates a significant inter-frame change. Therefore, if the total frame difference measure $\widetilde{D}_T(F_k,F_c)$ between the last key frame F_k and the current frame F_c is higher than a certain threshold τ, the current frame F_c is chosen as a new key frame.

3. Experimental Results

Performance of the proposed method is evaluated on the public Open Video Project (www.open-video.org) dataset which consisted of 50 videos containing fades and changes of illumination and camera/object motion [6]. The user summaries for each video in this dataset are also available. Based on this dataset, we compared our method with OV [3], DT [4], STIMO [5], and VSUMM [6] which is the clustering-based method using color features and k-means algorithm. The difference between VSUMM_1 and VSUMM_2 in Table 1 is that one key frame is selected per either cluster or keycluster which is larger than the average cluster size.

Table 1 Comparisons of video summarization performance.

	OV [3]	DT [4]	STIMO [5]	VSUMM_1 [6]	VSUMM_2 [6]	Proposed
R	0.70	0.53	0.72	0.85	0.70	0.82
P	0.55	0.65	0.55	0.69	0.72	0.74
F	0.62	0.58	0.63	0.76	0.71	0.78

In the system parameter settings, the sampling rate was 1 frame/sec, and N_s was set to $8\times6 = 48$ with a section size of 40×40. Moreover, the parameters a_c and a_h were set to 20, the weights w_C and w_H were set to 0.5, and the thresholds τ_c, τ_h, and τ were set to 0.3, 0.15, and 0.5, respectively.

For performance evaluation, we adopted the standard metrics "Recall" and "Precision". "Recall" is the ratio of the number of matching key frames to the total number of key frames from user summary. "Precision" is the ratio of the number of matching key frames to the total number of key frames from automatic summary. The higher both Recall and Precision are, the better the performance of the automatic summary is. However, there is a trade-off between Precision and Recall. The increase of Precision by selecting very few key frames leads to the decrease of Recall. Otherwise, the increase of Recall by selecting too many key frames leads to the decrease of Precision. Thus, to overcome this trade-off, we additionally adopted the F-measure combining both Precision and Recall into a single metric by a harmonic mean for evaluating the performance of the automatic summaries as follows.

$$\text{F} = 2\times\frac{\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}} \tag{9}$$

A high value of F-measure thus indicates a high value for both Precision and Recall.

The comparison results of average performance between the proposed method and the most prominent state-of-art methods are presented in Table 1.

The results indicate that while VSUMM_1 shows the highest Recall, its Precision is relatively low compared to VSUMM_2 and our method because of selecting too many key frames. On the other hand, the proposed method achieves the highest F-measure and Precision compared to all the other methods while achieving a competitive performance in Recall by providing sufficient video summaries. Considering these results, it is possible to conclude that the proposed method is more effective than other methods. Figure 2 shows an example of video summary result of the proposed method.

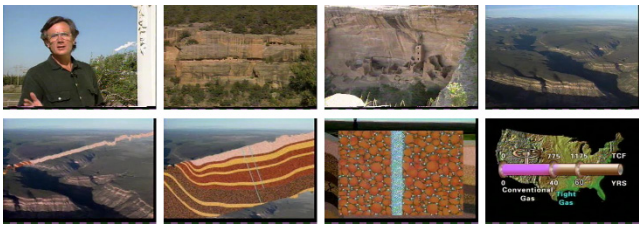

Fig. 2 Video summary of the proposed method for the video "The Future of Energy Gases, segment 05"

4. Conclusion

In this paper, we proposed a novel technique for key frame extraction based on chaos theory and color information for automatic video summarization. The proposed frame difference measure combining the two complementary measures, i.e. the chaos-based measure and the color histogram-based measure, enabled both efficient and robust key frame extraction even under the presence of fades and changes in illumination and camera/object motion. Experimental results generally demonstrated the effectiveness and improvement of the proposed method over previous methods.

Acknowledgments

Research was supported by Korea University.

References

[1] B.T. Truong and S. Venkatesh, "Video abstraction: a systematic review and classification," ACM Transactions on Multimedia Computing, Communications, and Applications, vol.3, no.1, Feb. 2007.

[2] R.M. Jiang, A.H. Sadka, and D. Crookes, "Advances in video summarization and skimming," Recent Advances in Multimedia Signal Processing and Communications, vol.231, pp.27–50, Springer Heidelberg, Berlin, 2009.

[3] D. DeMenthon, V. Kobla, and D. Doermann, "Video summarization by curve simplification," Proc. ACM International Conference on Multimedia, Bristol, United Kingdom, pp.211–218, Sept. 1998

[4] P. Mundur, Y. Rao, and Y. Yesha, "Keyframe-based video summarization using Delaunay clustering," International Journal on Digital Libraries, vol.6, no.2, pp.219–232, April 2006.

[5] M. Furini, F. Geraci, M. Montangero, and M. Pellegrini, "STIMO: STIll and moving video storyboard for the web scenario," Multimedia Tools and Applications, vol.46, no.1, pp.47–69, Jan. 2010.

[6] S.E.F. de Avila, A.P.B. Lopes, A. da Luz Jr., and A. de Albuquerque Araújo, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method," Pattern Recognition Letters, vol.32, no.1, pp.56–68, Jan. 2011.

[7] H.O. Peitgen, H. Jürgens, and D. Saupe, "Chaos and fractals: new frontiers of science," Springer Science & Business Media, New York, 2006.

[8] R. Basri and D.W. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Trans. PAMI, vol.25, no.2, pp.218–233, Feb. 2003.

[9] A. Hampapur, T. Weymouth, and R. Jain, "Digital video segmentation," Proc. ACM International Conference on Multimedia, San Francisco, USA, pp.357–364, Oct. 1994.

[10] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Communication Technology, vol.15, no.1, pp.52–60, Feb. 1967.

[11] M.E. Farmer, "A chaos theoretic analysis of motion and illumination in video sequences," Journal of Multimedia, vol.2, no.2, pp.53–64, April 2007.

[12] M.E. Farmer, "A comparison of a chaos-theoretic method for preattentive vision with traditional grayscale-based methods," Proc. IEEE Conference on Advanced Video and Signal-Based Surveillance, Klagenfurt, Austria, pp.337–342, Aug. 2011.