Video Analysis Algorithm based on Saliency Region Detection from Selected Key-frames

Min Soo Ko[1] and Hyok Song[2]
[1, 2] Korea Electronics Technology Institute
25, Saenari-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-816, Korea
E-mail: [1]kmsqwet@keti.re.kr, [2]hsong@keti.re.kr

Abstract: In the past few years, the amount of the internet video has grown rapidly, and it has become a major market. Reliable extraction of information from the video is one important step for efficient video management. In this paper, effective saliency region detection and key-frame selection algorithms for a video analysis is proposed. The first step for video analysis is the partitioning of a video sequence into scene. And we select key-frame from each scenes. In order to extract reliable information for the video, we determine saliency region from the key-frames. The experimental results show that the proposed algorithm performs much better than conventional algorithms.

Keywords-- **Video analysis, Saliency map, Scene change detection, Keyframe selection**

1. Introduction

Through the rapid growth of the Internet, the volume of digital video data has been significantly increasing. But the whole of video content may not be of interest to the user. Also we still have limited tools and applications to describe, organize, and manage video data. So research of effective algorithms are greatly needed which will lower the costs of video managing, as well as improve the efficiency and accessibility of digital videos. To handle huge amount of video efficiently, many algorithms about video segmentation and abstraction have emerged to analyze the stored digital videos [1].

In this paper, we propose saliency region detection and key-frame selection algorithms for a video analysis. Fig. 1 shows concept of proposed algorithm.

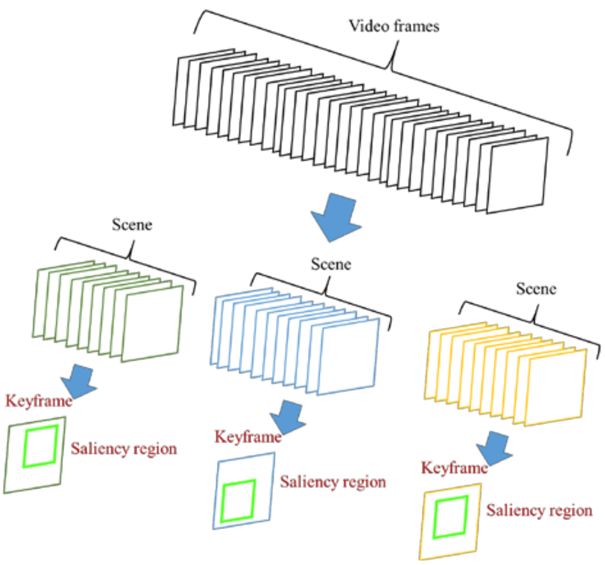


Figure 1. Concept of proposed algorithm

2. Proposed Algorithm

In this section, we describe the proposed algorithm. Fig. 2 shows a block diagram of the proposed algorithm. Two major components in our approach are key-frame selection and saliency region detection.

At first, we perform scene change detection algorithm in order to seperate the input video into scenes. And we select key-frames from each scenes. Next, we generate saliency map from the key-frames. Then, we determine saliency region by labelling the generated saliency map.

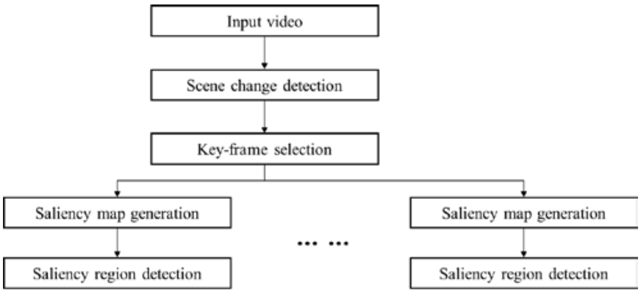


Figure 2. Block diagram of the proposed algorithm

2. 1 Scene change detection & Key-frame selection

Key-frame refers to the image frame in the video which is representative and able to reflect the summary of a video content. By using the key-frame it is able to express the main content of video data clearly. And It reduce the amount of memory needed for video data processing and complexity greatly. Thus key-frame selection is an efficient algorithm for video analysis [2].

In the scene change detection based key-frame selection, a video is segmented into a number of scenes and one or more key-frames are selected from each scenes.

In the proposed algorithm, we use Eigen values for calculating the dissimilarity between the fixed size consecutive frames.

First, We convert the input frame image using SVD-based image compression algorithm. The fundamental concept of the SVD-based image compression algorithm is to use a smaller number of rank to approximate the original matrix. This operation can be expressed in the Eq. (1).

$$A = U\Sigma V^T = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} \quad (1)$$

$$= u_1\sigma_1 q_1^T + u_2\sigma_2 q_2^T + \cdots + u_n \sigma_n q_n^T$$

where, A is a matrix containing the input image data. And U denotes left sigular vectors and V denotes right singular vectros. Σ denotes diagonal matrix to include sigular values. The higher singular values contribute most to the sum, and hence contain the most information about the image. Also the lower singular values containing negligible or less important information can be discarded without significant image distortion. So we use SVD-based image compression algorithm to remain the significant image information. Fig. 3 shows the re-constructed image using the SVD-based compression algorithm[3].

(a) (b)

Figure 3. Re-constructed image using the SVD-based compression algorithm
(a) original image (b) re-constructed image

In order to calculate dissimilarity, we generate the covariance matrix using the re-constructed images as in Eq. (2). As shown in Eq. (3), the behavior of the Eigen values of covariance matrix over certain frames is analyzed to detect the scene boundary[4].

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ C_{21} & C_{22} & \ddots & C_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{bmatrix} \quad (2)$$

$$C = P \Lambda P^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix} \quad (3)$$

where, C denotes the covariance matrix covariance matrix over certain frames. And P denotes the Eigen vectors, Λ denotes the Eigen values. The minimum the Eigen value chosen is used to detect the dissimilarities between the frames. If the minimum the Eigen value exceeds the threshold value, the $(k\text{-}1)^{th}$ frame will be scene boundary. And We select the middle frame of the scene such as key-frame of current scene.

2. 2 Saliency region detection

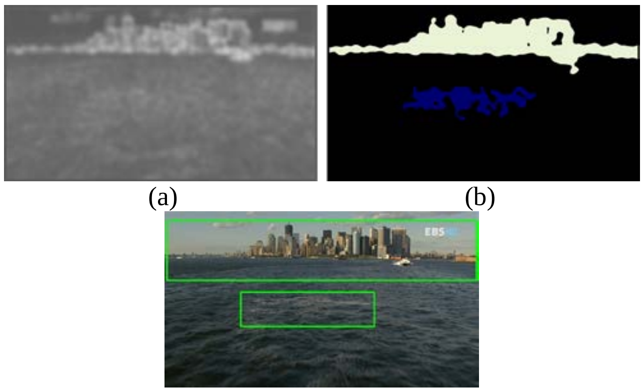
During the several years, a number of saliency models have been proposed for extracting regions of interest in image/video. Saliency detection models use to the image processing applications for various such as object segmentation, content-aware image retargeting and image retrieval.

In the proposed algorithm, saliency values for each pixels are calculated by quantifying the self-information of each local image patch. We use determined 79 basis functions from large set of natural images. We calculate independent coefficients for every local neighborhood in the all of the image region. Any given coefficient may be readily converted to a probability by looking up its likelihood. Thus we determine probability value for any single coefficient based on a probability density estimate in the form of a histogram. All of the individual likelihoods corresponding to a particular local region convert to the joint likelihood. In order to determine saliency value for each pixels, the joint likelihood is translated into Shannon's measure of Self-Information. Fig. 4 shows the result of saliency region[5].

(a) (b)

Figure 4. Result of salency map
(a) original image (b) saliency map

To detect saliency region in the saliency map, we use labelling algorithm. And we only extract region over a certain size in the labelled regions. Fig. 5 shows result of saliency region.



(a) (b)

(c)

Figure 5. Result of salency region
(a) saliency map (b) labelling map
(c) result of saliency region

3. Experimental Result

We compared the proposed algorithm's performance with a conventional algorithms. Fig. 6 shows comparison of saliency maps generated for some test images [6,7]. The experimental results show that the proposed algorithm performs much better than conventional algorithms.

Fig. 7 shows result of proposed algorithm. As shown in Fig. 7, it can be seen that through the proposed algorithm to the saliency region of the video is extracted well.

4. Conclusion

In this paper, we proposed an improved algorithm to analyze a video based on saliency region detection from key-frames. In the proposed algorithm, we tried to select key-frame using Eigen values. We also tried to detect saliency region using self-information measure from each key-frames.

Sometimes proposed algorithm fails to detect the fades and dissolves transitions. In our experiments, the threshold is set empirically but in future it can be made adaptive. In the future, we will develop a more robust algorithm which can detect any transitions with high accurate results.

Acknowledgements

This research was funded by the Ministry of science, ICT and Future Planning Korea in the GCS program (R0190-15-1115) 2015.

References

[1] C. Ngo, Y. Ma and H. Zhang, “Video summarization and scene detection by graph modeling,” in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp. 296-305, Feb. 2005.

[2] M. Cooper and J. Foote, “Discriminative techniques for keyframe selection,” 2005 IEEE International Conference on Multimedia and Expo, pp. 4-, 2005.

[3] S. Kahu, R. Rahate, “Image compression using singular value decomposition,” International Juournal of Advancement in Research & Technology, Vol. 2, Issue 8, pp. 244-248, Aug. 2013.

[4] V. Benni, R. Dinesh, P. Punitha, and V. Rao, “Keyframe Extraction and Shot Boundary Detection Using Eigen Values,” International Journal of Information and Electronics Engineering, Vol. 5, No. 1, pp.40-45, Jan. 2015.

[5] N.D.B. Bruce, J.K. Tsotsos, “Saliency Based on Information Maximization,” Advances in Neural Information Processing Systems, 18, pp. 155-162, June 2006.

[6] L. Itti, C. Koch, E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259, Nov. 1998.

[7] B. Schölkopf, J. Platt, T. Hofmann, “Graph-Based Visual Saliency,” Advances in Neural Information Processing Systems 19:Proceedings of the 2006 Conference, pp.545-552, 2007.

Figure 6. Comparison of saliency maps generated for some test images

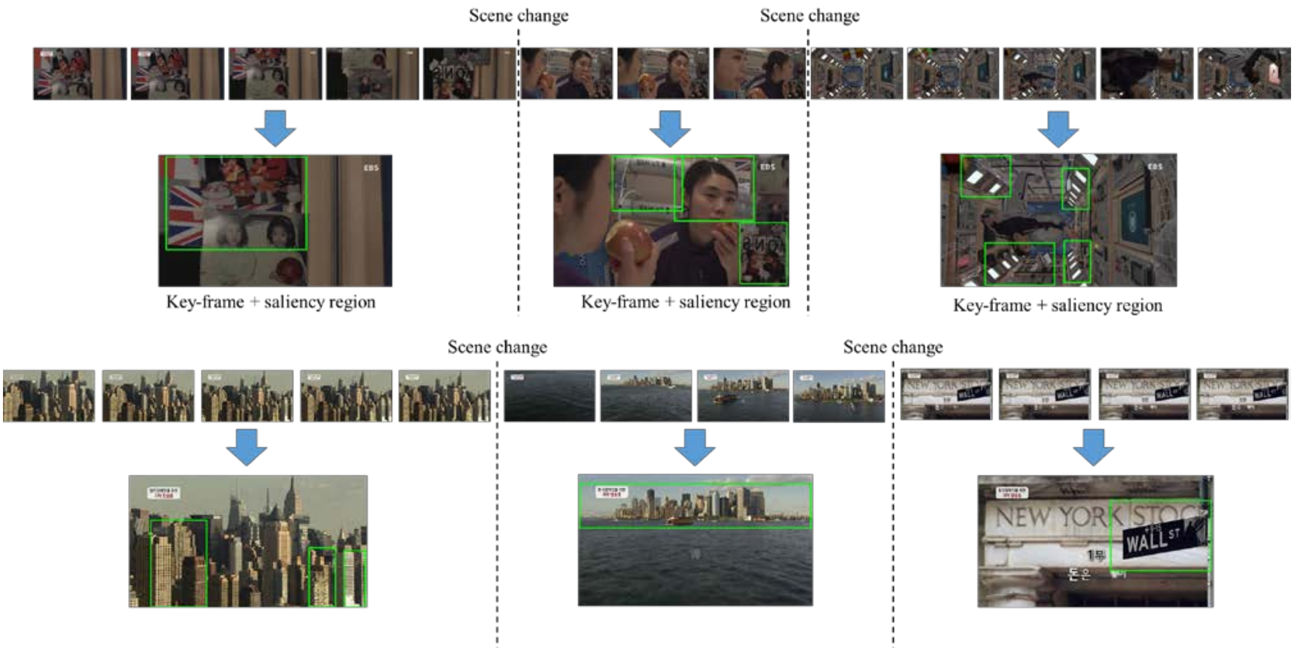


Figure 7. Result of proposed algorithm