

A Summarization Method for Medical Documentary Videos

Giseok Choe*, Hyerin Kim, Jongho Nang
Department of Computer Science and Engineering
Sogang University, Seoul, Korea
gschoe@sogang.ac.kr, hyerin16@gmail.com, jhnang@sogang.ac.kr

Abstract— This paper proposes a medical documentary video summarization method based on its content characteristics that are specific to this genre. The shots in “Mysteries of the Human Body”[1], that is a medical documentary video produced by KBS, were put into the categories of studio, interview, illustration, and others in the shot-classification stage. To perform this task, some useful algorithms[2] of face detection, text area detection, and graphic element detection were employed. In the subsequent summarization stage, the shots were hierarchically aggregated to video sentences and video paragraphs[3], then evaluated with an algorithm using six rules that score them as summary candidates. During the shot discarding stage, the shots with low scores were iteratively discarded from summary shot candidates. To evaluate the performance of our summarization method, the print edition, “Korean 100-year-old secret of health (Korean edition)”[4], of the source video was used as the ground truth. Our experiment show that the precision and recall of our proposed algorithm are 0.67 and 0.70, respectively.

Keywords—video summarization; documentary video;

I. INTRODUCTION

Documentary videos are rich in information compared to other video genre, and video summary can contribute to easy video browsing for viewers [5]. However, the summary video generated by previously proposed summary algorithms [5, 6, 7] may not be satisfiable, because they are general ones so that the structural characteristics of medical documentary video could be used. This paper proposes a video summarization algorithm for a particular video genre, namely, medical documentary video. We chose “Mysteries of the Human Body” as our data set. The summary result of the algorithm was compared to the print edition of the series, “Korean 100-year-old secret of health (Korean edition) [4]”.

II. PREVIOUS WORKS

A. Structure of Medical Documentary Videos

Choi et al. [3] surveyed medical documentary videos and reported that a typical medical documentary has many shots with speech and its structure can be described similar to that of a written essay:

- (1) *Shot*: an aggregation of video contiguous frames,
- (2) *Video Sentence (VS)*: a series of shots covered in a single speech,

- (3) *Video Paragraph (VP)*: an aggregation of VS collected as a content chunk

We found out that this hierarchical structure was identical to the structure of the book titled “Korean 100-year-old secret of health (Korean edition)[4]”.

B. Shot Types in Medical Documentary Videos

Lee et al. [2] described the four main shot types that appear in “Mysteries of the Human Body” as follows:

- (1) *studio*: shots taken by show hosts in the studio,
- (2) *interview*: shots of interviewees,
- (3) *illustration*: infographics and charts for explanation,
- (4) *others*: other shots, mostly taken outside the studio

Shots can be put into these four different categories by using the detection methods for face, text-region, and graphics elements as explained in the following section.

III. DESIGN OF SUMMARIZATION ALGORITHM

The following describes how we designed the video summary algorithm.

A. Shot Type Classification

Table I shows how shot types are classified in our summary algorithm based on the results of face detection, text-region detection, and graphics element detection.

Table I. Shot type classification using the results of face detection, text-region detection and graphical element detection

Shot Type	Face Detection	Text-region Detection	Graphic Element Detection
Studio	True	Don't care	Graphic Picture
Interview	True	True	Real Picture
Illustration	False	Don't care	Graphic Picture
Others	Don't care	Don't care	Real Picture

B. Video Summarization Algorithm

Rules in selecting summary shots are described below. We adopted six scoring rules for shot selection:

- (1) *Shots that make well-distributed collection are assigned higher scores.*
- (2) *Diverse shots make better summary result.*

This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for SW in Excellence (R7718-16-1004) supervised by the IITP (Institute for Information & communications Technology Promotion)

*Corresponding author: Giseok Choe



Fig. 1. Example of a hierarchical summary.

- (3) Shots that appear in the later section tend to summarize the content better.
- (4) Shots with large-type text tend to contain important information.
- (5) VPs with a studio shot contain important information.
- (6) VSs with an interview shot contain important information.

Rules (1), (2), and (3) are widely used in conventional video summary algorithms while Rules (4), (5), and (6) are devised for medical documentary videos. An evaluation function was written that reflects the six rules described above. Then another function was written that discards VSs with lower rankings in score. By applying this function in iteration, a hierarchical summary result was obtained as illustrated in Fig. 1.

IV. EXPERIMENT AND ANALYSIS

Objective evaluation of video summarization algorithm has been difficult to achieve because the ground truth is not easily defined[6]. Fortunately, the producer of the “Mysteries of the Human Body” has published its print edition[4], which we could use as our ground truth for summary. Since this book was written by the series producer and their medical advisors, independent of our work, we argue that the print edition could serve as an objective summary of the documentary video series.

A. Dataset

Ten chapters in ‘Korean 100-year-old Secret of Health’ [4] were selected, each of which has a match in between Episodes 294 and 537. Based on the book content, short and long summary videos for each episode were manually produced for ground truth. Original videos were edited to around 26% in length for longer summary and then to around 7% for shorter summary.

B. Result and Analysis

Fig. 2 compares the length of the automatically generated summary videos versus original videos. Table II lists the accuracy of shot-classification result.

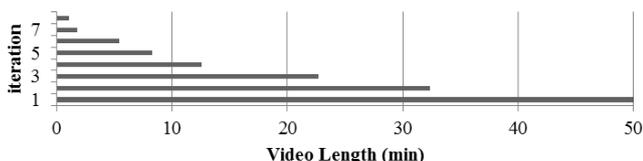


Fig. 2. The length of the automatically generated summary video

Table II. Result of shot type classification

	Studio Picture	Interview Picture	Graphic Picture	ETC
Accuracy	0.87	0.98	0.85	0.95

The produced longer summary was compared with the result at 4th iteration, and shorter summary with 7th iteration. Longer summary videos were compared to results after 4th iteration and shorter summary videos after 7th iteration. In order to test the validity of the six rules we proposed, another evaluation function was written that uses only the first three rules. Fig. 3 shows that the results in terms of recall and precision were superior when all of the six rules were used than when only the first three rules were used. Inclusion of genre-specific rules has contributed to making this difference in precision and recall.

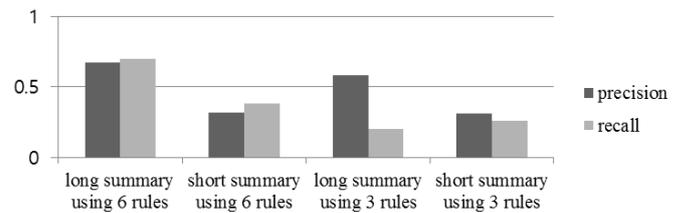


Fig. 3. The result of precision and recall using different rules.

V. CONCLUSION

This paper proposes a video summary algorithm that takes into consideration the characteristics of input video genre. Our future work will explore the possibility of including more genres for video summary.

REFERENCES

- [1] J. Lee, *Mysteries of the Human Body*, KBS in Korea, 2002.
- [2] J. Lee, I. Heo, and J. Nang, “Graphic Shot Detection Method in Documentary Video,” in *Proceedings of the 2015 Korea Computer Congress*, 2015, pp. 1687-1689.
- [3] J. Choi, H. Hong, and J. Nang, “Design and Implementation of Documentary Video Annotation System for Efficient Search Using Taxonomies : Focused on KBS <Mysteries of the Human Body>,” in *Proceedings of the 2015 Korea Computer Congress*, 2015, pp. 1699-1701.
- [4] KBS, *Korean 100-year-old secret of health (Korean edition)*: Bi Tabukseu, 2011.
- [5] B. Chen, J. Wang, and J. Wang, “A Novel Video Summarization based on Mining the Story-Structure and Semantic Relations among Concept Entities,” *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 295-312, 2009.
- [6] X. Zhu, J. Fan, A. K. Elmagarmid, and X. Wu, “Hierarchical Video Content Description and Summarization using Unified Semantic and Visual Similarity,” *Multimedia Systems*, vol. 9, no. 1, pp. 31-53, 2003.
- [7] J. Jeong, J. Nang, and H. Cha, “A Domain-Independent and Personalized Video Abstraction Algorithm,” *Journal of Information Science and Engineering*, vol. 20, no. 6, pp. 1183-1196, 2004.

