Sensors and Materials, Vol. 32, No. 2 (2020) 599–610
MYU Tokyo

S & M 2125

Automatic Route Video Summarization Based on Image Analysis for Intuitive Touristic Experience

Yuki Kanaya,[1,2] Shogo Kawanaka,[1,2,3] Hirohiko Suwa,[1,2*]
Yutaka Arakawa,[1,4] and Keiichi Yasumoto[1,2]

[1]Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, Japan
[2]RIKEN, Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan
[3]Research Fellow of JSPS, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo, Japan
[4]JST Presto, 7 Gobancho, Chiyodaku, Tokyo, Japan

(Received September 11, 2019; accepted January 20, 2020)

Keywords: automatic route video summarization, image analysis, intuitive touristic experience, decision support, tourist information application

Currently, many tourists search for and watch tourism videos on the Internet when planning a sightseeing tour. In order to quickly plan a sightseeing route, a shorter playback time of tourism videos is desirable. For this purpose, time-lapse playback would be effective. However, the faster the playback is, the lower the degree of comprehension of the viewers will be. In this paper, we propose a novel time-lapse-based video summarization method without the substantial loss of information important for viewers to plan a tour route. In the proposed method, we focus on scene changes in the video. We extract scenes with a certain level of change compared with previous scenes as important (slowly played back) in the summarized video, while other scenes are fast-forwarded. We investigated the appropriate playback speed of sightseeing videos. As a result of a questionnaire, we found that a playback speed between ×4 and ×8 was the most effective for viewers to understand the sightseeing information for tour route planning. In addition, to evaluate the effectiveness of our proposed method, we conducted experiments with 20 participants using a sightseeing video taken in Kyoto. Comparing the video summarized with our method and that summarized manually (by voting for necessary/unnecessary scenes), our method identified the important scenes with an F-measure of 62.22%.

1. Introduction

Sensing technologies have become important in many fields as seen in the trend of cyber-physical systems (CPS), machine-to-machine (M2M) systems, and the Internet of Things (IoT). Those technologies are also used for services in tourism. Various information such as texts, photos, maps, and videos collected in a whole city by various IoT devices are used for guiding, recommendation, and planning.[1–11] Tourists can watch many videos about sightseeing on the Internet through social networking services or YouTube. Videos are useful to plan a sightseeing tour because they include richer touristic information than the texts, maps, and photos in a tourism guide.

*Corresponding author: e-mail: h-suwa@po.wind.ne.jp
https://doi.org/10.18494/SAM.2020.2616

ISSN 0914-4935
https://myukk.org/

According to a study by Google,[12] 65% of leisure travelers are inspired by online sources, most notably through social/video sites and searches, while 42% of travelers are inspired to travel by YouTube content. It means that at least 42% of tourists watch videos to choose a sightseeing spot. However, it is hard to find suitable video content for each sightseeing spot on the Internet. It is even harder to select videos that meet the demand and preference of each tourist because tourists' requirements vary greatly. In our previous work, we proposed a video summarization system to support users planning a sightseeing tour.[13] Figure 1 shows the procedure of the tourism video summarization system.

This system consists of 5 steps: 1. User's Data Collection, 2. Tour Route Creation, 3. Consumer Generated Media (CGM) Collection, 4. Summarized Video Creation, and 5. Tour Route Decision. This system uses CGM, which includes photos and videos taken by tourists. These contents may not be accurate but will reflect the real situation of tourist spots. This system creates tour routes taking into account the user's preference and summarized videos along the routes. A short summarized video is made by compressing each of the segments in the original video corresponding to spots and movements between spots according to the compression rate of each segment determined by the user's preference in this system. Users can experience a virtual tour by watching the summarized videos, and they can plan and adjust their whole tour route easily.

On the other hand, it is desirable that the summarized video is reasonably short because a long video may bore viewers. There are many studies of video summarization allowing users to watch a video quickly and efficiently.[14,15] When summarizing a video, extracting scenes by characteristic frames or sounds and calculating the importance of these scenes are general processes. In particular, news[14] and sports programs[15] have scene changes with multiple cameras or characteristic sounds, and important scenes in these videos can be clearly identified. Sightseeing videos were taken by tourists; however, they do not include featured scene switching or sounds unlike news programs or sports videos. Also, we cannot use these existing methods because scenes to be extracted in sightseeing videos are not defined.

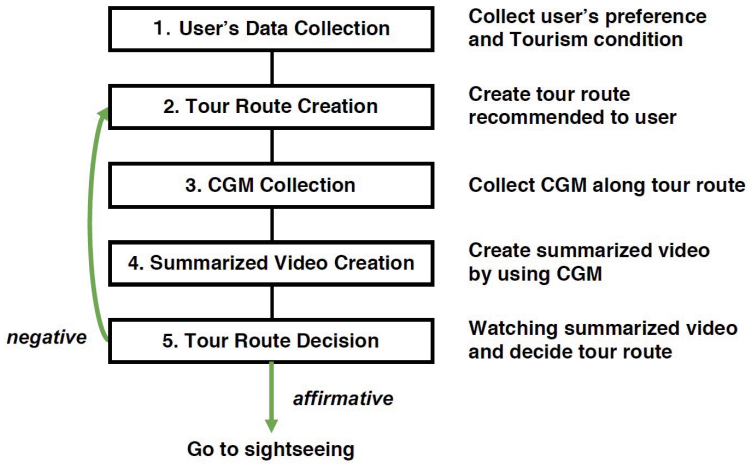

Fig. 1. (Color online) Procedure of video summarization system.

In this paper, targeting tour videos that show movements between tour spots and the situation of each spot in a tour, we define important scenes of sightseeing videos and propose a method of video summarization that extracts and plays back the important scenes and fast-forwards other scenes. Our method first calculates the color histograms of the frames in the video. By comparing these histograms between consecutive frames, we identify changing points between scenes. Finally, the video is summarized by fast-forwarding scenes except around these changing points. The reason why we do not cut all the frames with lower importance (lower changing) is to give a virtual touristic experience as if the tourist were walking along the actual sightseeing route.

In an implementation, the summarized video should be played back at an appropriate speed so that viewers can understand the touristic information in the video in order to avoid a decrease in users' comprehension degree as the playback rate increases.[16] Therefore, we investigated the relationship between comprehension degree and playback speed in tour videos, aiming to obtain the appropriate playback speed for summarized videos.

As a result, we found the playback speed between ×4 and ×8 is the best for tour videos. Also, we compared summarized videos made with our method and those with manual summarization (based on the manual labeling of important/unimportant scenes) using 20 participants to evaluate the effectiveness of our method. As a result, our method identified important scenes with an F-measure of 62.22%. Moreover, over 70% of participants answered that the summarized video made by our method was effective for planning a tour.

2. Related Work

There are many studies related to tourism.[1–11] In order to support tour planning, Kurata *et al.*[1] and Hidaka *et al.*[2] have proposed planning support systems. However, those systems only show a tour route on a map, and tourists cannot understand the tour route intuitively. Therefore, we focus on videos. Tourists many have watched many videos to obtain information about sightseeing spots on the Internet before going on sightseeing tours. However, it is difficult to find an optimized video from the massive numbers of videos on the Internet. To supply an optimized video matching each tourist, curation is needed.[17] Curation is to collect and organize various information, share them with new values, and provide users with high-value information.

In our previous study, we proposed a video summarization system[13] that aims to create curation videos using the sightseeing videos taken by ordinary people (CGM) when tourists (users) plan their sightseeing tours. This system can make a short summarized video by compressing each of the segments in the original video corresponding to spots and movements between spots according to the compression rate of each segment determined by the user's preference (lower compression rate for more important segments). To make a short and comprehensible summarized video, in addition to our previous method,[13] we need a new method for omitting unnecessary scenes in the original tour videos. For this purpose, it is necessary to determine the degree of importance of each scene to detect unnecessary ones.

Existing methods applied to news programs[14] and sports programs[15] detect important scenes easily by utilizing the fact that these videos have apparent scene changes with switching among multiple cameras or characteristic sounds (e.g., before switching to a new report). Many existing methods for summarizing these videos exist.[14,15,18,19] However, we cannot use these methods because the tour videos taken as CGM are typically one continuous shot and do not have apparent scene changes (camera switching) or characteristic sounds (between scene changes).

Some studies tried to extract important scenes from one-shot videos taken in sightseeing areas.[4] Zhang *et al*.[4] summarized a video using location information. They assumed that the scenes while stopping by famous sightseeing spots (from location information) are important and extracted these scenes from the video. Morishita *et al*.[20] proposed a method of extracting scenes where cherry blossoms are present by utilizing color histogram and fractal dimension analyses in video frames. Okamoto and Yanai[21] supposed that the important scenes of walking route guidance are street corners and summarized videos using the optical flow of consecutive video frames and detected street corners. As stated above, the importance of scenes in sightseeing videos varies greatly depending on the purpose. Unlike the above studies, our proposed method calculates the importance of scenes to make a summarized video for a virtual touristic experience.

When making summarized videos for tours, it is important to play back all the tour scenes containing both important and unimportant scenes. Otherwise, location information will be lost on viewers. Therefore, we apply a time-lapsing (fast-forwarding) method to fast-forward unimportant scenes. The resulting summarized video plays back important scenes slowly and others quickly. In this approach, we need to know the optimal fast-forward/time-lapsing play-back speed for the easy understanding of important scenes since it is known that the comprehension degree decreases as the playback speed increases.[16] Our proposed method makes a short video also taking into account the playback speed to obtain various information easily.

3. Video Summarization Method

We target tour videos (CGM) posted to SNS or other sharing services such as YouTube by ordinary users. We assume that each tour video is taken as one shot (cut) and consists of a sequence of segments called scenes, where each scene reflects similar situations (e.g., walking along a street and looking up at a building). We define points between consecutive scenes as changing points. Watching a long scene will bore the viewer because there are no big changes in the scene. Thus, we employ an approach to play back the beginning frames of a scene slowly and fast-forward (time-lapse) the remaining frames of the scene.

3.1 Overview of video summarization algorithm

The overview of extracting changing points is shown in Fig. 2 and described as follows: First, as shown in Fig. 2(a), all frames in the video are quantified using a color histogram (a

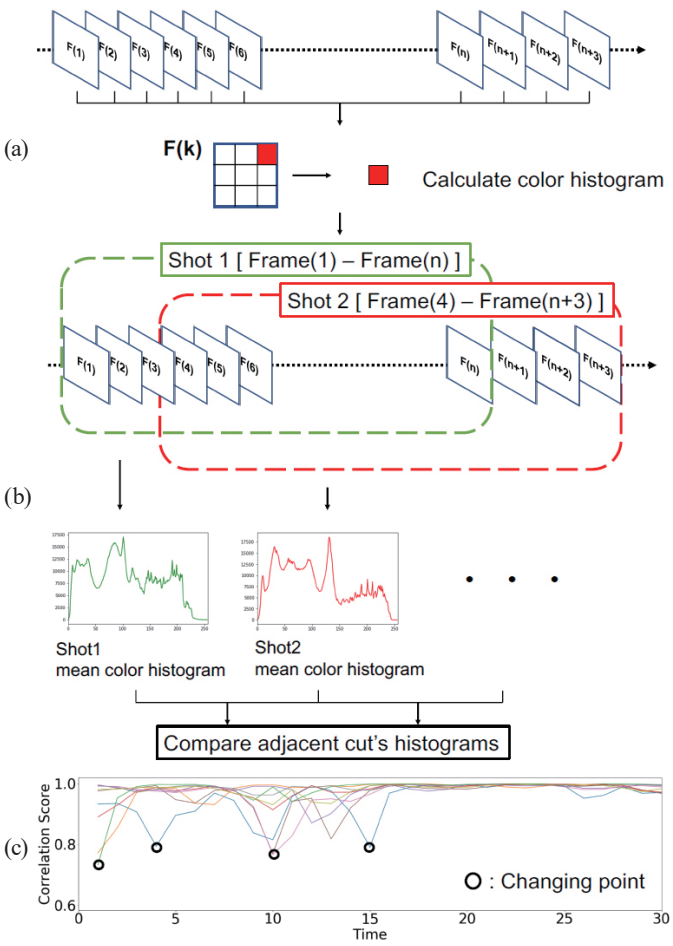

Fig. 2. (Color online) Overview of extracting changing points.

histogram is obtained for each of the 3 × 3 areas of video frames). Then, as shown in Fig. 2(b), the average of the color histogram is obtained over n frames in a sliding-window manner where a window consists of n frames (called a shot) and the shifting width is δ. Here, n and δ are predetermined constants and $\delta < n$. Finally, while comparing the average of the color histogram with adjacent shots, changing points are extracted by finding the lowest similarity (correlation) points [Fig. 2(c)].

Algorithm 1 shows our proposed algorithm. All frames in the video are divided into 9 areas (line 1). 256-level gray-scale histograms for all areas are calculated for all frames (line 2). Here, the total number of frames in the video is denoted by N. Let δ denote the number of frames for a shifting window (shot). The average of a color histogram in n frames, denoted by H_s^i, is calculated for all shots by shifting δ frames (lines 3–9). C_s (s = 1, 2, ..., $\lfloor (N - n)/\delta \rfloor$) is the correlation coefficient, obtained by comparing the adjacent shots H_s^i and H_{s+1}^i (line 10). Shots with more than one area whose coefficients fall in the bottom p% of values are selected as changing points and output as the set CP (lines 11 and 12).

Algorithm 1
Extracted changing points algorithm.

Require: Sightseeing video : v, Number of all frames : N, Number of frames in a *shot* : n, Percentage of extraction : p, Number of frame slide : δ
Ensure: Set of changing points : CP
1: Divide all frames in v to 9 areas.
2: For each i $(1 \leq i \leq 9)$ and for each frame j $(1 \leq j \leq N)$, calculate a color histogram h_j^i
3: $t \leftarrow 1$
4: $s \leftarrow 1$
5: **while** $t < N$ **do**
6: For each i $(1 \leq i \leq 9)$, calculate H_s^i by averaging $h_t^i, ..., h_{t+n-1}^i$
7: $t \leftarrow t + \delta$
8: $s \leftarrow s + 1$
9: **end while**
10: For each i $(1 \leq i \leq 9)$ and s $(1 \leq s \leq \lfloor N/n \rfloor)$, calculate correlation C_s^i between H_s^i and H_{s+1}^i.
11: $CP \leftarrow \{s \mid \text{more than one areas such that } C_s^i \text{ is in lower } p\%\}$
12: Output CP

3.2 Segmentation of shots

A shot indicates a set of *n* frames. The value of color histograms is unaffected by small changes of the scene or frames because this value is the average in a shot. The reason for using the average is to disregard the effects of crowds or camera shakes. It is also desirable that the division of a frame is small to improve the processing time. Preliminary experiments showed that changing points are affected by small changes if the frames are not divided. When we tried several division patterns (numbers) of frames, taking into account the processing time and detection performance, we found that the division of a frame into 9 areas is the best. In addition, tour videos used by our method are typically taken while walking. Video scenes do not change largely between 1 and 5 s of playback time because the normal human walking speed is about 1 m/s. In this case, it is appropriate that the number of frame slides δ is only between the numbers of frames in 1 and 5 s, considering the number of frames per second (if the FPS is 30, δ is between 30 and 150; if the FPS is 60, δ is between 60 and 300).

3.3 Histogram correlation of adjacent shots

A 256-level gray-scale color histogram is used in our method. A histogram indicates a distribution of the luminance level of pixels. For each screen area *i*, the average of a color histogram in shot H^i (i = 1, ..., 9) is calculated using Eq. (1) when the value of the color histogram of each frame is h_j^i [represented by a vector of 256 values (occurrences) in each area].

$$H^i = \frac{\sum_{j=1}^{n} h_j^i}{n} \tag{1}$$

Here, n is the number of frames in a shot. Similarly, H^i is calculated for all shots. In the proposed method, the correlation coefficient C^i is calculated using Eq. (2) when histograms of the two adjacent shots H^i and H'^i are given. Let $x^i_{k,l}$ and $y^i_{k,l}$ denote the lth element value of the kth frames in the histograms of H^i and H'^i, respectively. Then,

$$C^i = \frac{\sum_l \sum_{k=1}^{n} \left(x^i_{k,l} - \overline{x^i_l}\right)\left(y^i_{k,l} - \overline{y^i_l}\right)}{\sqrt{\sum_l \sum_{k=1}^{n} \left(y^i_{k,l} - \overline{y^i_l}\right)^2} \sqrt{\sum_l \sum_{k=1}^{n} \left(x^i_{k,l} - \overline{x^i_l}\right)^2}} \cdot \tag{2}$$

Here, $\overline{x^i_l}$ and $\overline{y^i_l}$ are the arithmetic means of $x^i_{k,l}$ and $y^i_{k,l}$ in the shot with n frames, respectively. C^i is calculated for all adjacent shots using Eq. (2).

3.4 Detection of changing points

Changing points are detected by using C^i calculated in Sect. 3.3. C^i (i = 1, 2, ..., 9) exists along the time axis because all frames are divided into 9 areas. For each area i, the bottom p% of values among all shots are selected as candidates of changing points. For each shot, when more than one candidates is selected in 9 areas of the shot, we define this shot as a changing point. Here, p is empirically chosen depending on the desired length of the summarized video.

3.5 Creation of summarized video

There are similar views between two changing points. To emphasize the beginning of a scene, we set the playback to a lower speed (a bit faster than the normal speed) around changing points, while playing back the remaining parts by fast-forwarding (time-lapsing).

3.6 Use case

Figure 3 shows the image of our use case. The system consists of map information, sightseeing spot information, and a summarized video. The user selects the spots he/she wants to go to from the map. The system builds a route based on the user's selection and displays a summarized video along that route.

4. Evaluation

We conducted quantitative and qualitative experiments to evaluate the effectiveness of our method. We first confirmed the appropriate playback speed of sightseeing videos. We then evaluated the relevance of our scene extraction method compared with the ground truth annotated by participants. Finally, we evaluated the usefulness of the summarized video created by our method through a user study. In this experiment, we used a video taken in Kyoto, Japan, whose duration is 6 min 58 s. We considered that this video is suitable for this experiment

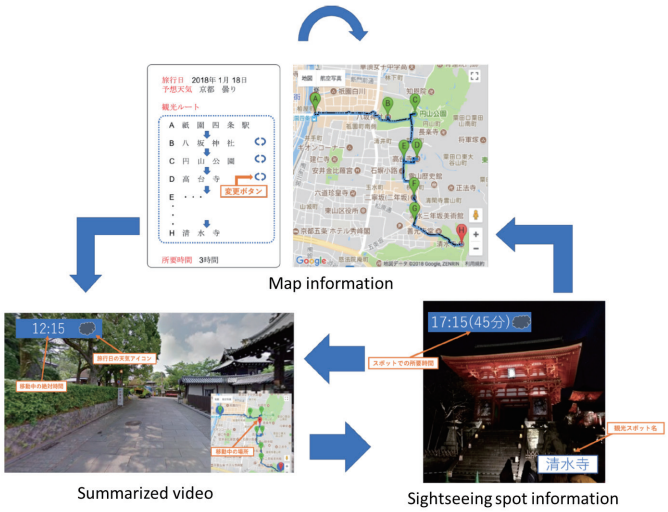

Fig. 3. (Color online) Image of our use case.

because it has various scene changes such as street corners, street stores, and crowds. We recruited 20 participants (all graduate students in twenties, male: 16, female: 4).

4.1 Evaluation of playback speed

The purpose of this experiment was to evaluate the appropriate playback speed of sightseeing videos. Participants compared videos with different speeds with respect to the comprehensibility of (i) the distance of the route, (ii) street corners in the route, (iii) stores in the route, and (iv) the atmosphere of the route. Participants evaluated them with a 7-level Likert scale. In addition, we asked them to evaluate the feeling while watching each video with the same 7-level Likert scale. We prepared ×4, ×8, ×16, ×25, ×40, ×50, and ×75 speed videos and the original speed (×1) video. Participants watched the videos first at the original speed and then at the higher speeds, and finally, they evaluated each of them. In this experiment, the order of videos shown was randomized to remove biases.

4.2 Quantitative evaluation of scene extraction

The purpose of this experiment was to evaluate the accuracy of the video summarization. We prepared 84 video segments by dividing the original video by 5 s intervals. The participants determined whether each segment is necessary (1) or unnecessary (0) by watching videos, supposing that they were planning a sightseeing tour. We evaluated the classification accuracy of the proposed method by using their answers as ground truth. The original video was taken while walking. We determined 5 s as the appropriate length of each segment because scenes do not change largely in 5 s. We empirically determined δ as 60 to obtain the average of 1 s because the original video's FPS was 60. Also, we empirically determined that $n = 300$ and $p = 15$.

4.3 Evaluation result on playback speed

Figure 4 shows the results for the questions regarding playback speed. Here, comprehensibility becomes 1.0 when watching the video at the original speed. As playback speed increases, the degree of information comprehension decreases. In particular, the level of understanding street store information was lower than the level of understanding other information. To increase the degree of store information comprehension, more information is needed, for example, we can add the text of the stores in the videos. We found that for the atmosphere of the route, an increase in video speed has a smaller effect on other information. From the results in Fig. 4, we consider that the appropriate playback speed is approximately between ×4 and ×8 when summarizing sightseeing videos.

Figure 5 shows the results of how participants feel about each video length. They answered 1 if they felt the video was too quick and 7 if they felt it was too slow. Playback speeds of ×4 and ×8 have average scores of 4 (mid-point), meaning that the participant found the video length to be most acceptable. However, some participants preferred the video length when playback speeds were ×25, ×40, ×50, and ×75. This suggests that we may need to change the playback speed depending on the viewer and/or the length of the original video. Furthermore, some participants answered that the feeling of length may change depending on their preference. Thus, when we suggest a sightseeing movie, it is important to select the most interesting scenes for watchers. Also, we can make a summarized video more satisfactory by changing the video speed according to the watcher's desired information.

4.4 Extraction accuracy

Figure 6 shows the results of participants determining what parts are necessary or unnecessary in creating a sightseeing video. Many of the participants answered that the scenes where a cameraman turned the corner (at time 70–80 s and 270–280 s) and where many street stores appear on the scene (at time 350–390 s) are necessary. Many participants selected

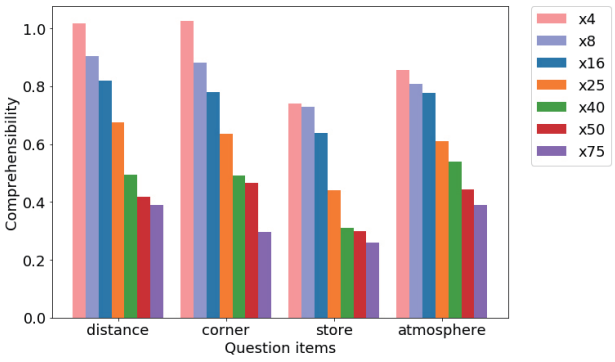

Fig. 4. (Color online) Comprehensibility of each information while watching videos.

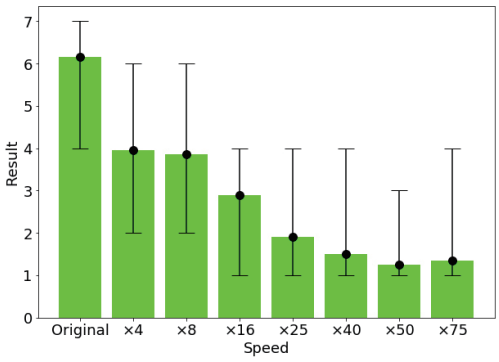

Fig. 5. (Color online) Results of how participants felt about video length. Lines indicate the range between the lowest and highest scores.

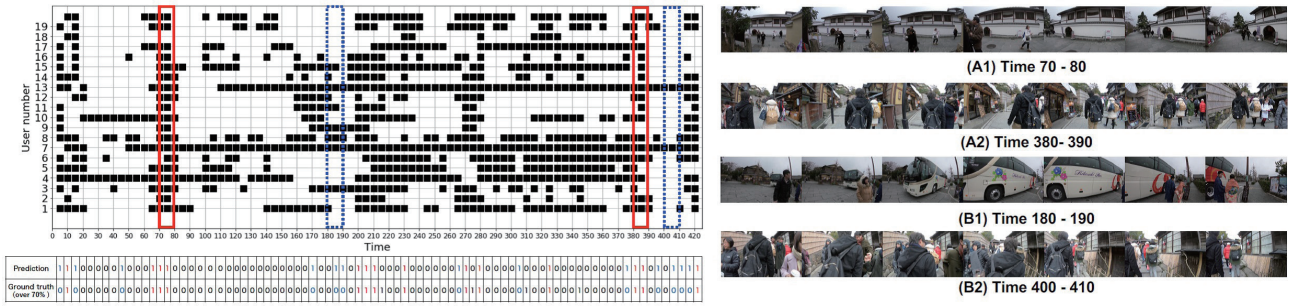

Fig. 6. (Color online) Comparison between manual and automatic (our method) video summarizations.

the latter half of the video as necessary owing to the presence of many stores and the better sightseeing atmosphere. This suggests that the information on street stores and sightseeing atmosphere are important to watchers.

Figure 6 also shows the comparison between manual and automatic (our method) summarizations. We compared changing points extracted as necessary by the proposed method with the parts that 70% or more participants selected as necessary. The bottom left chart of Fig. 6 shows the classification results. As seen in the chart, our method identified necessary scenes with an F-measure of 62.22%. In the left upper chart of Fig. 5, examples of correct extractions of necessary scenes (A1, A2) are marked by a solid red line, while those of incorrect extractions (B1, B2) are marked by a dotted blue line. We see that our method extracted street corners well like scene A1. Also, the scenes reflecting the presence of street stores like scene A2 were extracted correctly. However, our method extracted a scene like B1 where a bus is present as necessary since the frame change is large, but many participants selected this scene as unnecessary because it has no sightseeing information. Furthermore, scene B2 is an example of undesirable extraction, where a scene of crowded people was extracted because our method is affected by the color of clothes worn by people. Certain situations where the cameraman could not go straight because of crowds were incorrectly extracted as necessary owing to a large frame change.

We asked the participants why they selected particular unnecessary scenes. The participants said, “Similar scenes were continued,” “Couldn’t understand the route because of the crowd,” and “Showed scenes unrelated to sightseeing.”

5. User Study of Summarized Video

We made a summarized video using the findings in Sect. 4. The same 20 participants watched the summarized video. We used the same video as used in the previous experiments. We summarized this video by using the results in Sect. 4.4 and the length of the summarized video was 41 s (the original video was 6 min 58 s). Because the playback speed between ×4 and ×8 was the best according to Sect. 4.3, necessary scenes were played back at ×4, while unnecessary scenes were played back at ×32. All participants answered a questionnaire after watching the summarized video. Table 1 shows the results of the questionnaire. Questions

Table 1
Results of questionnaire in user study.

Item	Question	Answer (1: worst, 7: best)							Average	Deviation
		1	2	3	4	5	6	7		
Q1	Could you understand this route intuitively?	0	0	0	2	7	7	4	5.65	0.93
Q2	Are necessary scenes selected appropriate?	0	2	1	2	4	8	3	5.20	1.51
Q3	Is this video effective for planning sightseeing?	0	3	2	1	8	5	1	4.65	1.50

Table 2
Results of questionnaire in user study 2.

Item	Question	Video 1		Video 2		Video 3	
		Avg.	SD	Avg.	SD	Avg.	SD
Q4	Could you image this route intuitively?	4.94	1.39	5.10	1.55	4.28	1.81
Q5	Are necessary scenes selected appropriate?	4.7	1.45	4.84	1.30	4.32	1.54
Q6	Would this video support planning sightseeing?	4.80	1.47	4.96	1.43	4.50	1.69

were answered on a scale of one (worst) to seven (best). The average results for the three questions were 5.65 (Q1), 5.2 (Q2), and 4.65 (Q3).

Over 75% of participants answered that necessary scenes selected by our method were appropriate (Q2). This result confirms that the summarized video made by our method is effective.

Furthermore, three types of videos were created and compared to validate the usefulness of this method. The created videos were a summarized video based on a video recorded by the author (Video 1), a summarized video based on a video recorded by a non-author (Video 2), and a video edited manually by a production company (Video 3). Each of these three videos was watched by 50 people, who also answered a questionnaire with answers on a scale of one (worst) to seven (best).

Table 2 shows the results. According to the a result of one-way ANOVA, there was no significant difference between the three videos except for between Videos 2 and 3 in Question 4 ($p < 0.05$). The results show that the proposed method is effective for videos recorded by non-authors and has almost the same effect as manual editing.

6. Conclusions

In this paper, we proposed a method of video summarization based on scene changes. We implemented video summarization by fast-forwarding scenes with small scene changes. We evaluated the playback speed of the fast-forwarded sightseeing video and found that a speed between ×4 and ×8 is the best. Also, our method correctly identified necessary scenes in a sightseeing video compared with those selected as necessary by over 70% of participants, with an F-measure of 62.22%. Over 75% of participants answered that the summarized video was effective for planning a sightseeing tour. As a result, we believe that our method is effective in summarizing sightseeing videos. As part of future work, we will try to improve detection accuracy and apply our method with various videos taken in various sightseeing spots.

Acknowledgments

This work was in part supported by JSPS KAKENHI JP16H01721.

References

1 Y. Kurata, Y. Shinagawa, and T. Hara: Workshop on Tourism Recommender Systems (2015).
2 M. Hidaka, Y. Matsuda, S. Kawanaka, Y. Nakamura, M. Fujimoto, Y. Arakawa, and K. Yasumoto: 2nd Int. Workshop on Smart Sensing Systems (IWSSS'17) (2017).
3 D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, and N. Vathis: Pers. Ubiquitous Comput. **21** (2017) 137.
4 Y. Zhang, H. Ma, and R. Zimmermann: Int. Conf. Multimedia Modeling (Springer, 2013) 380.
5 X. Lu, C. Wang, J. M. Yang, Y. Pang, and L. Zhang: Proc. 18th ACM Int. Conf. Multimedia (ACM, 2010) 143.
6 M. Korakakis, P. Mylonas, and E. Spyrou: 11th Int. Workshop on Semantic and Social Media Adaptation and Personalization (SMAP) (IEEE, 2016) 59.
7 Y. Arakawa: Proc. 2014 Int. Workshop on Web Intelligence and Smart Sensing (ACM, 2014) 1.
8 C. Y. Sun and A. J. Lee: Decis. Support Syst. **101** (2017) 28.
9 Y. Zhang, H. Ma, and R. Zimmermann: Int. Conf. Multimedia Modeling (Springer, 2013) 380.
10 Y. Zhang, L. Zhang, and R. Zimmermann: ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **11** (2015) 24.
11 F. Jing, L. Zhang, and W.-Y. Ma: Proc. 14th ACM Int. Conf. Multimedia (ACM, 2006) 599.
12 Ipsos MediaCT: The 2014 Traveler's Road to Decision, Google Travel Study (2014).
13 Y. Kanaya, S. Kawanaka, M. Hidaka, H. Suwa, Y. Arakawa, and K. Yasumoto: 3rd Int. Workshop on Smart Sensing Systems (IWSSS'18) (2018).
14 I. A. Zedan, K. M. Elsayed, and E. Emary: in Advances in Soft Computing and Machine Learning in Image Processing (Springer, Cham, 2018) p. 89.
15 K. Fujisawa, Y. Hirabe, H. Suwa, Y. Arakawa, and K. Yasumoto: Int. J. Multimedia Data Eng. Manage. (IJMDEM) **7** (2016) 36.
16 H. Jin, Y. Song, and K. Yatani. Elasticplay: Proc. 2017 ACM Conf. Multimedia (ACM, 2017) 1164.
17 K. Yasumoto, H. Yamaguchi, and H. Shigeno: J. Inf. Process. 24 (2016) 195.
18 S. Jai-Andaloussi, I. El Mourabit, N. Madrane, S. B. Chaouni, and A. Sekkaki: Int. Conf. Computational Science and Computational Intelligence (CSCI) (IEEE, 2015) 398.
19 S. Jai-Andaloussi, A. Mohamed, N. Madrane, and A. Sekkaki: Int. Symp. Big Data Computing (BDC) (IEEE, 2014) 1.
20 S. Morishita, S. Maenaka, D. Nagata, M. Tamai, K. Yasumoto, T. Fukukura, and K. Sato: Proc. 2015 ACM Int. Joint Conf. Pervasive and Ubiquitous Computing (ACM, 2015) 695.
21 M. Okamoto and K. Yanai: Pacific-Rim Symp. Image and Video Technology (Springer, 2013) 431.