2017 IEEE 31st International Conference on Advanced Information Networking and Applications

Movie Summarization based on Alignment of Plot and Shots

Xuehshan Li
*Graduate School of Systems
and Information Engineering
University of Tsukuba
Tsukuba, 305-8573, Japan*

Takehito Utsuro
*Graduate School of Systems
and Information Engineering
University of Tsukuba
Tsukuba, 305-8573, Japan*

Hiroshi Uehara
*Graduate School of Systems
and Information Engineering
University of Tsukuba
Tsukuba, 305-8573, Japan
Corporate Sales and Marketing Division
NTT DOCOMO, INC.,Tokyo, 100-6150, Japan*

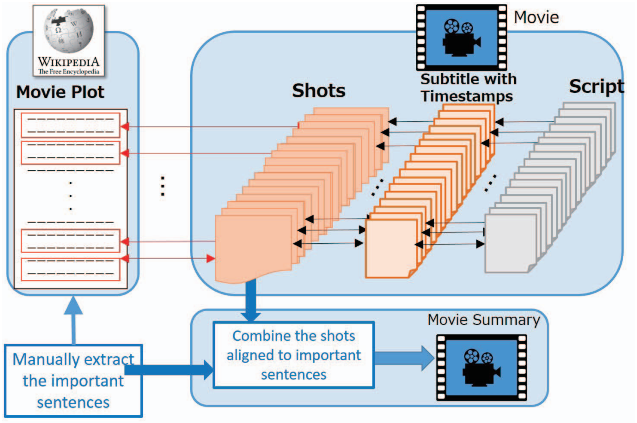


Figure 1. Flow of movie summarization

***Abstract*—This paper proposes a method of assisting movie summarization using plot information. A plot of a movie available at Wikipedia contains a major story of the movie. From such a plot of a movie, we extract several important sentences as the content of summary. For summarizing movie, the key work is finding the best alignment between sentences of plot and shots which are segmented from a movie. There are two cues used to measure the similarity between a sentence and a shot. One is based on character appearing in both sentence and shot, another is based on words matching. Then an alignment method based on dynamic programming is apply to optimize the alignment. Finally an experiment on movie *Roman Holiday* and *Alice in the wonderland* show the effectiveness of this method.**

***Keywords*-movie summarization; alignment; character identification; dynamic programming**

I. Introduction

Nowadays, as a kind of hit entertainment, large amounts of movies have been produced. An efficient way to select a piece of movie that suit one's taste is required. To this end, movie summarization like trailers are popular leading to increasing necessity of producing movie summarization. However, trailers mainly show the intense scenes for audiences, and contribute little to understanding the whole story of the movie. Our focus, as shown in Figure 1, the flow of movie summarization assist, is to extract the shots which are corresponding to the important sentences from Wikipedia item, and to merge the extracted shots as movie summary as well. Our work is of great interest to those:

1) Who want to write a report of a movie that has been watched already.
2) Who want to recommend or propagate a movie to others.
3) Who have trouble selecting a new movie to watch.

This paper is presented as follows. First related work is summarized in Section II and the data are introduced in Section III. Followed by an overall introduction of process in Section IV, we discuss the automatic alignment between plot and shot in Section V. At the beginning part Section V-A, we introduce a novel way to identify the movie character by using only textual data, subtitle with time information and script with names of characters. Then in Section V-B, we introduce an alignment method based on dynamic programming, which employs two cues of similarities. One is character identification similarity and another one is words similarity. The result and evaluation are presented in Section V-C. Finally, we present conclusion and future perspective in Section VI.

II. Related Work

A video summarization is a summary as an abstract of a movie. Based on the alignment between subtitle and movie. Liang et al. [1] proposed a method of scene segmentation by aligning the names in subtitle and face images in the movie. Yi et al. [2] processes movie summarization by extracting semantical subtitles which have been converted to vectors. In Tsoneva et al. [3], at the first scene segmentation is implemented by the alignment between script and subtitles with time information, then movie summarization is processed by ranking the scenes by importance. Moreover, in Tapaswi et al. [4], shots are aligned to every plot by two cues of similarities: character similarity and words similarity.

1550-445X/17 $31.00 © 2017 IEEE
DOI 10.1109/AINA.2017.127

Table I
MOVIE EXAMPLES

Movie	Number of plot sentences	Number of important sentences	Number of shots after automatically segmented
Roman Holiday	53	11	649
Alice in the wonderland	47	10	1,581

III. RESOURCES

A. Movie Plot from Wikipedia

The movie plot posted in Wikipedia, generally from hundreds to thousands words, contains the most importance content of movie story. Figure 2 shows a part of plot of the movie *Roman Holiday* in Wikipedia.

B. Shot Segmentation of Video

As the bottom part in Figure 4 shows the continuous images show the concept of shot in movie which indicates the uninterrupted sequence between two edits or cuts. Generally, contents in a shot, for instance, a static view, continuous movement of the car, and a series actions of drinking water, rarely changes. Here, we focus on internally stable shots, and regard every shot as a static image allowing us to align it with plot very efficiently. For segmenting the movie to shots, we apply the movie segmentation tool proposed in Apostolidis et al. [5]. In addition, for the shots longer than one minute, we segment it twice to avoid over-long shots.

C. Subtitle with Time stamps and Script

Subtitles, usually displayed at the bottom of the screen, are derived from either a transcript or screenplay of the dialog or commentary in films, television programs, video games. The subtitle we use is a textual file with playing time information in the beginning of every line of dialog. A script, also known as screenplay, records all the dialog in the movie and descriptions of characters actions and background environment and or so. The names of characters in the movie are written in the script where followed by the dialog contents.

IV. ASSISTING METHOD OF MOVIE SUMMARIZATION

Figure 1 shows our method of assisting movie summarization by aligning plot sentences and shots. At first, we collect a movie's plot from that Wikipedia entry. Then manually extract the important sentences based on some predefined principles. The movie will also be segmented by an effective tool in Apostolidis et al. [5]. A 2-hour movie usually can be segmented to hundreds of shots. Our goal is to extract those shots which are aligning to important sentences of plot then combine them together to make its movie summarization. Figure 3 shows the manual correspondence result between important sentences and shots. To this end, there are two cues used to align the sentences of plot and shots. One cue is character both occur in plot and shots, another cue is word co-occurrence. We present the detail of alignment in Section V.

A. Extraction of important sentences from movie plot

Specifically, numbers of sentences contained in a movie plot is somehow too many. For example plot of *Roman Holiday* contains 53[1] sentences. Thus we select important sentences based on principles below.

1) The important sentences constitute the main structure of the story
    a) Opening
    b) Starting point of main story
    c) Scene changing
    d) End

    are selected.
2) In order to grasp the details of the progress of the story, the sentences other than the four types of important sentences mentioned before are also selected. Upper part of Figure 4 shows the selection of important sentences and the shots that the 5th important sentence should be aligned to.

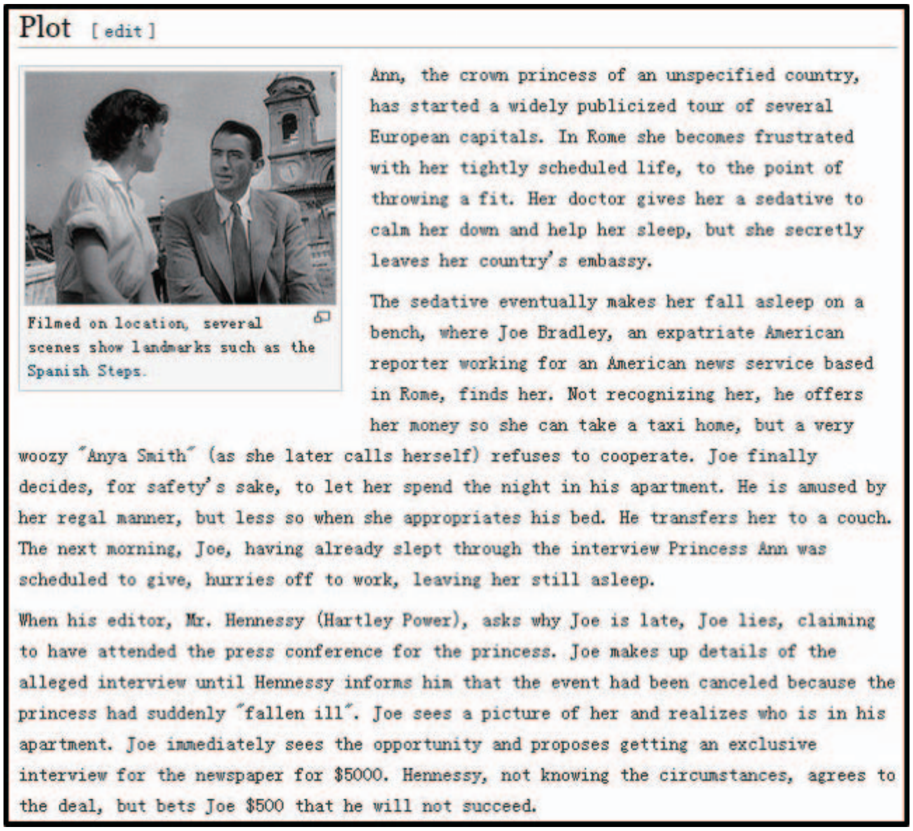
V. AUTOMATIC ALIGNMENT BETWEEN PLOT AND SHOT

In this section, we apply the method from Tapaswi et al. [4] to achieve a good alignment between shots and plot sentences. Character similarity and words similarity are two powerful cues to measure how similar a plot sentence and a shot in the movie are . Then we combine these two similarities and give the words similarity a coefficient to adjust its weight to optimize the fusion similarity. Finally, we address this alignment problem by adopting dynamic programming method. As a result, every shot is aligned to a sentence.

A. Character Identification of Shot

In the field of computer vision, movie character identification is a popular problem which is solved by image processing techniques. However, the computing process is sometimes time-consuming, and the results are not appealing. Therefore, we make full use of text data related to the

[1] For *Roman Holiday*, the number of sentences in the plot posted in Wikipedia is actually 39, but better evaluation results were obtained by dividing the long sentence into two sentences. Thus the number of plot sentences is 53 in evaluation. The movie *Alice in the wonderland* is also handled in the same way.



Plot [edit]

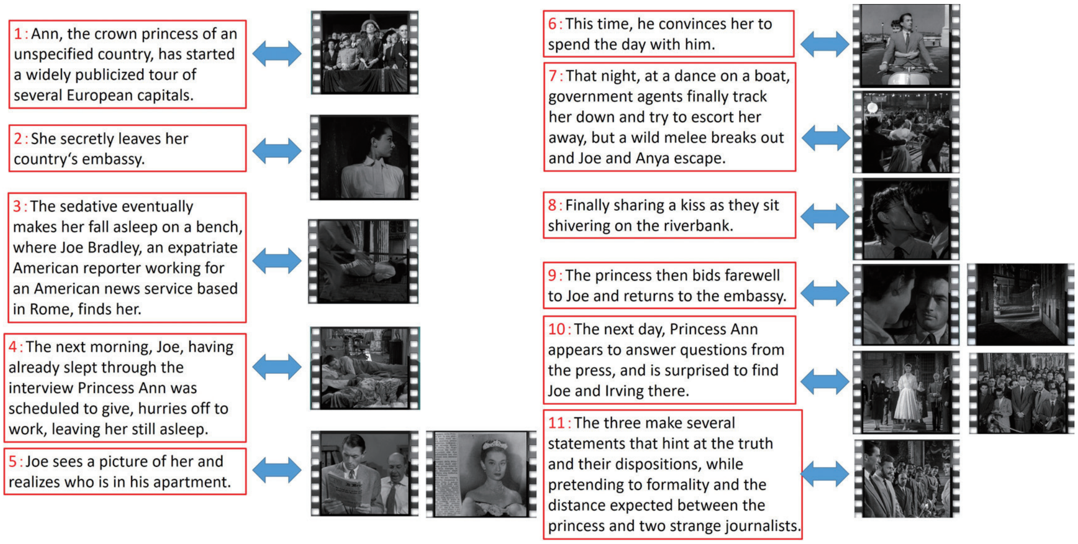
Ann, the crown princess of an unspecified country, has started a widely publicized tour of several European capitals. In Rome she becomes frustrated with her tightly scheduled life, to the point of throwing a fit. Her doctor gives her a sedative to calm her down and help her sleep, but she secretly leaves her country's embassy.

Filmed on location, several scenes show landmarks such as the Spanish Steps.

The sedative eventually makes her fall asleep on a bench, where Joe Bradley, an expatriate American reporter working for an American news service based in Rome, finds her. Not recognizing her, he offers her money so she can take a taxi home, but a very woozy "Anya Smith" (as she later calls herself) refuses to cooperate. Joe finally decides, for safety's sake, to let her spend the night in his apartment. He is amused by her regal manner, but less so when she appropriates his bed. He transfers her to a couch. The next morning, Joe, having already slept through the interview Princess Ann was scheduled to give, hurries off to work, leaving her still asleep.

When his editor, Mr. Hennessy (Hartley Power), asks why Joe is late, Joe lies, claiming to have attended the press conference for the princess. Joe makes up details of the alleged interview until Hennessy informs him that the event had been canceled because the princess had suddenly "fallen ill". Joe sees a picture of her and realizes who is in his apartment. Joe immediately sees the opportunity and proposes getting an exclusive interview for the newspaper for $5000. Hennessy, not knowing the circumstances, agrees to the deal, but bets Joe $500 that he will not succeed.

Figure 2. A part of plot of the movie *Roman Holiday*



1: Ann, the crown princess of an unspecified country, has started a widely publicized tour of several European capitals.

2: She secretly leaves her country's embassy.

3: The sedative eventually makes her fall asleep on a bench, where Joe Bradley, an expatriate American reporter working for an American news service based in Rome, finds her.

4: The next morning, Joe, having already slept through the interview Princess Ann was scheduled to give, hurries off to work, leaving her still asleep.

5: Joe sees a picture of her and realizes who is in his apartment.

6: This time, he convinces her to spend the day with him.

7: That night, at a dance on a boat, government agents finally track her down and try to escort her away, but a wild melee breaks out and Joe and Anya escape.

8: Finally sharing a kiss as they sit shivering on the riverbank.

9: The princess then bids farewell to Joe and returns to the embassy.

10: The next day, Princess Ann appears to answer questions from the press, and is surprised to find Joe and Irving there.

11: The three make several statements that hint at the truth and their dispositions, while pretending to formality and the distance expected between the princess and two strange journalists.

Figure 3. Manual correspondence result between important sentences of plot and shots.

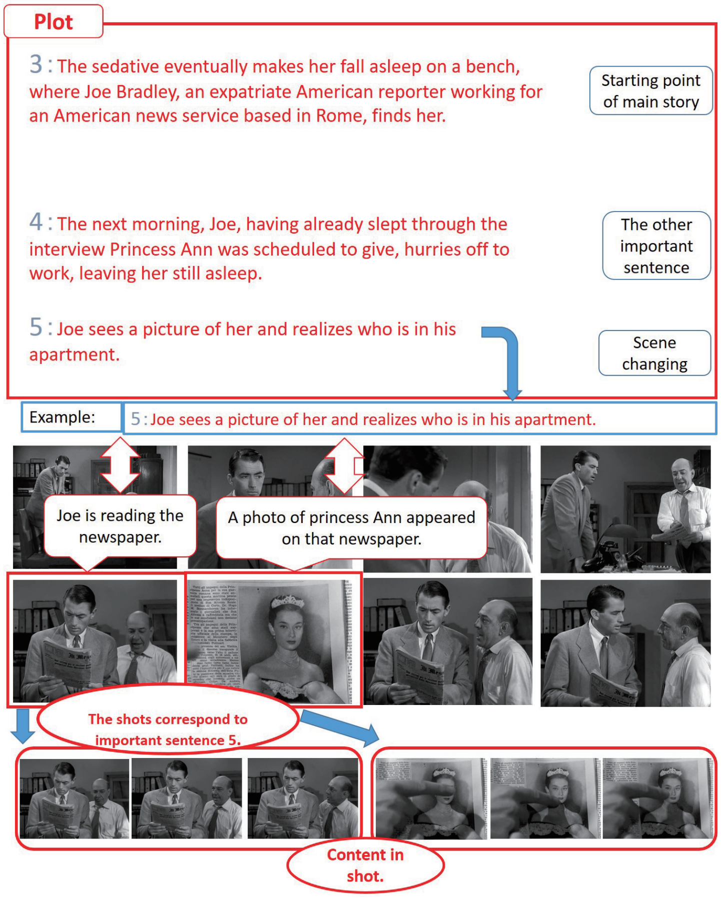

Plot
3 : The sedative eventually makes her fall asleep on a bench, where Joe Bradley, an expatriate American reporter working for an American news service based in Rome, finds her.
Starting point of main story
4 : The next morning, Joe, having already slept through the interview Princess Ann was scheduled to give, hurries off to work, leaving her still asleep.
The other important sentence
5 : Joe sees a picture of her and realizes who is in his apartment.
Scene changing
Example:
5 : Joe sees a picture of her and realizes who is in his apartment.
Joe is reading the newspaper.
A photo of princess Ann appeared on that newspaper.
The shots correspond to important sentence 5.
Content in shot.

Figure 4. An example of alignment between plot sentences and shots in movie *Roman Holiday*

movie to identify the character only by subtitle and script data introduced in Section III.

1) Alignment between Subtitle with Time stamps and Script: Both subtitle (dialogs with time stamps) and script (dialogs with names) share dialogs content. Thus we can align sentences in subtitle and script by matching words so that we know when (time stamp) who (name) spoke.

2) Alignment between Shots and Subtitle with Time stamps: We can easily align subtitle sentences to shots only according to the time stamps.

B. Alignment used Dynamic Programming

In this paper, we apply the method proposed by Tapaswi et al. [4] to automatically align plot sentences and shots. The main formula in Tapaswi et al. [4] is as below. Similarity between a sentence s_i of plot and a shot t_j is defined as f_{fus}, then dynamic programming is applied by the f_{fus} value to optimize alignment between sentences of plot and shots. Formula (1) contains two similarities. The first one is the similarity of character co-occurring in both plot sentence and shot, another is the similarity of words (without stop-words) co-occurring in both plot sentence and shot. α is the weight parameter to adjust two similarities.

$$f_{fus}(s_i, t_j) \quad = \quad f_{id}(s_i, t_j) + \alpha \cdot f_{subtt}(s_i, t_j) \qquad (1)$$

1) Identification Cue: For f_{id} the similarity of characters occurring in sentence s_i of plot and characters occurring in shot t_j, we first apply a simple voting method as formula (2) shows.

$$\text{align}(c, d) = \begin{cases} 1 & (c = d) \\ 0 & (others) \end{cases} \qquad (2)$$

c is the character occurring in shot t_j, C is a set of all characters in movie, d is the name wrote in sentence s_i of plot, D is a set of all names wrote in plot. This formula is based on the alignment between sentences of plot and shots introduced before. Then use this we consider the similarity of characters in sentence s_i of plot and shot t_j as formula (3).

$$f_{id}(s_i, t_j) \;=\; \sum_{k=j-r}^{j+r} \sum_{c \in C_j} \sum_{d \in D_i} \text{align}(c, d) \cdot I(c) \quad (3)$$

Time length of a shot is somehow short (most are shorter than 10 seconds) that character easily appear in the neighbor shots so we set a range of shots to limit this deviation. C_j is a set of all characters appear in the range of shots. D_i is a set contains all names in plot sentence s_i. For a particular character c, the frequency of its occurrence is negatively related to its importance. It is like the TF-IDF as below.

$$I(c^*) \;=\; \frac{\log\big(\max_{c \in C} n_{FT}(c)\big)}{\log\big(n_{FT}(c^*) + 1\big)}$$

$n_{FT}(c)$ is frequency of c in movie.

2) Word Co-occurrence Cue: After the alignment between subtitle and shots, we can measure the word similarity between sentence s_i in plot and shot t_j by counting the number of matches between words v in sentence s_i with w in the subtitle sentences that are assigned to shot t_j.

$$f_{subtt}(s_i, t_j) = \sum_{v \in s_i} \sum_{w \in subtt \in t_j} \text{word-match}(v, w)$$

$$\text{word-match}(v, w) = \begin{cases} 1 & (v = w) \\ 0 & (v \neq w) \end{cases}$$

C. Evaluation

We apply the proposed method to evaluate *Roman Holiday* and *Alice in Wonderland.* The number of plot sentences and shots on the two movies are shown in the Table I. However, in order to estimate the upper bound of the performance, we manually rewritten the pronoun in the plot to the corresponding character name and then evaluate.

1) When all shots are aligned to all plot sentence of plot: In Tapaswi et al. [4], in spite of the number of candidate shots being tens of times of the number of plot sentences, the constraint that all shots are aligned to all plot sentences is imposed. In Tapaswi et al. [4], in order to minimize the adverse effect caused by this constraint, an upper bound on the number of shots that can correspond to one sentence in the plot is determined by the equation below. In this paper, we also set the same upper bound as the method in Tapaswi et al. [4]. In that method, N_T is the total number of shots and N_S is the total number of plot sentences, the average number N_T/N_S of shots aligning to a plot sentence multiplied by the parameter k, then the result z is the upper bound of the number of shots aligning to a plot sentence.

$$z = k \cdot N_T/N_S$$

In this paper, for *Roman Holiday*, summarization performance is optimal when $k = 5$ in the case of $N_T = 649$ and $N_S = 53$, and the upper bound is $z = 61$. Also for *Alice in the Wonderland*, $k = 5$ is also applied as the optimal parameter, thus the upper bound $z = 168$ in this movie case of $N_T = 1,581$ and $N_S = 47$.

2) When at most three shots are aligned to one sentence of plot: Instead of aligning all the shots to all plot sentences, we introduce a method of carefully selecting a small number of shots which are aligned to each plot sentence.

Specifically, unless the number of system output shots aligning to the plot sentence s_i is 3 or less, before aligning them to the plot sentences, three consecutive shots in the system output shots will be selected by the following equation. N_i is the number of system output shots which gained from the procedure of the previous section. Then, the result $T_3(s_i)$, the shot sequence with the maximum sum of f_{fus}, is taken as the shots that should be aligned to the plot sentence.

$$T_3(s_i) = \underset{l=0\sim N_i-2}{\text{argmax}}\,\Big(\sum_{l=j}^{j+2} f_{fus}(s_i, t_l)\Big)$$

The alignment between the plot sentences and the shots made by the two methods of the previous section and this section is shown in the Figure 5 and the Figure 6. In the Figure 5 , the number of shots aligning to the tenth plot sentence is the maximum of 51 by the method of the previous section. Therefore, the length of the vertical line representing the number of shots is the longest. On the other hand, since the number of shots aligning to the plot sentence is no more than 3 according to the procedure in this section, the shot position is indicated by a dot. As can be seen from these results, even if the movie summary result is greatly shortened by the method of this section, the content of the original movie can be grasped is kept almost the same.

D. Automatic Evaluation Result

In order to automatically evaluate the movie summary, we manually select no more than 3 shots aligning to one important sentence. In the automatic evaluation of the movie summary result, for each plot sentence, if the difference between the position of any shot in the shot sequence of the summary result and the position of any shot in the correct shot sequence no more than 1, it is determined that the shot sequence aligned to the plot sentence is correct. The evaluation results for *Roman Holiday* and *Alice in the Wonderland* are shown in Table II, Table III, and below. Evaluation of each important plot sentence can be seen in Table IV and Table V.

1) When all shots are aligned to all plot sentence of plot

$$\textit{Roman Holiday}: \; 6/11 = 54.5\%$$
$$\textit{Alice in the Wonderland}: \; 8/10 = 80.0\%$$

Table II
SUMMARY RESULTS AND EVALUATION RESULTS (WHEN ALL SHOTS ARE ALIGNED TO ALL PLOT SENTENCE OF PLOT)

Movie	Time length of the movie	Time length of the summary result	Summarization rate	Precision
Roman Holiday	1:58:11	00:18:53	15.98%	54.5%
Alice in the Wonderland	1:15:10	00:14:43	19.58%	80.0%

Table III
SUMMARY RESULTS AND EVALUATION RESULTS (WHEN AT MOST THREE SHOTS ARE ALIGNED TO ONE SENTENCE OF PLOT)

Movie	Time length of the movie	Time length of the summary result	Summarization rate	Precision
Roman Holiday	1:58:11	00:03:03	2.58%	18.2%
Alice in the Wonderland	01:15:10	00:01:23	1.84%	50.0%

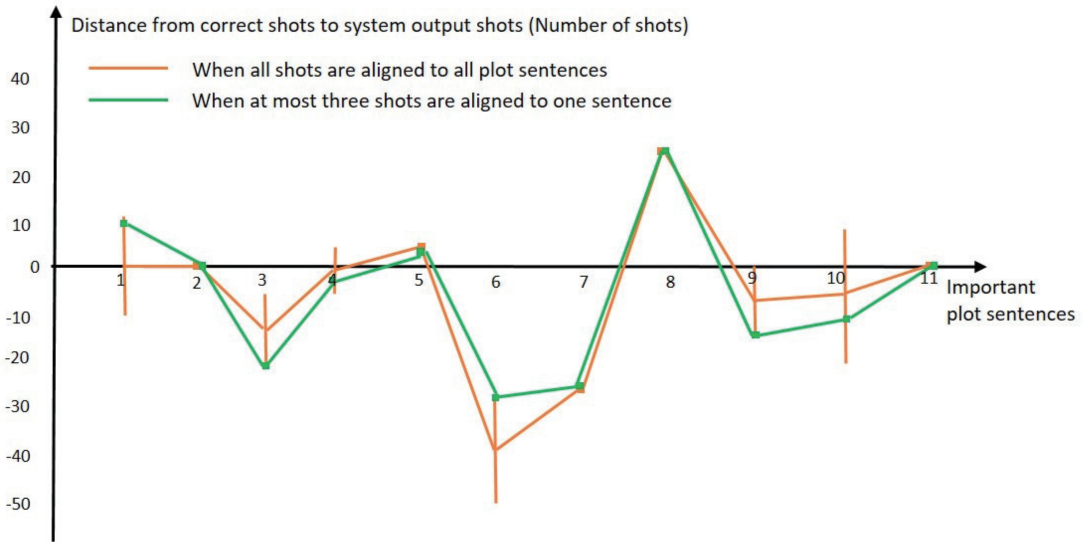


Figure 5. Distance from correct shots to system output shots (*Roman Holiday*)

2) When at most three shots are aligned to one sentence of plot

$$\textit{Roman Holiday}:\ 2/11 = 18.2\%$$
$$\textit{Alice in the Wonderland}:\ 5/10 = 50.0\%$$

E. Subjective Evaluation Result

Next, as a subjective evaluation of the movie summary result, when viewing the video of the shots aligned to each plot sentence, it is determined whether or not main information on each plot sentence has been obtained as a judgement criterion, and a subjective evaluation of the movie summaries is shown in Table VI, Table VI and below.

1) When all shots are aligned to all plot sentence of plot

$$\textit{Roman Holiday}:\ 8/11 = 72.7\%$$
$$\textit{Alice in the Wonderland}:\ 8/10 = 80.0\%$$

2) When at most three shots are aligned to one sentence of plot

$$\textit{Roman Holiday}:\ 6/11 = 54.5\%$$
$$\textit{Alice in the Wonderland}:\ 7/10 = 70.0\%$$

As shown in this subjective evaluation results, improvement can be found in these two cases compared with the automatic evaluation results. From these results, it was found that movie summary including certain information can be made by the proposed method.

VI. CONCLUSIONS

In this paper, we present a method for movie summarization by aligning the plot sentences and movie shots. The plot is from Wikipedia and the shots are the uninterrupted cuts segmented from movie. In the future, based on the method

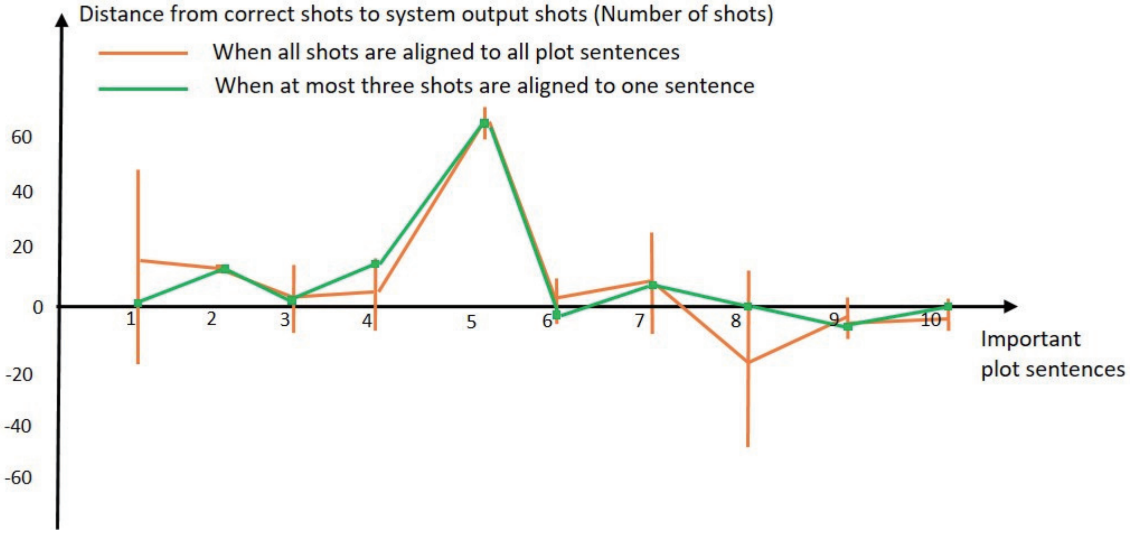


Figure 6. Distance from correct shots to system output shots (*Alice in the wonderland*)

Table IV
AUTOMATIC EVALUATION RESULTS: *Roman Holiday*

Important plot sentence	Number of correct shots selected manually	System output shots (No upper bound)	Evaluation	System output shots (With upper bound)	Evaluation
1	17	1∼33	Y	29,30,31	N
2	89	88	Y	88	Y
3	110	90∼101	N	90,91,92	N
4	164	159∼167	Y	160,161,162	N
5	195,198	201	N	201	N
6	401	325∼359	N	354,355,356	N
7	462	438	N	438	N
8	477	510	N	510	N
9	535,536,537	511∼536	Y	511,512,513	N
10	597,604	560∼611	Y	569,570,571	N
11	638	637	Y	637	Y
Total	—	—	54.5% (6/11)	—	18.2% (2/11)

Table V
AUTOMATIC EVALUATION RESULTS: *Alice in the wonderland*

Important plot sentence	Number of correct shots selected manually	System output shots (No upper bound)	Evaluation	System output shots (With upper bound)	Evaluation
1	21,22,23	3∼75	Y	24,25,26	Y
2	59,60,61	76	N	76	N
3	85,86,87	78∼103	Y	87,88,89	Y
4	127,128,129	122∼147	Y	144,145,146	N
5	149,150,151	209∼220	N	211,212,213	N
6	377,378,379	374∼388	Y	374,375,376	Y
7	617,618,619	612∼648	Y	628,629,630	N
8	1280,1281,1282	1232∼1296	Y	1280,1281,1282	Y
9	1568,1569,1570	1557∼1570	Y	1561,1562,1563	N
10	1578,1579,1580	1572∼1581	Y	1578,1579,1580	Y
Total	—	—	80.0% (8/10)	—	50.0% (5/10)

Table VI
SUBJECTIVE EVALUATION RESULTS: *Roman Holiday*

Important plot sentence	Number of correct shots selected manually	System output shots (No upper bound)	Subjective Evaluation	System output shots (W upper bound)	Subjective Evaluation
1	17	1∼33	Y	29,30,31	Y
2	89	88	Y	88	Y
3	110	90∼101	N	90,91,92	N
4	164	159∼167	Y	160,161,162	Y
5	195,198	201	Y	201	Y
6	401	325∼359	Y	354,355,356	Y
7	462	438	N	438	N
8	477	510	N	510	N
9	535,536,537	511∼536	Y	511,512,513	N
10	597,604	560∼611	Y	569,570,571	N
11	638	637	Y	637	Y
Total	—	—	72.7% (8/11)	—	54.5% (6/11)

Table VII
SUBJECTIVE EVALUATION RESULTS: *Alice in the wonderland*

Important plot sentence	Number of correct shots selected manually	System output shots (No upper bound)	Evaluation	System output shots (With upper bound)	Evaluation
1	21,22,23	3∼75	Y	24,25,26	Y
2	59,60,61	76	N	76	N
3	85,86,87	78∼103	Y	87,88,89	Y
4	127,128,129	122∼147	Y	144,145,146	N
5	149,150,151	209∼220	N	211,212,213	N
6	377,378,379	374∼388	Y	374,375,376	Y
7	617,618,619	612∼648	Y	628,629,630	Y
8	1280,1281,1282	1232∼1296	Y	1280,1281,1282	Y
9	1568,1569,1570	1557∼1570	Y	1561,1562,1563	Y
10	1578,1579,1580	1572∼1581	Y	1578,1579,1580	Y
Total	—	—	80.0% (8/10)	—	70.0% (7/10)

in Tapaswi et al. [4], we desire to pursue a more precise alignment between plots and shots by applying the method proposed in Sidiropoulos et al. [6].

REFERENCES

[1] C. Liang, Y. Zhang, J. Cheng, C. Xu, and H. Lu, “A novel role-based movie scene segmentation method,” in *Advances in Multimedia Information Processing — PCM2009*, ser. LNCS. Springer, 2009, vol. 5879, pp. 917–922.

[2] H. Yi, D. Rajan, and L.-T. Chia, “Semantic video indexing and summarization using subtitles,” in *Advances in Multimedia Information Processing — PCM2004*, ser. LNCS. Springer, 2004, vol. 3331, pp. 634–641.

[3] T. Tsoneva, M. Barbieri, and H. Weda, “Automated summarization of narrative video on a semantic level,” in *Proc. Semantic Computing*, 2007, pp. 169–176.

[4] M. Tapaswi, M. Bäuml, and R. Stiefelhagen, “Aligning plot synopses to videos for story-based retrieval,” *International Journal of Multimedia Information Retrieval*, vol. 4, no. 1, pp. 3–16, 2015.

[5] E. Apostolidis and V. Mezaris, “Fast shot segmentation combining global and local visual descriptors1,” in *Proc. ICASSP*, 2014, pp. 6583–6587.

[6] P. Sidiropoulos, V. Mezaris, I. Kompatsiaris, H. Meinedo, M. Bugalho, and I. Trancoso, “Temporal video segmentation to scenes using high-level audiovisual features,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 8, pp. 1163–1177, 2011.