An Interactive Cooking Support System for Short Recipe Videos based on User Browsing Behavior

Takuya Yonezawa
Kwansei Gakuin University
etm25735@kwansei.ac.jp

Yuanyuan Wang
Yamaguchi University
y.wang@yamaguchi-u.ac.jp

Yukiko Kawai
Kyoto Sangyo University
Osaka University
kawai@cc.kyoto-su.ac.jp

Kazutoshi Sumiya
Kwansei Gakuin University
sumiya@kwansei.ac.jp

***Abstract*—Recently, short recipe videos such as Kurashiru and DELISH KITCHEN have become popular. These short recipe videos can help people learn many cooking skills in a brief time. However, it is difficult for users to understand all cooking operations by viewing these videos only once. These short recipe videos do not consider users' cooking skills (cooking levels) since anyone may view the same video. Therefore, in this work, we propose an interactive cooking support system for short recipe videos by extracting and weighting cooking operations for each cooking genre based on user browsing behavior. The system then recommends various supplementary recipe videos based on the weights of cooking operations and user browsing behavior. Also, the system provides a user interface, called *Dynamic Video Tag Cloud* for visualizing the supplementary recipe videos, and the supplementary recipe videos can be dynamically changed based on the user browsing behavior. As a result, users can intuitively and easily understand cooking operations suited to their cooking favorites. Finally, we verified the effectiveness of the weighting of cooking operations and discussed the usefulness of our proposed user interface using the SUS score.**

***Index Terms*—cooking videos, cooking recipes, supplementary information, user browsing behavior.**

I. Introduction

Nowadays, an increasing number of people are staying at home more often, the use of cooking recipe sites such as Cookpad[1] and Rakuten Recipe[2] has become more widespread. The advantages of these sites are convenient and accessible to many users because they can browse a large number of recipes for free. On the other hand, with the spread of mobile devices, recipes using not only text recipes but also videos such as cooking TV programs and short recipe videos are becoming popular. Short recipe videos such as Kurashiru and DELISH KITCHEN, mainly focus on dishes that are easier to cook than the ones featured in cooking TV programs. These short recipe videos do not have audio commentaries such as those added to cooking TV programs, only cooking operations, ingredients, or simple explanations are presented in open captions. And short recipe videos taken by the fixed-point camera are edited to be about 1 minute long. Therefore, viewers can get a quick overview of the cooking points of short recipe videos. In this work, we focused on short recipe videos with open captions that viewers can easily grasp the main points of cooking in a short time.

[1]https://cookpad.com/
[2]https://recipe.rakuten.co.jp/

However, it is difficult to cook while browsing only the short recipe video. Because the video is playing too fast, and some necessary cooking operations are presented as images. As extracting necessary cooking operations, Akiguchi et al. [1] could detect complex cooking operations based on temporal information of cooking TV programs, but it is not appropriate for short recipe videos. Also, existing cooking TV programs and short recipe videos do not consider differences in users' cooking levels. In particular, short recipe videos have time constraints.

In this work, we aim to extract and weight cooking operations from short recipe videos for each cooking genre to support browsing short recipe videos based on user browsing behavior. It is useful to recommend supplemental information and relevant information to the short recipe videos according to the cooking levels of users based on the weights of cooking operations and user browsing behavior. For this, we first extract cooking operations from text recipes included in short recipe videos, and we weight the cooking operations for each cooking genre based on the *TF-iDF* method and user browsing behavior (pause, rewind, skip). Next, we determine which cooking operations need to be supplemented based on their weights, and extract and visualize their supplementary information and relevant information to users. Here, supplemental information and relevant information are recipe videos. In this way, users can intuitively and easily understand cooking operations suited to their cooking levels.

The remainder of this paper is structured as follows. The next section describes an overview of our proposed system and reviews studies on cooking recipe recommendations and studies analyzing cooking videos. Section III explains how to weight cooking operations. Section IV describes how to extract and visualize the supplementary information. Section V shows our evaluation experiments of the proposed method using the real dataset. Finally, Section VI concludes this paper and presents future works.

II. System Overview and Related Work

A. An Interactive Cooking Support System

In this work, we propose an interactive cooking support system for short recipe videos by extracting and weighting cooking operations for each cooking genre based on user browsing behavior. To determine the user's cooking level,

978-1-6654-1865-2/21/$31.00 ©2021 IEEE
DOI 10.1109/MIPR51284.2021.00011

Fig. 1. Comparing our proposed system to other methods.

we first ask the user to select the cooking experience as an experienced person or beginner. To extract the user's preference, the user needs to choose a genre of favorite dishes. Then, supplementary recipe videos based on the user's cooking level are presented as shown in Fig. 1.

Our proposed user interface is called *Dynamic Video Tag Cloud*, it presents an existing short recipe video on the left screen, and users can experience browsing behavior such as "pause", "rewind" or "skip" when browsing a short recipe video. Supplementary recipe videos are shown as "video tag cloud" on the right screen, and users can select any supplemental recipe video to browse from the video tags. The "video tag cloud" shows thumbnail images of supplementary recipe videos for important cooking operations, and sizes of thumbnail images according to the weights of cooking operations, such as "chop", "diagonal cut" or "put oil". And the weights of the cooking operations are decreased outward from one of the cooking operations, which is located at the center of the right screen. For example, if "chop" is determined as the most important cooking operation for the user in the short recipe video "curry" (the left screen), the thumbnail image of the supplementary recipe video "chopped carrot" is shown at the center of the right screen. Then, this system can help users to understand the cooking operations intuitively. The flow of our proposed interactive cooking support system is shown in Fig. 2.

A: Main Screen asking users to select the cooking experience and a genre for favorite dishes.

B: Menu Select asking users to select a short recipe video.

C: Video Tag Cloud showing the selected short recipe video with their supplementary recipe videos in the video tag cloud.

D: Supplementary Video showing the selected supplementary recipe video.

Fig. 3 shows the flow chart of our proposed method. Firstly, a user selects a short recipe video that the user wants to browse. At this moment, a cooking genre of the short recipe video is recorded in "Browsing DB" to analyze the user's favorite genre, since we are weighting cooking operations based on cooking genres. Secondly, the weighting methods for cooking operations are described in Section III. One is the weighting based on *TF-iDF* and the other one is the weighting based on user browsing behavior. Basically, *TF-iDF* is simple but it is superior to Word2Vec for the cost of the dynamic calculation. In the case of using Word2Vec, it is sufficient to modify *TF-iDF* to calculate the weights. Furthermore, each extracted cooking operation is generated as a query, and the supplementary video of the corresponding cooking operation is also extracted from the Web. Also, the thumbnail images to be presented on the generateVideo Tag Cloud is generated at this time. After that, the most important cooking operation is selected from the cooking operations above a threshold, and the corresponding video is located in the Video Tag Cloud.

For a moving image in the first place, the moving image should be the largest size and located at the center. Afterward, the other cooking operations should be arranged around the most important cooking operation, while gradually decreasing the size of the moving image in decreasing order of weights. Finally, the user's browsing behavior is performing while the user is browsing the short cooking video. The Video Tag Cloud to the user without any change if there is no browsing behavior. However, if the user's browsing behavior is pausing or rewinding the video when browsing the short cooking video, we assume that the user has no knowledge or skill in the corresponding cooking operation, we then calculate the weight for the cooking operation dynamically. Therefore, the cooking assistance that the user does not understand is displayed preferentially, thereby effectively supporting cooking. In this work, we defined the browsing behaviors handled by the system as follows: *Pause*, *Rewind*, and *Skip*. Moreover, the history of those browsing behaviors is recorded in the "User browsing behavior history DB".

B. Related work

In recent years, many studies have actively conducted on the search and recommendation of cooking recipes. In par-

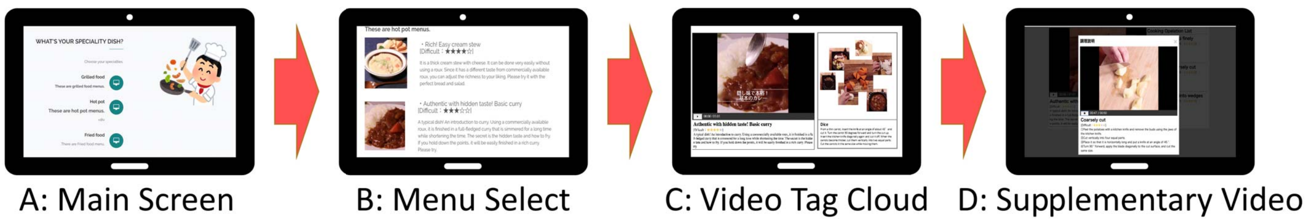

Fig. 2. The flow of our proposed interactive cooking support system for short recipe videos.

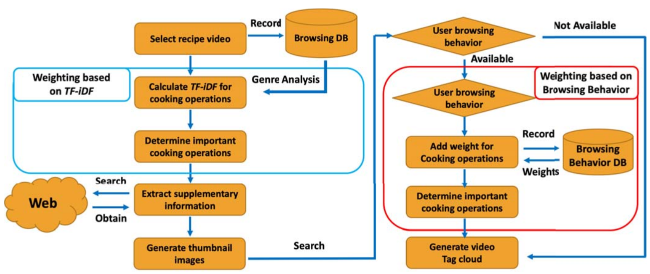

Fig. 3. System flow chart.

ticular, many studies seek the cooking difficulty of cooking recipes. Fujisaki and Miyoshi [2] are conducting research to recommend cooking recipes of appropriate difficulty to users who want to learn cooking skills. This research classifies the difficulty level of recipes into three categories: handiness, technical difficulty, and intelligibility of recipe sentences, and applies an algorithm for estimating the difficulty level of learning content to the estimation of each difficulty level. Makino et al. [3] focused on cooking behaviors that appeared in textbooks, and they determined the difficulty of the cooking operation described in the text. Iwamoto and Miyamori [4] pay attention to the cooking skills necessary for cooking, and propose a search system that can search by the difficulty level of each recipe. The system they proposed allows novice users to efficiently understand which recipes they can cook. Moreover, Kusu et al. [5] proposed a computation method of degrees of difficulty for searching cooking recipes based on the given set of cooking operations. These studies calculated the degrees of cooking difficulty for cooking recipe text. But in this work, we calculate the degrees of difficulty for the cooking operations in short recipe videos. Yajima and Kobayashi [6] proposed a system to recommend recipes that are considered “easy” in the individual’s situation. Our work is similar in that it deals with simple dishes. However, it differs in that it recommends information about the cooking behavior of the recipe.

Recently, in addition to textual recipe sites, there has been an increasing number of posts on recipe video sites, and studies on analyzing cooking videos as cooking recipe text are actively underway. Doman et al. [7] proposed a method to classify cooking operations by defining cooking programs as cooking videos and analyzing the repeated cooking operations in the cooking video. But in this study, we extract supplementary information from the cooking operations in short recipe videos. Otaki and Takano [8] proposed a method to calculate the cooking costs, focusing on the three points of cooking operation, utensils, and ingredients included in the cooking recipe video, and considering the number of times each appears and the number of times the dish moved. The goal of our work is to calculate the importance of the frequency of cooking operations included in the short recipe videos. Ushiku et al. [9] proposed a method to generate recipes from videos by training a Faster R-CNN network for object recognition and an LSTM network for text generation, and combining them at runtime. In this research, the difference is that auxiliary recipe information is extracted from the recipe short video.

Several multimedia research also actively conducted cooking recipes. Maruyama et al. [10] proposed a cooking recipe recommendation system using image recognition on mobile devices. Our research also makes information recommendations in real time, but the difference is that the decision is made based on the viewing operation when the user views the short recipe video. Tanno et al. [11] proposed a new approach for food calorie estimation with CNN and Augmented Reality (AR)-based actual size estimation. As a result, it is possible to calculate the size more accurately than in the previous method by measuring the meal area directly, and the calorie estimation accuracy has improved. This research is about calorie calculation using AR, but it is relevant as research using recipes and multimedia as in this paper.

Furthermore, our previous work developed a system that determines the meanings of cooking operations in the text recipes in the past and presents the supplementary information of the recipe to the user in an appropriate way [12]. The system proposed here also shows the existing short recipe videos on the left screen while presenting supplementary recipe videos of necessary cooking operations on the right screen. The user interface proposed here was designed based on YouTube, an existing video service, as shown in the right image of Fig. 1. We also made some improvements to the proposed system in our previous work [13]. The difference between this system and our proposed system in [12] is to support users understand the cooking operations sensitively by extracting supplemental recipe information. In [13], our proposed system provides a new interface, called *Dynamic Video Tag Cloud* to present the thumbnail images of the supplemental recipe videos on the right screen. The sizes of the thumbnail images can be changed according to the degrees of difficulty of the cooking operations by our proposed difficulty calculation method. It enables the users sensitively to understand complex cooking operations in the short recipe video.

In this paper, we improve our developed *Dynamic Video Tag Cloud* in [13], as shown in the left image of Fig. 1, to dynamically changes the supplementary information by extracting the users' intentions based on user browsing behavior. To verify the effectiveness and the usefulness of our developed *Dynamic Video Tag Cloud*, we compare it with the existing service (Kurashiru) and our previous work (YouTube-style) in [12], as shown in Fig. 1.

III. Weights of Cooking Operations

In this section, we describe how to extract cooking operations for each cooking genre from short recipe videos, and explain how to weight cooking operations based on user browsing behavior.

A. Extracting Cooking Operations

To extract cooking operations from short recipe videos, we randomly selected 100 short recipe videos from Kurashiru, a video recipe site that handles short recipe videos. Then, we obtain words that represent cooking operations from the text recipes included in these short recipe videos by a Japanese morphological analysis using MeCab. However, there are various notations of the same cooking operations written in hiragana letters and Japanese Kanji characters. Therefore, to unify different written forms of the same cooking operation based on a preliminary survey of text recipes of 100 short recipe videos, we created a total of 53 rules to normalize the cooking operations.

B. Weighting Cooking Operations based on TF-iDF

By considering the cooking levels and favorite cooking genres of users, cooking methods are classified into genres to calculate the weights of cooking operations. In this paper, we classified cooking recipes into five types of genres based on genres in the recipe site: "fried food", "stir-fried food", "baked food", "rice bowl", and "boiled food". Then, we calculate weights for words of cooking operations by considering genres. For example, there are various complicated cooking operations in a stir-fried cooking recipe, such as chop, mince, and diagonal cut for ingredients such as vegetables. Since cooking operations with different tasks are often used in cooking recipes of the same genre, we calculate weights for cooking operations in each short recipe video based on the *TF-iDF* using the *TF* value and the *iDF* value for a word j of a cooking operation in a text recipe $r \in k$ of each short recipe video belonging to each genre k as follows:

$$\textit{TF(j, r)} = \frac{\#j \text{ in recipe } r}{\text{\#all cooking operations of genre } k} \quad (1)$$

$$\textit{iDF(j)} = \log \frac{\text{total \#recipes in genre } k}{\text{\#recipes contains } j+1} \quad (2)$$

Here, to avoid *iDF* value is divided by 0 when *iDF* is calculated for a word of a cooking operation that does not appear in a short recipe video once, the denominator is incremented by +1.

C. Weighting Cooking Operations based on User Browsing Behavior

To consider user browsing behavior for weighting cooking operations, we conducted a preliminary experiment for surveying the types and intentions of user browsing behavior while browsing short recipe videos. As a result, we found mainly three types of browsing behavior that are "pause", "rewind", and "skip". In particular, we found that most of the users who did "rewind" when they need to confirm cooking operations be careful. Also, many users did "pause" when they do not understand cooking operations. On the other hand, "skip" indicated that users do not want to browse sufficiently understood cooking operations. Therefore, we set weights for these three types as rewind (+1.0), pause (+0.8), and skip (-1.0). Then, we calculate weights for cooking operations in each short recipe video based on a sum of *TF-iDF* values by Eqs. (1)(2) and weights for three types. For example, if a user did "rewind" (+1.0) for a cooking operation "slice" and the *TF-iDF* value for "slice" is 0.222, then the weight for "slice" is 1.222 $(= +1.0 + 0.222)$.

IV. Visualization of Supplementary Information

In this section, we explain how to extract supplementary information of important cooking operations and generate the *Dynamic Video Tag Cloud*.

A. Supplementary Information Extraction

For generating the video tag cloud, cooking operations are determined based on ingredients and seasonings, which are to be supplemented by the difficulty level of the "cooking operation" in each short recipe video. In particular, we set a threshold value as the average value of the weights of cooking operations included in each recipe calculated by each of the methods proposed in Section III. On the other hand, cooking operations need to supplement when they have higher weights than the threshold, and we extract them as supplementary targets.

Then, we can select the top cooking operation of the ranking as a query from the determined supplement targets In this way, we acquire the top-level cooking video of search results from the Web by the selected cooking operation, and we generate the video tag cloud using the supplementary recipe video of the weighted supplement target. In the generated video tag cloud, we set the size of the auxiliary recipe video is according to the weight of the cooking operation that is the supplement target.

B. Dynamic Video Tag Cloud

Our proposed method extracts the auxiliary recipe videos from the Web and presents them in the video tag cloud on the right screen of the user interface. In this interface, the thumbnails of the supplementary recipe videos for each

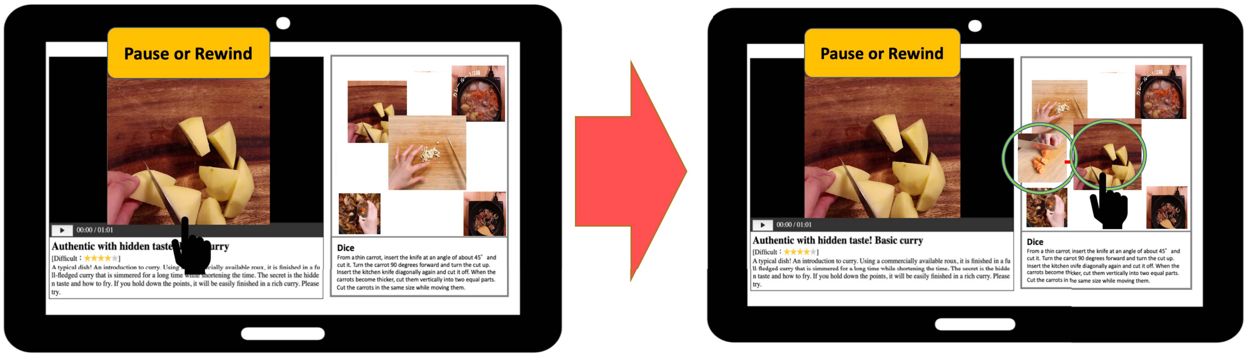

Fig. 4. Screen transition.

supplement target are located at the center of the video tag cloud while changing in size according to the weight of the supplement target. Also, the sizes of the thumbnails of the supplementary targets are decreasing from the center to the outside according to the weights of the cooking operations. It allows the user to select the auxiliary recipe video of the cooking operation that they want to check.

Therefore, in the Video Tag Cloud, when the user selects the supplementary recipe video presented on the right screen, some are the same cooking operation with other ingredients included in the recipe. We propose a system that recommends them as related information at the same time. During the extraction of the supplementary recipe videos from the Web using a query, the videos related to the corresponding cooking operations are extracted and generated as the related information. Also, we extract the supplementary recipe videos by the comprehension levels of the cooking operations based on user browsing behavior, and the thumbnails of the cooking operations are changed dynamically in the Video Tag Cloud. In the case of “pause” or “rewind”, it is determined that understanding the corresponding cooking operation is insufficient and is given a further weighting. If the user browsing behavior is “rewind”, the weight is increased by +1.0, and if the user browsing behavior is “pause”, the weight is increased by +0.8, based on the results of a preliminary experiment. Based on the above, the thumbnail sizes of video tags can change in the Video Tag Cloud. On the other hand, if the user browsing behavior is “skip”, it is determined that the cooking operation is not necessary for the user, and the weight is reduced by -1.0 to remove it from the Video Tag Cloud.

V. Evaluation

In this work, we conducted two experiments. In Experiment 1, we verified our supplementary target extraction method using 100 recipes. And in Experiment 2, we evaluated the usefulness of our proposed Dynamic Video Tag Cloud with the System Usability Scale (SUS) score [14].

A. Experiment 1: Verification of Supplementary Target Extraction

To verify the supplementary target extraction, we extracted cooking operations that need supplementing according to their weights and the given threshold. Table I shows the extracted cooking operations from four short recipe videos (R1∼R4) by our proposed method. From the results, we found cooking operations that no need to supplement, such as “Serve” and “Heat”, are extracted in common with the recipes in Table I. Also, we could extract cooking operations, such as “Dice” and “Shared”, from the recipe (R1) that need supplementary information.

Afterward, we considered that easy cooking operations appear more frequently in the recipes, and difficult cooking operations appear less in the recipes. Cooking operations that don’t need supplementary information such as “Put” and “Drain (Oil)”, which are also extracted from the recipes on the top with high weights, we should set the appropriate threshold in the future.

Next, we compared two genres of grilled foods such as the recipe (R3) and stewed foods such as the recipes (R2) and (R4), “Slice” with high importance on the top of grilled foods (R3), but it with low weight in stewed foods (R2) and (R4). Because of these differences, users who are good at stewed foods may not have much experience with “Slice” compared to users who are good at grilled foods, so it is necessary to supplement the information about “Slice” appears in stewed foods.

B. Experiment 2: Usefulness of Dynamic Video Tag Cloud

To evaluate the usefulness of our proposed system called *Dynamic Video Tag Cloud*, we applied the SUS score in this experiment. We compared our proposed *Dynamic Video Tag Cloud* (A) in this paper with two other user interfaces: one is the existing short recipe video service (Kurashiru) (B), and the other one is a YouTube-style user interface (C) in our previous work [12] (see Fig. 1). There were ten subjects who participated in this experiment, six subjects are beginners in cooking, and four subjects are experienced in cooking.

The question items of the SUS score are as follows:

- Q1 I think that I would like to use this user interface frequently.
- Q2 I found the user interface unnecessarily complex.
- Q3 I thought the user interface was easy to use.

TABLE I
RESULTS OF COOKING OPERATION EXTRACTION

R1: Cabbage and chicken with salted rice		
Rank	Operation	Weight
1	**Mix**	**0.254**
2	**Dice**	**0.152**
3	**Shared**	**0.152**
4	**Stop**	**0.142**
5	Drain (Oil)	0.121
6	Stir fry	0.108
7	Cut	0.108
8	Serve	0.096
9	Put	0.089
Threshold		**0.132**

R2: Tomato stew		
Rank	Operation	Weight
1	**Put**	**0.227**
2	**Stir fry**	**0.224**
3	**Add**	**0.140**
4	**Cut**	**0.132**
5	Scatter	0.101
6	Wrap	0.101
7	Heat in a microwave	0.101
8	Remove	0.085
9	Slice	0.078
10	Mince	0.078
11	Serve	0.075
Threshold		**0.122**

R3: Mini mayonnaise and corn pizza		
Rank	Operation	Weight
1	**Slice**	**0.222**
2	**Put**	**0.189**
3	**Lay out**	**0.178**
4	**Scatter**	**0.178**
5	Cut	0.150
6	Put on	0.136
7	Serve	0.132
8	Mix	0.128
9	Put	0.114
Threshold		**0.158**

R4: Minestrone		
Rank	Operation	Weight
1	**Put**	**0.236**
2	**Stir fry**	**0.207**
3	**Dice**	**0.177**
4	Season	0.140
5	Round slice	0.140
6	Put oil	0.123
7	Heat	0.112
8	Slice	0.108
9	Serve	0.104
10	Boil	0.091
Threshold		**0.151**

TABLE II
COMPARISON OF THREE USER INTERFACES USING THE SUS SCORE

Method	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score	Grade	Adjective Rating
A (*Dynamic Video Tag Cloud*)	4.0	**2.7**	3.7	2.1	3.4	2.3	3.1	**2.4**	3.9	**3.0**	64.0	D	Poor
B (Kurashiru)	3.5	**2.7**	3.7	**2.2**	2.5	2.2	**4.0**	2.0	**4.1**	2.9	64.5	D	Poor
C (Youtube-style)	**4.1**	2.2	**4.4**	1.8	**4.0**	**2.4**	3.5	1.7	4.3	2.3	**74.8**	B	Good

Q4 I think that I would need the support of a technical person to be able to use this user interface.
Q5 I found the various functions in this user interface were well integrated.
Q6 I thought there was too much inconsistency in this user interface.
Q7 I would imagine that most people would learn to use this user interface very quickly.
Q8 I found the user interface very cumbersome to use.
Q9 I felt very confident using the user interface.
Q10 I needed to learn a lot of things before I could get going with this user interface.

The subjects asked questions with a 5-point Likert scale. The adjective ratings for the SUS Score are the following: A) Excellent: >80.3, B) Good: >68.0-80.3, C) Okay: >68.0, D) Poor: >51.0-68.0, E) Awful: >51.0.

Table II shows the average scores of each interface. (A) is our proposed *Dynamic Video Tag Cloud* in this paper, (B) is the existing short recipe video service (Kurashiru), and (C) is a YouTube-style user interface in our previous work. As a result, the average score of (A) is 64.0, (B) is 64.5, and (C) is 74.8, (C) got a good rating that was the most useful user interface for the subjects. The results show that the proposed *Dynamic Video Tag Cloud* (A) is in the lowest rank. The reason for these results is that it takes time to get used to the proposed system because of its unique operability, as shown by the scores in Q8 and Q10. In comparing (A) and (C), we found that (A) is superior in terms of functionality, it is possible to supplement the users. Besides that, we could measure the time until to understand the cooking operations. From the results, we found that eight out of ten subjects took less time to understand (A) than it did in (B).

We considered that these factors are due to the difference in the flow of information retrieval that is necessary for the user to understand the cooking operations. In the existing service, users need to understand the cooking operations themselves after viewing the whole video. For this reason, it takes a long time to understand all cooking operations in the short recipe video. On the other hand, in the proposed method, if the user pauses while viewing the video, the supplementary information can be recommended to the user. Therefore, we considered that the time it takes for the user to understand the cooking operations to be shorter than that of the existing service.

Based on the above, we confirmed that our proposed *Dynamic Video Tag Cloud* could efficiently and effectively recommend supplementary information to the users.

VI. CONCLUSION

In this paper, we proposed a method to compute the weights of cooking operations in a short recipe video and set a threshold to extract cooking operations that need to supplement. Moreover, we also developed a user interface, called *Dynamic Video Tag Cloud*, which provides the users with supplemental recipe videos from the extracted information. As a result of comparing the user interfaces of our proposed methods with the existing service in the evaluation, our developed *Dynamic Video Tag Cloud* could shorten the time until the user understands the cooking operations.

In the future, we plan to do two main tasks. The first one is to set detailed thresholds to improve the accuracy of the cooking operation extraction. We consider that it will be possible to increase the range of cooking levels by the users. The second one is to improve the user interface for the users to understand the cooking operations easily with verification of the usefulness of the supplementary targets of cooking operations.

ACKNOWLEDGMENT

The research was partially supported by ROIS NII Open Collaborative Research 2020-(20FC01).

REFERENCES

[1] I. Akiguchi, Y. Wang, Y. Kawai, and K. Sumiya, “Determining difficulty of cooking recipe videos based on time characteristics,” in *Proceedings of the 10th Forum on Data Engineering and Information Management (DEIM Forum 2018)*. Tsukuba, Japan: DBSJ, 2018, pp. C4–4.

[2] Y. Miyoshi, Y. Fujisaki, K. Suzuki, K.-i. Shiota, and R. Okamoto, “Estimating the difficulty of cooking recipes by analyzing user–recipe relationship in the social network,” in *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Association for the Advancement of Computing in Education (AACE), 2015, pp. 420–425.

[3] N. Makino, K. Kusunoki, and K. Hatano, “A difficulty calculation method for cooking recipe search based on cooking behaviors,” in *Proc. of the 9th Forum on Data Engineering and Information Management (DEIM Forum 2017)*. Tsukuba, Japan: DBSJ, 2017, pp. C4–5.

[4] J. Iwamoto and H. Miyamori, “Proposal of a search system for cooking recipes considering difficulty of cooking,” in *Proc. of the 1st Forum on Data Engineering and Information Management (DEIM Forum 2012)*. Hyogo, Japan: DBSJ, 2009, pp. E1–3.

[5] K. Kusu, H.-I. Choi, T. Kambara, T. Kinoshita, T. Shioi, and K. Hatano, “Searching cooking recipes by focusing on common ingredients,” in *Proceedings of the 19th International Conference on Information Integration and Web-Based Applications & Services*, ser. iiWAS ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 95–101. [Online]. Available: https://doi.org/10.1145/3151759.3151797

[6] A. Yajima and I. Kobayashi, ““easy” cooking recipe recommendation considering user’s conditions,” in *Proc. of the 1st Forum on Data Engineering and Information Management (DEIM Forum 2009)*. Shizuoka, Japan: DBSJ, 2009, pp. E5–6.

[7] K. Doman, C. Y. Kuai, T. Takahashi, I. Ide, and H. Murase, “Video cooking: Towards the synthesis of multimedia cooking recipes,” in *Advances in Multimedia Modeling*, K.-T. Lee, W.-H. Tsai, H.-Y. M. Liao, T. Chen, J.-W. Hsieh, and C.-C. Tseng, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 135–145.

[8] K. Otaki and K. Takano, “A method for estimating cooking cost from cooking vidoe based on cooking motions, tools, and food materials,” in *Proceedings of the 79th National Convention of IPSJ*. Tokyo, Japan: IPSJ, 2017, pp. 539–540.

[9] A. Ushiku, H. Hashimoto, A. Hashimoto, and S. Mori, “Procedural text generation from an execution video,” in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP 2017)*, 2017, pp. 326–335.

[10] T. Maruyama, M. Akiyama, and K. Yanai, “Recipe recommendation system using ingredient image recognition,” in *IEICE Technical Report*, March 2012, pp. 43–48.

[11] R. Tanno, T. Ege, and K. Yanai, “Ar deepcaloriecam v2: Food calorie estimation with cnn and ar-based actual size estimation,” in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST ’18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3281505.3281580

[12] T. Yonezawa, Y. Wang, Y. Kawai, and K. Sumiya, “A cooking support system by extracting difficult scenes for cooking operations from recipe short videos,” in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2225–2227. [Online]. Available: https://doi.org/10.1145/3343031.3350608

[13] ——, “Dynamic video tag cloud: A cooking support system for recipe short videos,” in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, ser. IUI ’20, 2020, pp. 122–123.

[14] J. Sauro, “Measuring usability with the system usability scale (sus),” 2011.