Cooking Navi: Assistant for Daily Cooking in Kitchen

Reiko HAMADA
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo, 113-8656, Japan
reiko@mtl.t.u-tokyo.ac.jp

Jun OKABE
Institute of Info. Technology
Nihonbashi, Chuo-ku, Tokyo,
103-0024, Japan
jokabe@iit.jp

Ichiro IDE
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya, 464-8603, Japan
ide@is.nagoya-u.ac.jp

Shin'ichi SATOH
Nat'l Institute of Informatics

Shuichi SAKAI
The University of Tokyo

Hidehiko TANAKA
Institute of Info. Security

ABSTRACT

We are developing a cooking navigation system, which helps even a novice user to cook several recipes in parallel without failure, while improving an advanced user's skill further. To realize this, the system optimizes the cooking procedure considering the following restrictions: (1) Duration of cooking, (2) Accuracy of cooking, and (3) Learning effect, by providing appropriate instructions to user's at the right timing, making full use of multimedia information. The users should be able to cook perfectly and comfortably just by following the text, video and audio provided by the system. According to the result of a preliminary experiment, all users from novice to experienced cooks could finish two dishes in parallel while enjoyeing the cooking very much. The result of a questionnaire shows the effectiveness of the multimedia navigation that we propose.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; H.5.2 [**Information Interfaces and Presentation**]: User Interface; J.m [**Computer Applications**]: Miscellaneous; K.3.1 [**Computers and Education**]: Computer Uses in Education

General Terms

Verification, Human Factors

Keywords

Assistance to human activities, Cooking

1. INTRODUCTION

We are developing a cooking navigation system, which helps an inexperienced user, a novice, to cook without failure, and at the same time improves an advanced user's skills. This system enables even a novice to execute complicated tasks such as cooking several recipes in parallel. On the other hand, the system is designed so that the users should also enjoy cooking. In order to realize such a system, we treat cooking tasks in each cooking recipe as a flow structure.

Each cooking task consists of Action Units (AUs) such as "cut onions", "fry onions", and so on. In addition, AUs have a logical order; for example, "fry eggs" should usually come after "break eggs", and also have a time restriction; for instance, "fry onions for 5 minutes." Moreover, AUs have to be merged correctly to complete a dish; for example, "fried bacon and onions are mixed together in a frying pan." The flow structure describes such ordinal restrictions and relations between AUs. In this system, we deal with actual cooking activities as an optimization problem that allocates the above AUs appropriately under given restrictions on the resources, such as the user's cooking skills and the kitchen environment.

The system optimizes the cooking procedure considering the following restrictions: (1) Duration of cooking, (2) Accuracy of cooking, and (3) Learning effect. The system optimizes all these constraints to provide appropriate instructions to the users at the right timing by full use of multimedia information. The user should be able to cook perfectly and comfortably just only by following the text, video and audio provided by the system. This system should be a suitable application for ubiquitous kitchens such as those proposed in [1, 5].

2. COOKING AS AN OPTIMIZATION PROBLEM

Fig. 1 shows an example of the flow structure which is derived from a recipe. In this structure, ingredients are processed at each AU and merged so as to compose a dish at the end. From this structure, restricted and un-restricted orders could be distinguished clearly. Directly linked orders can not be reversed, while un-restricted AUs can be processed in parallel. However, most of the AUs have to be allocated sequentially for a user because he/she can only execute one operation at a time.

In many cases, two or more recipes have to be cooked at the same time. But as shown in Fig. 2 (a), if we cook them sequentially, the first dish would become cold when the last one is finished. Therefore, as shown in Fig. 2 (b), we have to

MM'05, November 6–12, 2005, Singapore.

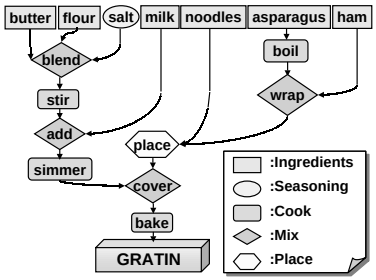

Figure 1: A flow structure representing a recipe.

merge multiple recipes and rearrange the sequence of AUs to finish all dishes at the same timing. (Usually this is difficult especially for novice cook.) Some AUs can be executed in parallel even if there is only one person to cook, and as a result the duration of the cooking would be reduced. For example, while we are waiting for the rice to absorb water, another task can be executed in parallel.

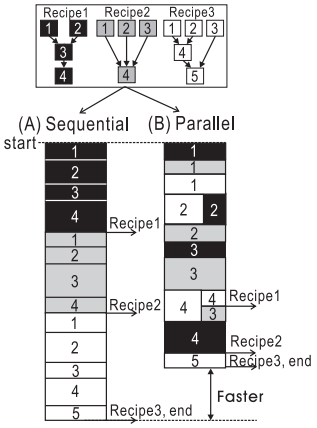

Figure 2: Rescheduling for cooking multiple recipes.

The resources involved in the cooking optimization problem are human resource and kitchen resource. The human resource is the number of cooks. The more cooks there are, the higher the parallelism can be. On the other hand, as kitchen resources, we introduce "gas stoves", "cutting boards", "sinks" and "tables", each of them correspond to heating tasks, cutting tasks, tasks related to water, and other tasks respectively. On optimization of the cooking problem, allocation of AUs to these resources is a strong restriction.

In addition, to realize the three aims of optimization mentioned in the Section 1, the system considers the following conditions.

Duration

This is optimized by considering the total cooking duration and the timing of serving. Then the system tries to optimize flow structures satisfying the following conditions: (1) Minimize the duration as much as possible, and (2) Finish all dishes in almost the same timing.

Accuracy

To optimize the cooking accuracy, users have to be navigated properly. In many cases the reason of a failure comes from the lack of information; mere text information is not enough to fully describe cooking motions or conditions of ingredients. The system makes use of every medium in the cooking navigation. First, visual information from cooking video instructs what he/she has to do in the actual environment. Next, audio can provide information even when he/she is occupied with a cooking task, too busy to watch the monitor. At last, text information shows detailed information that the video or audio can not express such as the name and amount of the ingredients.

Each AU which is defined as appropriate unit of action is displayed for a user one by one for a certain navigation. Such multimedia step-by-step navigation will help users to fully understand each process.

Learning Effect

If an appropriate navigation is provided depending on the user's behavior and progress, the user can learn cooking skills naturally by watching a cooking video that always instructs exactly what he/her has to do at the moment, and then high learning effect can be expected.

To realize this, scheduling of the process is performed in real-time to follow the changes of the situation. By this, the actual duration of each process can be adjusted according to the user's speed. Even if there is some interruption during the process or the user is much slower than expected, the system waits for the user by extending the duration of the AU. Furthermore, the user can always select the next cooking motion against the recommendation by the system. In such cases, the system would re-perform the optimization to deal with the changed situation.

3. IMPLEMENTATION

Details to realize the features mentioned in the previous Section is introduced in this Section.

3.1 Overview of the System

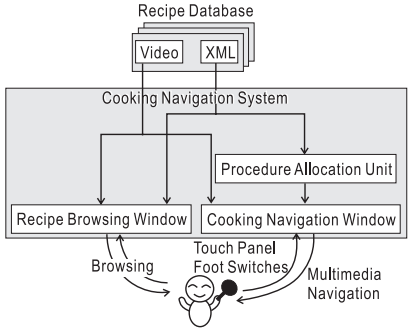

Figure 3: Overview of the "Cooking Navi" system

Fig. 3 shows the overview of the system.

1. Input data (Recipe Database)
   This system requires as input, video file(s) and an XML file with recipe texts and metadata. They have to be prepared beforehand, so the metadata is created either manually or automatically.
2. Procedure Allocation Unit
   Optimizes and reschedules cooking procedures referring to the flow structure.

3. Recipe Browsing Window (Fig. 4)
   This is a window to browse and select multimedia recipes by displaying corresponding video and recipe texts.
4. Cooking Navigation Window (Fig. 5)
   While users cook, they are provided with multimedia navigations from this window.

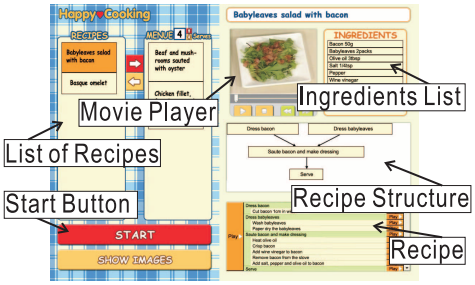

Figure 4: Recipe Browsing Window

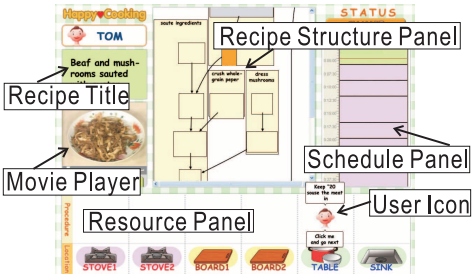

Figure 5: Cooking Navigation Window

3.2 Data Structure

3.2.1 *User Environment Data*

Users are requested to input their kitchen environment; the number of stoves, cutting boards, tables, sinks, and their layout.

3.2.2 *Recipe Data*

– Action Unit (AU)

"Action Unit" is defined as a unit of cooking action. The role of an "Action Unit" is similar to a "verb" in a text. It is often more coarse so it might correspond to two or more verbs. Each AU corresponds to a "Shot" in a video. Procedure optimization is performed by this unit.

The AU structure is displayed in the center of the Cooking Navigation Window (Fig. 5). The node that the user is executing is highlighted (colored orange) and nodes that have been finished turns to gray to indicate the status to the users. At the bottom of the Cooking Navigation Window is shown the cooking navigation field. The kitchen resources are laid out according to the user's environment and a character that represents the user is located where the AU is allocated. A speech balloon displays the description of the AU to instruct the user. While the user is executing an AU, the corresponding video shot is played repeatedly. When the user finishes the action, he/she clicks the character icon and then the next AU will appear. If there is another AU that the user is executing in parallel, the remaining time is displayed, which works as a kitchen timer.

In the input data, an "Action Unit" contains the following information.

- Text description.
- Dependency to the next AU.
- Corresponding video "Shot".
- The resource which the motion should be allocated.
- Default duration time for the AU.
- Flag that indicates whether the AU can be processed in parallel or not.

– Action Block

An "Action Block" is a series of AUs which handle the same ingredient at the same resource. The corresponding unit in the video is defined as a "Scene", which is a group of shots. The Action Block structure is displayed in the center of the Browsing Window (Fig. 4). Users can understand the overview of the recipe from this field. When each node is clicked, the node is highlighted and the corresponding "Scene" of the cooking video is played repeatedly. In the input data, an "Action Block" contains the following information.

- Text description.
- Dependency to the next Action Block.
- List of AUs that the Action Block contains.
- Corresponding video "Scene".

3.2.3 *Problems in Data Creation*

We have the following problems on automatic generation of the meta data: (1) Analysis of text recipe to extract and structuralize AUs and Action Blocks. (2) Supplement of words to the description of AUs to make complete cooking instructions.(3) Association of AUs with Shots.

We have already proposed an automatic structuring method of recipe texts in [3], and a multimedia integration method to associate texts with video according to the cooking scene in [2, 4]. However, for the following experiment, we created correct data by hand to evaluate the system. Especially, association in the scene level is too coarse for navigation, so fine grain integration in the shot was necessary.

3.3 Optimization Algorithm

The recipe procedure is optimized and rescheduled according to the *List Scheduling Algorithm* which is often used for parallel processing and compiler optimization[6]. The evaluation function is defined as the sum of the earliest allocation time and the path length. In any of the following cases, an weight α is added to the evaluation function: (1) The user can stay in front of the same kitchen resource from the previous node. (2) The two continuous nodes in the same path is also allocated continuously.

3.4 Input Devices

In a kitchen, a simple interface that does not have a problem with users with dirty or wet hands is desirable. In a preliminary experiment in the following section, users used a foot switch as a replacement to mouse buttons and/or a touch panel with a waterproof touch pen.

Fig. 6 shows a user cooking with the system in her kitchen. In this example, the user used a pen touch panel as an input device.

Figure 6: A user cooking with "Cooking Navi".

Table 1: Results of the Questionnaire.

User level	(a)	(b)	(c)	(d)	(e)	(f)
Experienced	10	10	2.0	9.5	10	9.5
Intermediate	10	7.5	4.0	10	10	7.7
Novice	8.5	9.3	4.0	8.3	9.0	9.3
Average	9.5	9.0	3.5	9.3	9.6	8.8

4. PRELIMINARY EXPERIMENT

We asked various users to actually cook with the assistance of our system. The number of subjects was 8, of which 2 of them were experienced, 3 were intermediate, and 3 were novice cooks. They cooked 2 recipes randomly selected among 4 recipes which are relatively difficult. After this experiment, they answered a questionnaire that contains the following questions (each question was answered on a scale of 0 to 10; most negative to most positive):

(a) How much did you rely on the video?
(b) How much did you rely on the audio?
(c) Did you feel that the provided procedure was appropriate?
(d) Did you feel comfortable with this system than with ordinary TV cooking programs?
(e) Did you feel comfortable with this system than with ordinary text recipes?
(f) Did you enjoy cooking with this system?

The results of the questionnaire is shown in Tab. 1. Each value is the average of the values answered by the subjects. Although it was almost the first time to cook for the 3 novice cooks, they succeeded to finish 2 dishes at the same time. In addition, 2 of them answered that they could not have finished them if it had been a ordinary text recipe. According to Tab. 1, the video and the audio is very important, even for experienced cooks. Most subjects evaluated that cooking with this software was much more comfortable than only with a video or a text recipe. Adding to that, all users answered that they enjoyed cooking very much with the software. Their comments were that "It was like playing a game", or "It was very comfortable because there was a video navigation and I did not wonder what and how I should do", and so on.

On the other hand, the rating for the rescheduled procedure was low in general. It was because the time optimized cooking procedure is often unnatural. For example, it is theoretically possible to heat multiple frying pans in parallel, but it is not safe to do so. To make the rescheduling algorithm more natural and comfortable, we are going to introduce constraint conditions derived from empirical knowledge on cooking in the future.

5. CONCLUSION

In this paper, we proposed a novel multimedia application system which provides multimedia cooking navigation. According to the result of the preliminary experiment, all users from novice (including those who have not cooked before) to experienced users could finish two dishes in parallel, while enjoying cooking more than usual. In the experiment, the effectiveness of multimedia assistance for cooking was confirmed. Users were dependent on every medium, and if just one medium was stopped in the midst of the experiment, they were bewildered and lost their way.

Many issues on multimedia analysis and indexing was derived from the software. For example, users tend to concentrate on audio especially when they are working, so synchronization of audio is required. Users rely on video more than texts, so high-level indexing accuracy is essential. Shot-based indexing was still too coarse for cooking navigation, so contents analysis within a shot is necessary. A new text analysis method is also necessary to create the navigation script automatically.

"Cooking" lies at the midpoint between a routine work such as engine maintenance and an artistic work such as painting or composition. The current simple rescheduling algorithm might not be suitable for the artistic side of cooking. In the future, we will examine how to support creativity based on users' non-restricted behavior.

6. ACKNOWLEDGMENTS

The part of this work was funded by Exploratory Software Project grant of Information-technology Promotion Agency, Japan. We would like to thank Yukitoshi Suzuki and Dr. Hidenori Tsuji at Institute of Information Technology, Inc., for their kind help in developing the system.

7. REFERENCES

[1] Counter intelligence. In *http://www.media.mit.edu/ci/*, 2004.

[2] R. Hamada, I. Ide, S. Sakai, and H. Tanaka. Associating cooking video with related textbook. In *Proc. ACM Multimedia 2000 Workshops*, pages 237–241.

[3] R. Hamada, I. Ide, S. Sakai, and H. Tanaka. Structural analysis of cooking preparation steps in Japanese. In *Proc. Fifth Intl. Workshop on Information Retrieval with Asian Languages*, pages 157–164, Oct. 2000.

[4] R. Hamada, K. Miura, I. Ide, S. Satoh, S. Sakai, and H. Tanaka. Multimedia integration for cooking video indexing. In *Proc. Pacific-Rim Conference on Multimedia*, volume II, pages 657–664, Dec. 2004.

[5] C. Kidd, R. Orr, G. Abowd, C. Atkeson, I. Essa, B. MacIntyre, E. Mynatt, T. Starner, and W. Newstetter. The aware home: A living laboratory for ubiquitous computing research. In *Proc. Second Intl. Workshop on Cooperative Buildings*, Oct. 1999.

[6] Y. Srikant and P. Shankar. *The Compiler Design Handbook: Optimizations & Machine Code Generation.* CRC Press, 2002.