Iterative Design of Adaptive Tabletop Dish Recommendation by Real-time Recognition of Dining Status

Tomoo Inoue[†]

[†]Faculty of Library, Information and Media Science, University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp

Abstract –This paper presents a series of smart dining tables named "Future Dining Table." The system recommends dishes to the user visually during dining according to his/her context. This is achieved by real-time recognition of the user's dining activity and the food remains, and by using its history. The system is supposed to be useful where repeated dish orders take place while the number of serve staffs is not enough. Evaluations are given in the recognition accuracies of eating action and food remains, which confirmed the system's practicality [1].

Keywords: Dining computing, Tabletop, Behavior recognition, Future Dining Table.

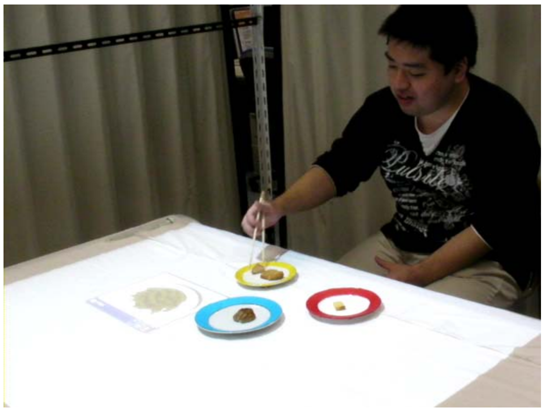


Figure 1: The system in use.

1 INTRODUCTION

Dining table, or the dining area, is a place where people get together and eat together. Regardless of whether it is at home or in public places, it is necessary for everybody because nobody can live without eating.

Information and communication technologies have become prevailing in many areas where we live. Dining table is one of the few places ICT would be deeply applied from now on.

We have been focusing on this research field as "Dining computing". We have applied the categorization of Computer Supported Cooperative Work (CSCW) to the dining computing and made the conceptual framework in 2007. In CSCW, the application areas are categorized in terms of time and space; "same time" or "different times", and "same place" and "different places." Because dining can be regarded as a kind of tasks, and dining together that involves multiple persons can be regarded as a kind of collaboration, the application areas of the dining computing can be categorized in the same way.

In this consideration, we can explore the key constituent of the dining computing applications. First, a typical CSCW application in "same time" and "same place" is an electric meeting room whose basic function is enhancing face-to-face meetings. So a typical corresponding dining computing application is a dining environment that enhances face-to-face dining. A smart dining table that understands the real-time context of the local users can be an application system of the category.

Second, a typical CSCW application in "same time" and "different places" is a desktop conferencing system whose basic function is providing a communication line when both users in different sites are in front of the cameras. So a typical corresponding dining computing application is a dining environment that connects users in different sites. A real-time tele-dining support system can be an application system of the category.

Third, a typical CSCW application in "different times" and "same place" is a physical bulletin board whose basic function is maintaining the connection between users in different times, e.g. by leaving a message. So a typical corresponding dining computing application is a dining environment that enables time-shifted communication and/or dining. Many families hold members with different living patterns and they sometimes have difficulty in finding shared dining time. This may be helped by such a system.

Fourth, a typical CSCW application in "different times" and "different places" is an e-mail system whose basic function is providing message transfer between users in different sites and times. So a typical corresponding dining computing application is a dining environment that enables time-shifted remote communication and/or dining. This may be useful for a family with some members live afar in different time-zones. It has the least constraints from time and space, but has least availability of live communication channels.

We call the dining environment in the dining computing "Future Dining Table," which includes all application systems of all categories. However not all the systems can be built at the same time. The adaptive tabletop dish recommendation systems that are presented in this paper are early application systems of the dining computing in real-time and face-to-face setting.

[1] The work reported in the paper was partially supported by the JSPS Grant-in-Aid for scientific research 22500104, the Research Grant of Telecommunications Advancement Foundation, and the Research Projects of Graduate School of Library, Information and Media Studies, University of Tsukuba.

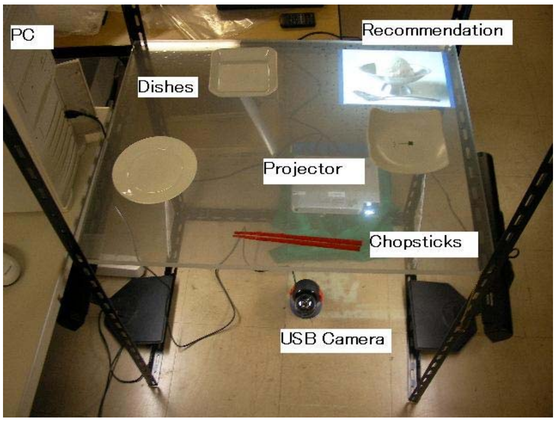


Figure 2: Appearance of the first FDT.

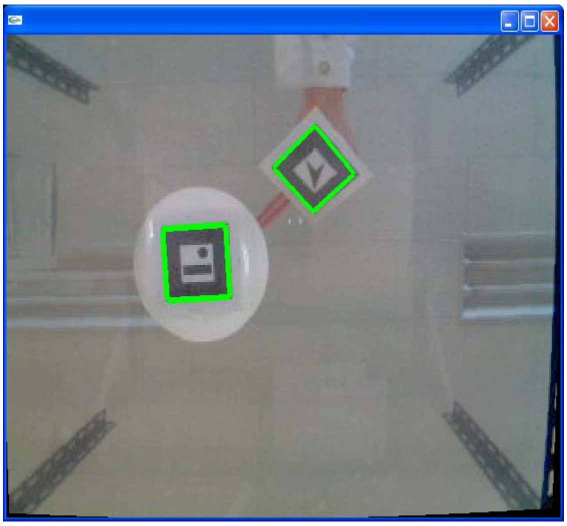


Figure 3: Visual markers on the bottom of a dish and on the user's hand.

So far, we have developed a few versions of FDT. It has evolved gradually. After the description of the related works in Chapter 2, this paper introduces those FDTs in Chapter 3 and Chapter 4. The system recommends dishes according to the dining status of the user (Fig. 1). The system has a USB camera as the input device, a PC as the information processing unit, a projector as the output device, and a table for dining and for information display.

In Chapter 5, the system's recognition accuracy of eating action was examined and compared to the previous system with different recognition method. The system's recognition accuracy of food remains was also examined in Chapter 6. These evaluations indicated the practicality of the system.

2 RELATED WORKS

There are other works in the dining computing besides our works.

The diet-aware dining table was a table that can track what and how much the users eat. To enable automated food tracking, the table was augmented with two layers of weighing and RFID sensor surfaces [1]. Although this table is a good example of the dining computing, it is not very easy to build such a special table. It is expensive and not portable.

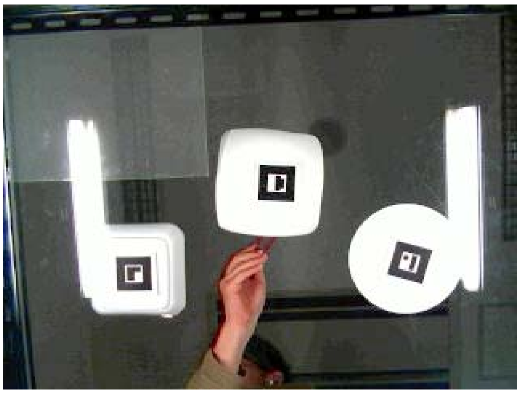
Then this table was applied to a dining game for better dietary behaviors of kindergarten children. The game was to color the picture of a child's favorite cartoon character. Each food item on the table corresponded to a particular crayon color, and the color would be drawn on the character when the corresponding food item was eaten. To make his/her favorite cartoon character colorful, the child was then motivated to eat and finish all the food items including those that he/she dislikes [2].

The Irodorin was a research prototype system that decorated a white dish plate with the projected colored light pattern according to the food color on the plate. This aims to make the food delicious [3]. This system was extended to the DiningPresenter, which showed drawings on and around a dish plate according to the food remains. The system also included the authoring tools in a kitchen [4]. This was for motivating children to eat up the food. The goal of the research was in this sense similar to [2]. The target user and the goal of our research in this paper is different, which aims to support the user's dining on the selection of dishes.

The pHotOluck was a system to help interpersonal communication during dining. It projected pictures on vacant dishes from above so that pictures gave clues to start conversation [5]. This research supports meal times in terms of communication but it does not use dining status or a user's behavior. We have also developed such a system for supporting communication [6], but this paper presents iterative design of FDT, which is unique with this system.

CoDine was a dining table embedded with interactive subsystems that augment and transport the experience of communal family dining. CoDine connected people in different locations through shared dining activities such as gesture-based screen interaction, mutual food serving, ambient pictures on an animated tablecloth, and the transportation of edible messages [7]. The goal of this research is to provide co-presence feelings for remote users, and this corresponds to the category of real-time tele-dining support systems. Another research on this category is Being Here System [8].

CU-Later was a system that played a recorded video of remote dining after a specific time shift when the local user was in front of the display placed on the dining table. So the local user could watch the video automatically when he/she was eating. The system recorded the local user's dining session as well when the video was played, so similarly the remote user could watch the local user's dining later on [9]. It is like a video mail exchange with automatic playback and recording. The goal of this research is to provide communication channels for the users in different time-zones, and this corresponds to the category of time-shifted tele-dining support systems. Another research on this category is KIZUNA system [10].



(a) Original image

(b) Extracted image

Figure 4: Hand recognition by image processing.

3 FUTURE DINING TABLE

3.1 First Version

FDT is a tabletop system and recommends dishes according to the dining status of the user.

The appearance of the first version is shown in Fig. 2 [11]. The table was a transparent 15mm thick acrylic board with 60cm depth and 75cm width, sealed by the transparent screen sheet for video projection so that the recommendation can be displayed on the table. Each dish on the table had a visual marker on its bottom. The markers were recognized from the image taken by a USB camera under the table. The user also put the visual marker on his/her hand, which makes him/her feel unnatural (Fig. 3). The recognition was not very robust with the roll of the wrist. The marker was sometimes occluded.

3.2 Second Version

To prevent these problems, the user's hand was recognized by image processing in the second version [12]. The hand was recognized by the background subtraction method. The image of the table was captured first as the background. To cope with the gradual change of the shooting condition, every frame was combined at the rate of 0.01 with the background. To cope with the change of the dish location and the change of the recommendation image, the background was updated. It was replaced by the foreground image when the foreground remained the same in 50 consecutive frames. Then the background was subtracted from the current frame, and the changed region was gained. After the opening to delete minor noise, the region with some area (more than 1000 pixels) was extracted by the labeling (Fig. 4). This is the hand recognition process.

However, the same visual markers were still used on the bottom of the dishes with the transparent table, and the camera was under the table, which some users commented might be an issue. Also, the food remains were not measured directly from actual food.

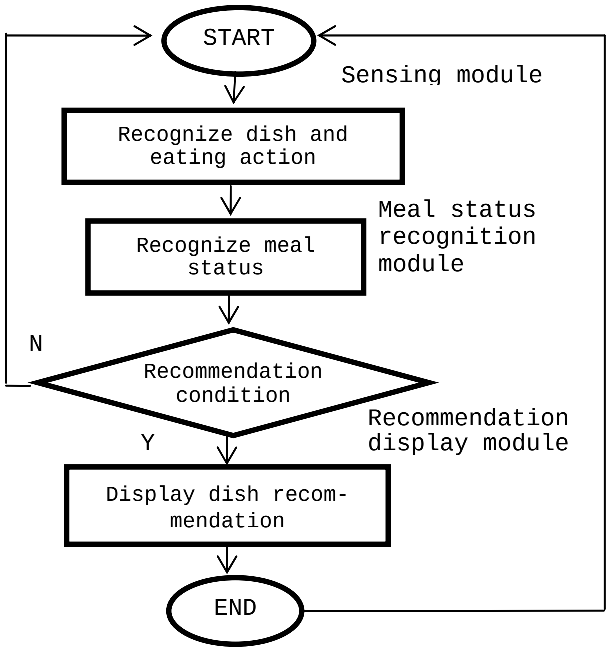


Figure 5: Software procedure.

3.3 Third Version

The third version has been modified to solve above-mentioned issues. A USB camera has been installed on the ceiling of the table and has recognized the dishes and the user's behavior. A white table has been used for the dining table with information display. A projector has also been installed on the ceiling of the table to project information on the table [13].

4 FDT SOFTWARE

The software of the FDT version 3 is explained in this chapter. It has been implemented by Microsoft Visual C++ on Windows OS. Figure 5 shows the procedure, which is constructed by the sensing module, the meal status recognition module, and the recommendation display module.

4.1 Sensing Module

Figure 6: Hand recognition from the ceiling.

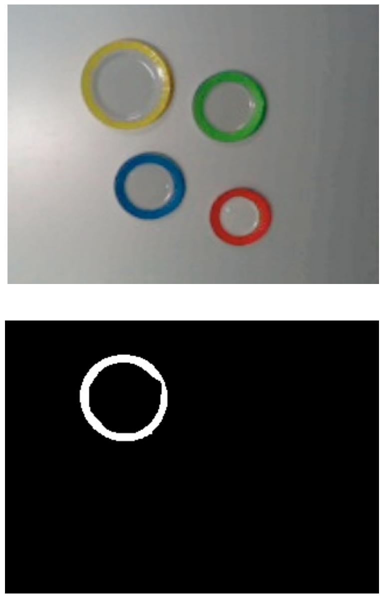


Figure 7: Dish recognition by colors.

The sensing module has been implemented with the Intel OpenCV. Four major processes are performed after obtaining the table image from the USB camera. They are hand recognition of the user, dish recognition, eating action recognition, and food remains recognition. When eating action is recognized, this triggers the meal status recognition module. When eating action is not recognized, the table image is replaced by its next frame. The frame rate of the camera is 9 frames per second.

Hand recognition

The finger tip of the user becomes most distant from the user's body and becomes close to the dishes on the table when eating action takes place.

The hand is recognized by the background subtraction method. The image of the table is captured first as the background. To cope with the gradual change of the shooting condition, every frame was combined at the rate of 0.01 with the background. To cope with the change of the dish location and the change of the recommendation image, the background is updated. It is replaced by the foreground image when the foreground remains the same in 25 consecutive frames. Then the background is subtracted from the current frame, and the changed region is obtained. After the opening to delete minor noise, the region that is more than 1000 pixels is extracted by the labeling. Figure 6 shows the hand recognition where the hand moves 23.3 cm/sec.

Dish recognition

The dish recognition is performed by the colors of the dish rims. Base colors of the round dishes are white with 5 different colors on the rims. The colored regions are extracted for each color and labeled to obtain the resulted region. Figure 7 shows the yellow dish recognition. The two points with the maximum and minimum x coordinates of the obtained region and the line between these two points can be gained. The line between the two points where the y coordinates are the maximum and minimum of the obtained region can be gained in the same way. The dish center can be gained as the intersection of the two lines.

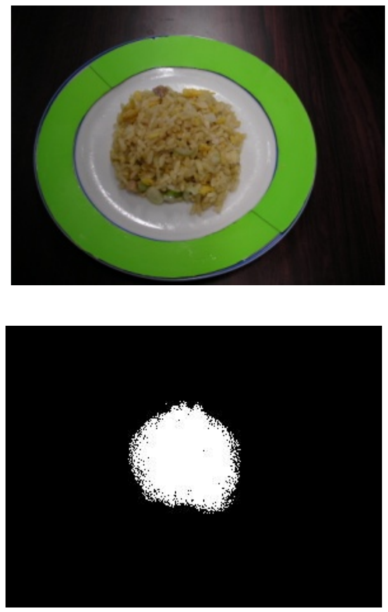


Figure 8: Food remains recognition.

Eating action recognition

Eating action recognition is performed by the distance between the dish center and the finger tip of the user. When the distance becomes less than the radius of the dish plus 10 pixels, it is determined that the eating action occurs empirically. It is because the user holds chopsticks with his/her hand but the chopsticks are too thin to get the image. The user picks up the food with the distance. Usually the system has multiple dishes on the table. If the finger tip becomes

Table 1: Example rules for the recommendation.

Status	Recommendation
Food remains is 25%	Closing food (Soba etc.)
Food remains is 5%	Desert (Ice cream etc.)
25% of 1st plate (Caesar salad etc.)	2nd plate (Gyoza etc.)
3 serial eatings of fries (Fried chicken etc.)	Same category plate (French fries etc.)
More than 6 drinkings (Beer) in the recent 10 eatings	Nibbles (Boiled soybeans etc.)

less than the "eating action" distances for multiple dishes in this case, it is treated the food is taken from the nearest dish. However this does not really happen because with the determined distance it is not realistic for the hand to occupy such a position.

Food remains recognition

One of the method to recognize food remains is to estimate from the number of eating actions [11][14]. This method needs the information of the number of eating actions to finish the dish. Then the food remains is estimated as the ratio of the current number of eating actions to that of finishing the dish. Because the number of eating actions to finish the dish can vary depending on the various factors such as individual, food, health condition, the estimation is not very accurate.

In welfare and medical care, meal has been recorded for management of health. This record is typically a written memorandum showing the given menu or sometimes the remains after meal. Because making a written meal record is time consuming and is not easy when there are many clients, automatic recording of meal has been researched. Recognition of video-recorded dining behavior is one such approach. Hand movement is detected and eating is recognized from the video of a dining room by applying hidden Markov model [15]. However, it does not recognize what the diners eat and the order of eating things.

Another method is to measure the weight of the foods. The food remains is calculated as the ratio of the current weight to the initial weight [1][16]. This method can be accurate but is not very easy to implement because the weight of each dish must be measured. Weight scales may be embedded to the dining table.

Our research employs image processing method. This method estimates food remains from the 2D images of the dish. The result may not be very accurate due to the use of 2D images, but may be obtained fairly easily and is directly based on the actual food.

Figure 8 shows the result. Non-white pixels are counted from the white dish area. This is compared to the number of the initial pixels and the rate is regarded as food remains.

Figure 9: Dish arrangement in the evaluation of eating action recognition.

4.2 Meal Status Recognition Module

Meal status recognition module receives the food remains for each dish and the eating action history. Then the meal status is estimated from these. The food remains basically show the whole progress of meal, while the eating action history shows more precise chronological food consumption trend. Any rule to recommend a dish from the combination of the food remains and the eating action history can be made.

Example rules have been implemented in the system as shown in Table 1. "Rule 1: Food remains are 25%," and "Rule 2: Food remains are 5%" are derived from the survey result of dish recommendation timing. Recommended dish was most felt like placing the order at 25% and 5% food remains in the survey. Also recommendation was least disturbing at 25% and 5% food remains [17].

4.3 Recommendation Display Module

After the meal status is recognized and the recommended dish at the time is determined by the recommendation rules, the recommendation is displayed on the table. The recommendation display module shows the image of the recommended dish dynamically depending on the existing dish positions.

The existing dish positions are known from the dish recognition. The recommended dish position can be determined in various ways and has been investigated. The positions that are near to other dishes and culturally right are known to be felt easy to eat [18]. The module uses this finding.

5 ACCURACY OF EATING ACTION RECOGNITION

5.1 Procedure

We have evaluated the accuracy of eating action recognition of the system. The same procedure was applied for the comparative evaluation to the first version which used the visual marker for the hand recognition and the current version which uses image processing for the hand recognition.

Table 2: Eating action recognition of the third version (%)

	Precision	Recall	F-measure
Subject A	78	88	82
Subject B	96	96	96
Subject C	100	75	86
Average	90	86	88

Table 3: Eating action recognition of the first version (%)

	Precision	Recall	F-measure
Subject D	95	79	86
Subject E	82	75	78
Subject F	68	54	60
Average	82	69	75

The subjects were three right-handed male university students. All were the first time users of the system. They were instructed to eat normally as in everyday. Three dishes were set on the table as shown in Fig. 9. A snack was chosen as the food on the dishes because the snack was of small pieces and we could control the number of eating actions easily. Eight pieces of the snack were set on each dish. This meant the subject conducted 24 actions. The behavior in the session was videotaped. The record of recognition by the FDT was compared with the videotaped behavior, which provided the correct answer.

5.2 Result

The results are shown in Table 2 and Table 3. Precision, recall and F-measure were used as the measures. These are originally from information retrieval, and have become generally used as the measures of such evaluation. Precision is defined as true positive / true positive + false positive. When an eating action from a false dish is recorded by the system, precision decreases. Recall is defined as true positive / true positive + false negative. When the system overlooks an eating action and does not record the action, recall decreases. F-measure is defined as the harmonic mean of precision and recall; F = 2 / (1/Precision + 1/Recall). It is used to represent the performance of both precision and recall in a single measure.

Because this type of evaluation is unique, we cannot discuss the results in comparison with those of other research. The precision was 90% and the recall was 86% on the average, resulting 88% of F-measure in the current version. Whereas the precision was 82% and the recall was 69% on the average, resulting 75% of F-measure in the first version. The change in the hand recognition method succeeded in improving the recognition rate.

Figure 10: Dishes used for food remains evaluation. (Top) Boiled soybeans. (Bottom) Fried rice.

6 ACCURACY OF FOOD REMAINS RECOGNITION

6.1 Procedure

We have also evaluated the accuracy of food remains recognition of the system.

The subjects were three right-handed male university students. They used FDT for dining. A single dish was served at one time. Two different dishes were served as shown in Fig. 10.

One was "boiled soybeans," which each piece was not very small and the number of pieces was countable. The quantity was measured in pieces with this dish. We set 20 pieces initially, which means 5% decrease for every eating action.

The other was "fried rice," which had different appearance from the former thus was expected to produce a different result. The quantity was measured by weight with this dish. A cooking scale with the minimum scale of 5g was used for weighting.

Both dishes were common in the supposed environment of the system. Each subject ate up these two dishes.

6.2 Result

The results are shown in Table 4 and Table 5. The remains recognition for fried rice was more accurate than that for boiled soybeans. The result of 5% in Table 4 shows the recognition bias clearly. The remains are recognized by the 2D image from the top. When the remains are recognized as 5%, the pixels of food are 5% of the initial pixels. This is only achieved when no overlapping of the food is found even if extraction of the pixels is accurate, and is often not realistic. In the experiment, only one piece of soybeans pod was 5% but the pixels for the one soybeans pod was clearly

Table 4: Food remains recognition for boiled soybeans (%)

	75%	50%	25%	5%
Subject A	60.0	40.0	10.0	0.0
Subject B	60.0	40.0	20.0	0.0
Subject C	75.0	35.0	10.0	0.0
Average	65.0	38.3	13.3	0.0

Table 5: Food remains recognition for fried rice (%)

	75%	50%	25%	5%
Subject A	53.3	46.7	30.7	5.3
Subject B	62.7	40.0	20.0	1.3
Subject C	73.3	60.0	37.3	9.3
Average	63.1	48.9	29.3	5.3

more than 5% of the initial pixels because the soybeans were piled up with overlapping on the dish. This is why the recognition of 5% always came after the actual 5%.

The reason why the difference in 75% was the biggest in Table 5 can be explained with the same recognition bias. Fried rice is also piled up on the dish. When the subject eats fried rice, he/she does not usually eat up a particular area, so does when eating a pizza. Instead, he/she often eats some upper part without finishing to the bottom. Because of the 2D image recognition, height decrease is not recognized, and the remains are recognized as the same in this case. Thus actual quantity becomes less than the recognized quantity by 2D image.

However, the differences between the recognized remains and the actual remains were within 5% when measured by the weight. Those two values were in good co-relation.

7 CONCLUSION

In this paper first presented the short introduction of the concept of the dining computing.

Second, a smart dining table system “Future Dining Table (FDT)” that has been developed iteratively for versions as an application system of the dining computing was explained. The system recognizes the user’s dining activity and the food remains in real time, and along this context recommends dishes to the user visually on the table during dining.

Then the evaluations of the recognition accuracies of eating action and food remains were explained, which indicated the system’s practicality.

According to the categories of the dining computing and other demands such as communication support, FDT will be further extended in the future.

ACKNOWLEDGMENT

This research was partially supported by the JSPS Grant-in-Aid for scientific research 22500104, the Telecommunications Advancement Foundation, and the Research Projects of Graduate School of Library, Information and Media Studies, University of Tsukuba.

REFERENCES

[1] K. Chang, S. Liu, H. Chu, J. Hsu, C. Chen, T. Lin, C. Chen, and P. Huang, ``The Diet-aware Dining Table: Observing Dietary Behaviors over a Tabletop Surface,’’ LNCS, Vol.3968, pp.366–382 (2006).

[2] T. Lin, K. Chang, S. Liu, and H. Chu, ``A Persuasive Game to Encourage Healthy Dietary Behaviors of Young Children,’’ Adjunct Proceedings of the 8th International Conference on Ubiquitous Computing (2006).

[3] M. Mori, K. Kurihara, K. Tsukada, and I. Shiio, ``An Augmented Table to Enrich Food Color,’’ Proceedings of the 70th National Convention of IPSJ, pp.4-245-246 (2008). (in Japanese)

[4] M. Mori, K. Kurihara, K. Tsukada, and I. Shiio, ``Dining Presenter: Augmented Reality System for a Dining Tabletop,’’ Supplemental Proceedings of the 11th International Conference on Ubiquitous Computing, pp.168-169 (2009)

[5] K. Nishimoto, K. Amano, and M. Usuki, ``pHotOluck: A Home-use Table-ware to Vitalize Mealtime Communications by Projecting Photos onto Dishes,’’ Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp.9-16 (2006).

[6] S. Kutsuwada, and T. Inoue, ``Toward a System for Facilitating Multi-party Conversation Based on Dynamic Group Recognition,’’ IPSJ SIG Technical Reports, Vol.2009-GN-072, No.6, pp.1-6 (2009).

[7] J. Wei, X. Wang, R. L. Peiris, Y. Choi, X. R. Martinez, R. Tache, J. T. K. V. Koh, V. Halupka, and A. D. Cheok, ``CoDine: An Interactive Multi-sensory System for Remote Dining,'' Proceedings of the 13th International Conference on Ubiquitous Computing, pp.21-30 (2011).
[8] M. Nawahdah, and T. Inoue, ``Building a high realistic media space by superimposing a remote person's figure on the local view,'' Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, pp.416-422 (2012).
[9] H. Tsujita, S. Yarosh, and G. D. Abowd, ``Cu-later: A Communication System Considering Time Difference,'' Proceedings of the 12th international conference adjunct papers on Ubiquitous computing, pp.435-436 (2010).
[10] M. Nawahdah, and T. Inoue, ``Development of KIZUNA system to realize time-shifted virtual co-dining,'' Proceedings of the 6th International Conference on Collaboration Technologies, in press (2012).
[11] Y. Seto, Y. Noguchi, M. Tosaka, and T. Inoue, ``Development of Another Dish Recommender Based on Dining Activity Recognition,'' IEICE Technical Report, Vol.107, No.554, pp.55-60 (2008). (in Japanese)
[12] Y. Seto, Y. Matsusaka, and T. Inoue, ``Realtime Dining Activity Recognition in Another Dish Recommender,'' IPSJ SIG Technical Reports, Vol.2009, No.3, pp.1-6 (2009). (in Japanese)
[13] Y. Kourai, Y. Otsuka, and T. Inoue, ``Adaptive Tabletop Dish Recommendation System by the Recognition of Realtime Dining Activity,'' IPSJ SIG Technical Reports, Vol.2010-GN-77, No.18, pp.1-8 (2010). (in Japanese)
[14] Y. Dong, A. Hoover, and E. Muth, ``A Device for Detecting and Counting Bites of Food Taken by a Person during Eating,'' IEEE International Conference on Bioinformatics and Biomedicine 2009, pp.265-268 (2009).
[15] J. Gao, G. A. Haupymann, A. Bharucha and D. H. Wactlar, ``Dining Activity Analysis Using a Hidden Markov Model,'' Proceedings of the 17th International Conference on Pattern Recognition, Vol. 2, pp. 915–918 (2004).
[16] T. Kawashima, Y. Tanisugi, and Y. Mitsudo, ``Dining Monitoring System Using Sensing Tray and ID-ware,'' IEICE Technical Report, Vol.106, No.285, pp.61-66 (2006). (in Japanese)
[17] Y. Seto, and T. Inoue, ``Study of Recommendation of Dishes for a Dish Recommender System,'' IEICE Human Communication Group Symposium, HCG2009-C5-1 (2009). (in Japanese)
[18] Y. Otsuka, and T. Inoue, ``Comparative Study Of Attractive Dish Arrangements for a Tabletop Dish Recommendation System,'' Proceedings of 2011 IEEE International Conference on Systems, Man and Cybernetics, pp.357-362 (2011).

(Received March 8, 2012)
(Revised June 15, 2012)

Tomoo Inoue is Associate Professor of the Faculty of Library, Information and Media Science of University of Tsukuba. His research interests include HCI, CSCW, and CSCL. He received Ph.D. in Engineering from Keio University in 1998. He is a recipient of awards including Best Paper Award, Activity Contribution Award and SIG Research Award from Information Processing Society of Japan (IPSJ). He has served a number of academic committees, currently including IEICE SIG Human Communication Science, IEICE SIG Multimedia on Cooking and Eating Activities, VRSJ SIG Cyberspace, IPSJ SIG Digital Contents Creation, IEEE TC CSCWD, and IEEE TC HCI.