Building a Recognition Process of Cooking Actions for Smart Kitchen System

Fong-Gong Wu and Tsung-Han Tsai

Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan
fonggong@mail.ncku.edu.tw, scott0510@hotmail.com

Abstract. Smart kitchen should be focusing its development on the actual interaction with users and the environmental objects rather than emphasizing on complicated instructions and feedback. Unfortunately, the current techniques can only be designed to identify motions and basic actions. The main purpose of this paper is to analyze and research user motions and actions involved in the process of cooking, including ingredient preparation, and to discover multiple action identification characteristics for the user and cooking utensils. By using the video analysis, ultimately, the project will use these characteristics to establish a reliable cooking-action database. Our study can distinguish between similar actions. The model is primarily used to identify, understand and differentiate the extent of the intellectuality of user motions. This model may be used in the future in the application to cooking support systems or other smart kitchen developments.

Keywords: Smart Kitchen, Human Behavior Taxonomies, Motion analysis, Video analysis, Decision Tree Learning.

1 Introduction

Nowadays, people's daily lives are closely related to technologies. Particularly our family lives with the accelerating pace of technological innovation, which has always been the focus of life. Also, many activities have been created within the family environment. If the environment can be made to reciprocate this behavior and respond to human behavior, it will lead to several advantages [1]. To provide useful devices which are most closely related to daily life, an intelligent family lifestyle has comprehensively seen the direction of future development worldwide.

Combined with various sensing technologies, the smart home system is also used to monitor and analyze the daily routine behaviors of residents. When unusual behaviors are sensed which are different to those being established in the system database, then residents might have some potential problems which needed to be further understood, like physiological or physical disease problems. Different systems may choose different sensing technologies based on its purpose. In general, technologies most commonly used by existing smart home systems are mainly divided into two main categories—Direct Environmental Sensing and Infrastructure Mediated Systems [2]. Direct Environmental Sensing takes advantage of certain facilities like camera or

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part III, LNCS 8515, pp. 575–586, 2014.

RFID to offer considerably useful information for identification of actual activities of human beings, yet the installation and maintenance costs are relatively higher. Infrastructure Mediated Systems, however, merely need to install sensors on certain existing facilities. Compared to the Direct Environmental Sensing which has to use a large number of simple binary sensors in an area, the Infrastructure Mediated Systems can relatively lower the complexity of installation and maintenance costs.

In order for the system to be able to correctly guide the user on cooking steps, it is necessary to first think about how to let the system know when the user has completed an action [3]. Three ways which can be used are as follows: (1) User notification; (2) Use of an IC tag; and (3) System recognition. The first way allows the user to directly notify the system by pressing a button so that the system knows that a certain action has been completed. The user then follows the directions for the next step.

The User centric Smart Kitchen System was created based on this design in which the system directly identifies the movements of the user and then gives support and feedback to the user if necessary. In general, the system is designed to identify hand locations, postures or cooking utensils. Food ingredients are not identified, as too many characteristics, such as colors, shapes, grains and textures make identification very difficult. There are several pieces of equipment which can be used for identification; besides the RFID identification system motioned above, an identification technique based on image identification is also available. The latter does not add any electronic tag on objects or the user, but directly turns a screenshot into pictures or images by using cameras or thermal imaging detection. Images or pictures used to take a screenshot could be a set of logistic, systematic totem. Not adding an image on the kitchenware or cooking tools for simple identification by the appearance is also workable. Many of the recent dining and auxiliary systems employ this type of identification technique to identify and recognize hand gestures and basic movements.

Smart kitchen should focus its development on the actual interaction with the user and the environmental objects, rather than emphasizing on complicated instructions and feedbacks. Those extra and unwanted motions do nothing but easily distract the user [4]. If the operation of smart equipment and the feedback approaches are inconsistent with the behaviors that the users are familiar with, or even force the user to have to re-learn and adapt to new ways of interaction, then such a design may contrast with the original idea for better quality or even causes more inconvenience. Nevertheless, as mentioned in the prior section, the development of existing smart kitchen systems focus relatively on the systematic application and the integration of the system and environment, like the cooking support system, for example, which emphasizes on the integral process planning of the cooking guide.

Cooking usually involves many complex actions, rather than merely simple ones like picking up or cutting. It is not enough for a smart system to identify hand movements [5]. Thus, in order to allow the system to better understand which movements and actions are being carried out by the user, it should first build an integrated action data set as references for the system to map user actions and movements.

The main purpose of this paper is to analyze and research user motions and actions involved in the process of cooking, including ingredient preparation, and to discover multiple action identification characteristics for the user and cooking utensils.

Ultimately, the project will use these characteristics to establish reliable Human Action Recognition Process. The model is primarily used to identify, understand and differentiate the extent of the intellectuality of user motions. This model may be used in the future and applied to the cooking support system or other smart kitchen developments, such as the auxiliary system of recipe amplification. In addition, because our study investigates the kitchen of smart home, our target environment is focused on home kitchen not commercial kitchen, which may be different not only in spatial allocation but also in cooking actions.

Miyawaki and Sano [6] developed a cooking navigation system utilizing the virtual agent, which is made by augmented reality tech to assist user to accomplish all cooking actions.

The scope of the research, which involves Human Behavior Analysis (HBA), widely ranges into applications from several fields, such as motion detection, background extraction and high-leveled abstraction behavior models [7]. Prior to conducting the behavior analysis, however, it is necessary to first define the relationship of layered behavior. There is plenty of literature related to HBA taxonomies [8] defined three layers of taxonomy. This first layer is called “action primitive” or “motor primitive”, an action layer which is made up of a series of different or repeated action primitives. If the layer is involved in a wider range, including objects or interaction between the user and the environment, then it is called an “action layer”. Let’s take the action of making coffee as an example. A single arm or hand motion is called “action primitive”, putting the teapot on the stove or picking up a cup from a table is called “action”, and the whole process of making coffee is so-called an “activity”.

Chaaraoui et al. [7] categorized HBA as being divided into four layers: motion, action, activity and behavior according to semantics and time frame. The layer of “motion” is mainly used to detect the “movement” or something like the measurement of eye position or head posture. At the “action” layer, it has not been merely used to differentiate human motions; rather, it includes the interaction between humans and objects. “Second” is used as a measurement unit; an “activity” is made up of all types of “actions”, which are measured anywhere from several seconds to several minutes, such as cooking or taking a shower. As a result, to better understand a user’s activity, it is necessary to identify and classify a series of actions. The final layer is called “behavior”, in which the time unit ranges from a single day to several weeks. It is used to detect the subject’s abnormal behaviors in advance, such as discovering whether the subject suffers from some symptoms of a disease, like Alzheimer's disease (AD), for example, through observing and analyzing lifestyle, habits, and routine behaviors.

The method of defining behavior based on different layers is very helpful for doing relevant research, as the researchers can effectively identify the preferred layers they are going to identify, and avoid unwanted ones. To summarize the above classifications, this study focuses on the exploration of the “action” layer, including the operational interactions between humans and utensils. The “activity” of “cooking” is made up of these series of “actions”.

2 Methods

2.1 Content Analysis

Frequent cooking actions must be listed first before being recognized and understood by the system. The research for this study is set in the home kitchen, and the target user is set to be those who frequently use the kitchen, regardless of gender. It is believed to be more appropriate to collect the cooking actions from general commercial recipe books. Later, the ideal target data can be analyzed and induced based on the Content Analysis. Content Analysis is a methodology of quantitative analysis which is based on the contents of the literature. It converts non-quantitative texts into quantitative data for the purpose of establishing meaningful classification items to analyze specific characteristics, features, or trends.

According to the definition, the research scope of this study is defined at the very beginning. A Western cuisine recipe book, which ranked number 1 in the current market, was selected. This recipe book provides 119 dishes. Any verb used to describe the cooking procedure of each recipe will be extracted and recorded into a form of Microsoft Excel. Because the difference between “Action” and “Motion” has been clearly defined previously in this study, this study will record and extract from the texts of the selected recipe book those actions which have an interactive relationship with the objects, such as shredding or stirring. As a result, some motions like stretching out or raising a hand or certain cooking activities such as boiling or roasting, will not be counted and included in this research.

2.2 Expert Interview

Those actions, which are extracted from the texts of the selected recipe book, may partially differ from those performed in real life, or actually are the same ones but merely with a different description in the text. Therefore, before determining the final target actions, this study will conduct an expert interview to evaluate and correct those action items listed previously.

This study constructed an open-ended interview with two professors who have professional backgrounds in cooking. Through these two professors’ expertise, the interview aims at determining which actions are suitable for the research scope of this study.

2.3 Recording Cooking Action Videos

When the action items are decided, video recordings will be conducted to record these actions. The first step for video recording is to set the environment as well as the recording equipment. The filming environment is set in a home kitchen, rather than in a laboratory. From this picture it is clear that the work area of the kitchen is divided into three parts: stove, countertop, and sink. Then the action demonstration and video recording will be conducted based on the work areas which correspond to different action items. This study chose subjects with cooking experience for this action demonstration because, first, the research scope is set in a home kitchen, and second, the action items are selected and established from the texts of recipe books.

Concerning the kitchenware, according to the action list we presented, a general kitchen knife is selected for the cut action, and other utensils are rod and turning shovel. The knife tip and shank are respectively tagged with a round green sticker (with a 16 mm), that makes it clearly stand out from the background of the environment. This round green sticker is used as a feature point for image analysis (As Figure 1). Moreover, the distance of two dots between the utensil tips and shank is different. This is used as a reference for determining the displacement of the z-axis (in an orthogonal relation with the image screen). The distance between the camera and the kitchenware is 90 cm.

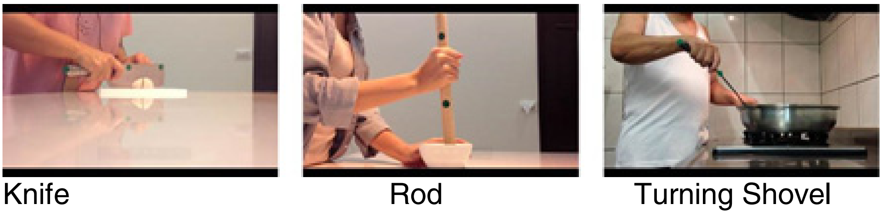

Fig. 1. Utensils

Concerning the video recording equipment, a camera with a resolution of 1920*1080 pixels at 30fps is selected, the distance of utensil and camera is 90 cm. When the recording environment and video equipment are all set, it's time for recording the cooking actions. At the beginning of the recording, a pure action, which simply operates the kitchenware without food, will be recorded. Later, the main purpose for the operators to hold the knife and repeat the chopping actions on the same side is to calculate errors in the video analysis, which is used to define the green color coordinate, followed by the implementation and recording of all kinds of actions. Every single action is operated by a single operator, with a single action lasting for 15 seconds. While filming, operators will be asked to hold the utensil vertically toward the countertop, and start the action after the recording has started for one second. When all of these files are saved, set and well organized, the next step will be to make the video set for the analysis of the action parameters.

2.4 Video Analysis

After completing the recording work, videos will be converted to Avi format which will be imported to the MATLAB for image analysis. First, according to the different environment, the green points on the utensils will undergo an RBG Coordinates Analysis, which will find the parameter threshold of the green points among red, blue and green colors. The purpose of doing so is to adjust parameter threshold of two green points captured by each video (Figure 2).

The steps of analysis are as follows: First, execute the file named RGB_ROI_ColorAnalysis_ReadVideo after turning on MATLAB. Second, type the file path which you want to analyze and then circle the green dots on the utensils by cross cursor tool, and the analysis results of RGB color spaces will be obtained after clicking. Then definite the color spaces of green dots according to the results.

Fig. 2. Color Analysis Process

Then, execute the file named Motion_Record by MATLAB, import the cooking action videos one by one according to classification, and the process of the analysis is as figure 3, recorded action videos will be analyzed for later establishing moving track diagrams of green points which correspond to a time axis, including five diagrams of analyzed results— Amplitude and Frequency on the X-axis Y-axis and Z - axis and variations of two-point distances. Also it will make a video output of the analysis process.

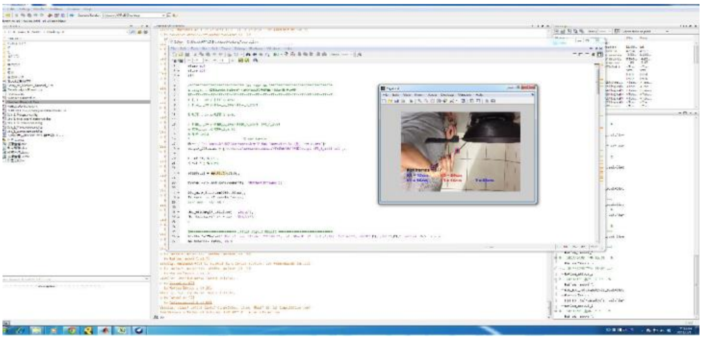

Fig. 3. Video Analysis Process

2.5 Data Analysis

After completing the analysis of all videos, each action group, such as cooking action groups for slicing or dicing, will be undergo a data analysis to identify the parameter difference of sub-actions in each group, define each parameter's threshold, and finally analyze and organize each action group as diagrams to establish a database of action parameters.

Attribute Analysis. There are several steps after analyzing action videos: First, we organize the peak data of the diagram into Microsoft Office Excel. Second, classify the data that might be Key attribute from the action videos into a table, columns are action classes and rows are attribute classes. And then, save the file as Comma separated Value (CSV) after completing the table for machine learning to proceed researching.

Decision Tree Learning. After building data table, we start to precede Decision Tree Learning by Weka. Weka is free software for data analysis and predictive modeling which is written by java and the developer is University of Waikato in New Zealand. There are many standard data mining tasks and operator methods to choose, and the operating interface is very easy for users. Plus, it is able to be operated in almost every system, including Linux, Windows, OS X and etc.

First, open the Weka, click “Explorer” and choose the database table which is csv file from the “Preprocess” label, and we can see some numerical value of table classes and data. And then move the cursor to “Classify” label, click the “Choose” button, choose the Decision Tree Learning methods. This research is completed by representative operator method called SimpleCart. At last, choose the “Start” button to start analyzing, and the results will be presented in the “Classifier Output” column.

The analysis results are the classification results of SimpleCART Decision Tree, different attribute data thresholds, total Correctly Classified Instances and Precision, and detailed accuracy by class, including Precision and Recall. At last, proceed Confusion Matrix by the analysis result and then present the final matrix.

3 Analysis and Results

3.1 Cooking Actions Taxonomies

The first half of this study mainly organizes general actions for cooking by first analyzing a Western cuisine recipe book through Content Analysis. Later, a preliminary action list is established and then given to cooking professionals for revision and to solicit suggestions for completing a final list of cooking action items.

In the first phase of conducting the Content Analysis, a total of 119 dishes will be recorded with 1,607 verbs. 45 different action items will be obtained after organization.

Table 1. The sub-actions list (Recipe)

Action	Sub-action					
Cut	Cut	Slice	Cut into two	Cut into stick	Julienne	Chop
	Mince	Dice	Cube	Gash	Cut into angularity	Cut into segments
	Cut into rods					
Press	Mush	Crash				
Frying	Frying	Shallow-Fry	Stir-Fry			

Then, these 45 action items will be grouped based on operational attributes. For example, those actions, such as Slicing, Cubing or Julienning, are all associated with a knife; therefore, all of them will be categorized under the general classification of cutting. Finally, a total of 27 action items will be organized in order to conduct a simple descriptive statistical analysis for calculating the total percentage of each item. Table1 showed sub-actions list of some actions. The Above results will later be given to two professors who are cooking experts for correction.

Based on the ideas and feedback given by cooking experts and the observation of some cooking videos, partial items with similar actions will be revised. Finally, a final list of actions will be completed as shown in Table 2.

Table 2. The sub-actions list (Final)

Action	Sub-action					
Cut	Cut	Slice	Cube	Dice	Julienne	Mince
Press	Mush	Crash				
Frying	Frying	Shallow-Fry	Stir-Fry			

3.2 Cooking Actions Video Set

After analyzing the color space, we will get three figures which show the color distribution of RGB individually. We definite the green dots spaces as Index_NonSelect=find(Normal_R>0.133 | Normal_G<0.602 | Normal_G>0.684 | Normal_B<0.275) according to the results data, red value for <0.133, green value for <0.602 and >0.684, blue value for 0.275 which is not detected, the location of green dots on the moving utensils is able to be identified precisely through this method. Because there is no certain testing place for this experience, the color space of green dots will be dissimilar under different circumstances in the video. To precisely catch the moving green dots, we will do color space analysis to adjust data before analyzing the videos recorded under different circumstances.

Fig. 4 showed the image of reference video which is purposed to analysis the chromaticity coordinates of green point and calculate the error of video analysis. The following is the video images of similar action groups.

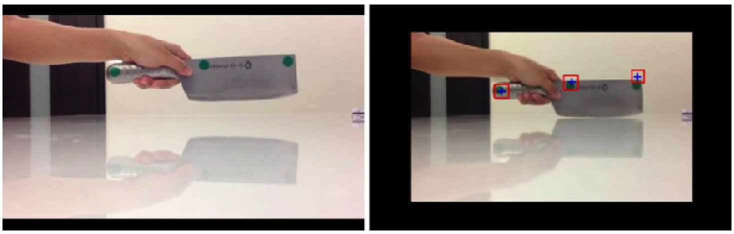

Fig. 4. Image of reference video

3.3 Video Analysis Results

Base on the correspondence of green point's locus and timeline in reference video, we can export four diagram and data of Amplitude on X Y and Z-axis, Frequency on X Y and Z-axis and variation of two-point distances. At first, we analyzed the reference video, try to correct the accuracy and calculate the error, and then analyzed the action video sets we recorded before.

According to the variation of two-point distance, divide the differences between maximum and minimum variation by two-point distance, then we can get the error of our video analysis, which is approximately 1.9%.

The following are the video analysis results of Cut action(cube slice shred dice and mince), which contained three figures, including the displacements of X-axis, displacements of Y-axis and variation of two-point distances. X-axis represented the vertical axis on the image perpendicular to countertop; Y-axis represented the horizontal axis on the image, as to Z-axis it's orthogonal to the image. We can see the amplitude differences from the first two figures. If the variations of two-point distances decrease, it means that the utensils are moving away from the camera, so we can see it represents the displacements of the Z-axis. Then it used the Fourier transformation to calculate each action's movement frequency on each of the three axis and find out the representative number of frequency. We can see the peak frequency of each action, the three axis number of peak frequency is its movement frequency.

3.4 Motion Elements Database

After video analysis, we will get the data of cooking movements. In addition to organizing amplitude and frequency of the data, we also observe the possible key differences from the action videos, and classify the representative data into tables, then precede the Decision Tree Learning analysis to get the result path.

In addition to calculating the general value of video analysis results such as amplitude and frequency; meanwhile, we also observe cooking movements to figure out the key attributes to separate two similar movements and then put them into the chart. From the analysis result of chart, we figure out that there are some observable differences in the frequency part; therefore, we use the frequency of all movements as presetting attribute in this research.

In the cutting classifications, we figure out that the displacement on the Z-axle of stripping is bigger than slicing in the same time observably. The displacement will be increasing because that the chunk-shaped ingredients are bigger than flake-shaped ones. The observed differences also reflect on the result chart of video analysis, which is the slope of amplitude on the Z-axle of the two movements. The greater the slope the more displacement per unit time, on the contrary, the smaller. At last, organizing the key data into charts and then built motion element database.

There is no significant difference in other movement classifications except for frequency; therefore, we use it on the three axles as attribute of the movement.

4 Discussion

It is important for smart kitchen systems to accurately identify the actions performed by users. It's also the core of cooking support systems. We started by analyzing the verbs occurring frequently in recipes and made classifications. A total of 119 dishes were recorded with 1,607 verbs. 46 different action items were obtained after organization. Actions that are performed by the same utensils were grouped together. We had classified 27 items, of which four items have sub-actions. Then we asked experts to refine this list, combining and modifying some action items, and finally getting a total of 26 action items, including four actions with sub-actions. These action items represent the most commonly used actions in cooking behavior. The second stage is to record the cooking action video set. Although there are already many available action video sets, since we have sorted out a list of actions in this study, we recorded them ourselves. This cooking action video set can also supply other studies with a resource for other image analysis methods.

The most important part was that our study has achieved significant results on the identification of the various actions. Based on the results, we analyzed the action's amplitude and frequency in the X, Y, Z-axis. First is the "Cut" action which occurs most often in the cooking process. Although they are all performed by a knife, they can be quite diverse, and the presentation of the ingredients is also very different. The sub-actions of Cut are Cube, Slice, Strip, Dice and Mince. According to the three amplitude diagrams of video analysis results, we can see these actions in the X-Axis and Y-Axis and the amplitude difference is quite significant. Cube is the highest, and Julienne Mince is the lowest. These differences can also be observed by eyes. Variation of two-point distances can be regarded as the displacement of the utensil movements in the Z-axis, the variation of the action obviously presents regularly decreasing, which means that action is far away from the camera. Also, we used the Fourier Transform to calculate the frequency value of each action in the three axis of the highest frequency. We are able to observe that the frequency of julienne and mince are higher than cube from the action videos and analysis data; however, the precise value of distinction depends on Decision Tree Learning. Discussing the results of machine learning from the result of decision Tree learning, we can separate Cube, Slice, Dice and Julienne precisely. The method of classification is showing in figure 5. First is the Frequency of X-axis, division value is 4.051, >4.051 for Cube and Slice, <4.051 for Dice and Julienne. Cube and slice have greater movements and lower frequency due to the bigger ingredients, while Dice and Julienne have smaller movements and higher frequency due to the smaller ingredients. From the research result, we can find out that Frequency of X-axis will be affected by the altitude of ingredients. Second is slope of Z-axis amplitude, the greater slope value means the bigger movements in the same time. Cube and slice are able to be distinguished by slope value 6.49, >6.49 for Cube, <6.49 for Slice, and the reason is that Cube has thicker ingredients, so as the movements. And Julienne and Dice are able to be distinguished by slope value 10.42. The incorrectly classified precision between Julienne and Mince is very high, so although total precision is 73.8462%, if we take of the Mince action data, total precision will increase to 96.1538%. Because it's very similar in ingredient's size and thickness between Julienne and Mince, so it's hard to define the attribute. The most obvious difference between them is the relationship of time. Julienne is the previous step before Mince; this difference can also be a reference of action recognition in smart kitchen system.

Then, the category of Press action is Crash and Mush, their movement is very similar, but there are only two sub-actions, so it only need one key attribute to distinguish them apart.

In addition to the difference of the amplitude and frequency of three-axis direction can separate the similar actions. Space and time differences are also other clues to help finishing a list of action items with all the actions corresponding to the relationship between the regional and the time listed. Therefore, we can use the system identification process in accordance with these relationships classified. First, the system can detect the location of the camera in the environment, such as near to the stove or countertop. It can first determine whether the next activity is cooking or preparing. Judging, again according to what the user came to the kitchen to use attached to the kitchen on the totem determining what kitchen utensils are used, such as round totem indicating frying, star totem indicating mixing. Finally, use the data threshold of motion element established in this study to determine the different actions performed by the same kind of utensils. With such a judgment process, the system can judge the actions of various dishes with higher precision.

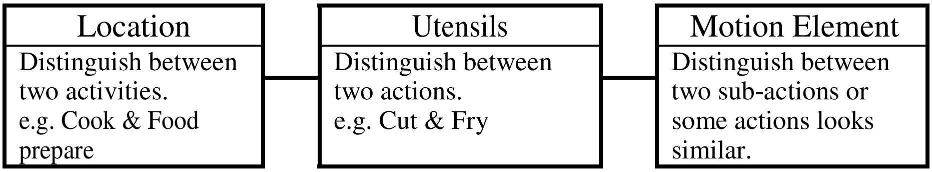

Fig. 5. System recognize process

These results can be used as the foundation for action recognition for smart kitchen systems in order to determine if the user is still running or has finished specific actions.

5 Conclusion

Smart kitchen should focus its development on the actual interaction with the user and the environmental objects, rather than emphasizing on complicated instructions and feedback. Our research's purpose is to analyze user's motion and action in the cooking process and ingredients preparation, and to discover multiple action identification characteristics for the user and cooking utensils. While many differences in cooking actions can be found by observation, how to let the system know user's action is very important for activity recognition system. Our study classified common used actions in various cooking activities.

Since the purpose of this study is to establish a set of identification process, one part of our recognize process is base on the use of different types of patterns on kitchen utensils for judgment, which are not available. This study uses green points for reference in video analysis, in addition to using the color as a reference; we can also change the use of different totems for this study.

The green point tagged on the front and near the end of utensils are to avoid the mask of hands and ingredients, it will have the range limits by using the color as the reference. If using totem for identification, we can use its characteristics of continuity to avoid masking problems. In addition, there is only a side shot with utensils, we use the size difference to analyze the z-axis data. In the future study, we can consider tagging the patterns on the top of utensils, and use two cameras for video recording. This may increase the accuracy of the z-axis data analysis, but may also increase the complication of the system analysis process.

Meanwhile, the research has also built a good database procedure to follow if there are additional actions to add or subdivide the present actions, such as thick slices and thin slices, to make this database more complete.

Acknowledgments. The authors would like to thank the National Science Council of ROC for financially supporting this research under Contract No. NSC100-2221-E-006 -204 -MY3.

References

1. De Silva, L.C., Morikawa, C., Petra, I.M.: State of the art of smart homes. Engineering Applications of Artificial Intelligence 25(7), 1313–1321 (2012), doi:10.1016/j.engappai
2. Ding, D., Cooper, R.A., Pasquina, P.F., Fici-Pasquina, L.: Sensor technology for smart homes. Research Support, U.S. Gov't, Non-P.H.S. Review. Maturitas 69(2), 131–136 (2011), doi:131-136. doi: 10.1016/j.maturitas. 2011.03.016
3. Hashimoto, A., Funatomi, N.M., Yamakata, T., Kakusho, Y., Minoh, K., Smart Kitchen, M.: Smart Kitchen: A User Centric Cooking Support System. Paper presented at the Proceedings IPMU (2008)
4. Borchers, J.: A Pattern Approach to Interaction Design. In: Gill, S. (ed.) Cognition, Communication and Interaction, pp. 114–131. Springer, London (2008)
5. Morioka, S., Ueda, H.: Cooking Support System Utilizing Built-in Cameras and Projectors. Paper presented at the The 12th IAPR Conference on Machine Vision Applications. Nara Centennial Hall, Nara (2011)
6. Miyawaki, K., Sano, M.: A Virtual Agent for a Cooking Navigation System Using Augmented Reality. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 97–103. Springer, Heidelberg (2008)
7. Chaaraoui, A.A., Climent-Pérez, P., Flórez-Revuelta, F.: A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. Expert Systems with Applications 39(12), 10873–10888 (2008), doi:10.1016/j.eswa.2012.03.005
8. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104(2-3), 90–126 (2006), doi:10.1016/j.cviu