Multiphase Activity Recognition System using Activity Scene

Tomoharu Imai^{a,*}, Kazuki Matsui^a

^aFujitsu Laboratories Ltd., 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, Kanagawa 211-8588, Japan

*Corresponding Author: imai.tomoharu@jp.fujitsu.com

Abstract

Recent years have seen an increase in user activity support services. Recognition of detailed user activities is essential to the provision services of high levels of convenience. However, this gives rise to the problem of the need for massive numbers of man-hours required to create feature values such as classifiers and rules necessary for recognition of on-site activities. We are developing primitive behavior grouping technology that can recognize activity partitions without teachers by automatically constructing feature quantities for behavior recognition. This technology presented the problem of degradation of the accuracy of recognition of activity partitions due to the effects of phenomena, such as temporary user movements or noise when applied in an actual environment in which log acquisition frequency was inhibited because of reduced power consumption. This time, we have proposed a multiphase activity recognition technology capable of highly-accurate recognition of activity partitions by defining the period of macro-granular user activities as the activity scene recognized from the state of the user surrounds and compared with activity partition information. This technology makes it possible to correct misrecognized information which was impossible to distinguish by single activity partitions and to realize improved accuracy in activity partition recognition in usage scenes that assumed an actual environment focused on white-collar users likely to exhibit changes in activities. The technology was applied to the activity recognition system focused on white-collar users and, as a result, an improvement in the activity partition recognition rate of 20% or more for our previous developed technology and performance equivalent to the conventional technologies using the most frequent value and the activity flow information were confirmed. We also confirmed that it is possible to realize recognition of activity partitions with high granularity equivalent to the schedule which could not be recognized by the

conventional technique with high accuracy.

Keywords: activity recognition, white-collar, scene detection, multiphase, Bluetooth.

1. Introduction

Recent years have seen an increase in user activity support services⁽¹⁻³⁾. Activity support services make use of smartphones, smart watches and various other sensors to recognize user activities in order to support users through the selection and provision of necessary information matched to user activities.

The realization of activity support services requires recognition of user activities, which can be broadly divided into the following two types depending on purpose and means of realization.

- General-purpose feature value base
- Specific-scene feature value base

Services using the general-purpose feature value base are furnished with a function that recognizes only activities specified in advance and provide users with services and information. Typical services under this type include information guidance services and reminder displays matched to user locations. These types of services realize activity recognition using only feature values of predefined activities.

Services making use of the specific-scene feature value base perform functions, such as detailed activity recognition and service provision that cannot be handled on a template base. For example, services such as fall detection in factories are provided by having users wear dedicated sensors and defining feature values for the detection of falls. This type of service is realized by recognizing activities by using dedicated sensors and defining feature values for each usage scene and activity⁽⁴⁻⁵⁾.

Both of the two approaches described above require the predefinition of feature values. However, defining feature

Fig. 1. Overview of Primitive Activity Grouping Technology

values for all tasks in white-collar businesses, for example, that feature large numbers of non-regular tasks, entails extremely high costs. Thus, we are in the process of developing an activity support service aimed at the white-collar segment based on an activity recognition technology different to the above two approaches⁽⁶⁾. This service that realizes activity recognition at low cost adopts what is known as “primitive activity grouping technology”. This technology is furnished with technological characteristics that enable recognition of activity partitions without elements such as feature value definitions or dedicated sensors by using sensor logs to automatically classify user activities and create feature values. Logs output from sensors or OS, for example, are converted into primitive activities known as “micro-activities” using existing recognition technology and then grouped, enabling automation of feature value definition. A technological overview of the activity recognition technique used by this service is shown in Fig.1.

Use of feature values automatically defined by primitive activity grouping technology makes it possible to create activity partition information that shows changes in user activities. For example, activity partition information in a work scene in a white-collar office is created as shown in Fig.2. Activity partition information shows the times when activities take place and group names. Information allocated with the same group name is determined to represent the same activity. Information of specific group content is allocated separately from sources such as feedback or schedule information from users.

Fig. 2. Example of Activity Partitions

2. Issues

Mobile terminals such as smartphones are battery operated and this frequently imposes severe restrictions on the use of sensors due to reduced power consumption. In such cases, logs from sensor are acquired at a rough granularity of once every few minutes. The necessary volume of information at this granularity cannot be acquired using existing activity recognition techniques, which renders application of such techniques impossible.

While primitive activity grouping technology is capable of activity recognition irrespective of the granularity, the technology presents the following problems at rough granularity.

- Degradation of recognition accuracy caused by temporary changes in activities
- Degradation of recognition accuracy caused by sensor noise

Formerly, there was a problem where a single activity in which a user is engaged may be recognized as multiple activities. For example, a user walking to write on a whiteboard during a meeting is recognized as a separate activity, with the result that a single meeting is mistakenly recognized as multiple divided meetings. During the grouping process, a set including the same micro-activity is recognized as a single group, causing occurrences of this kind of problem. Applying the process of simply deleting minor activities applied to address this phenomenon will produce problems such as erroneously deleting the act of moving between meetings, or continuous occurrence of detailed activities, rendering proper grouping impossible and giving rise to the problem of degraded recognition accuracy.

In the latter of the above two problems, temporary changes in the environment cause changes in sensor logs, unerring determination of which as noise is difficult, resulting in degraded activity recognition accuracy. For example, acquisition of logs when the condition of the radio wave targeted for sensing by a wireless sensor worsens will be misrecognized as a change in user activity. These problems are illustrated in Fig.3.

Techniques for making improvements to address problems with activity recognition can be broadly divided into the following three types.

1. Maximum frequency value acquisition
2. Creation of an activity flow in advance
3. Location recognition

In the first of the above techniques, the result of activity

Fig. 3. Example of misrecognition of Activity Partition Information

recognition output ⁽⁷⁾ at the maximum frequency during a fixed period is adopted as the result of recognition during the fixed period to improve recognition accuracy. Assuming the inclusion of misrecognition caused by temporary changes in logs and using the maximum frequency value makes it possible to exclude misrecognition. However, this technique also presents the problem of misrecognition as the same activity, even in situations where recognition as a different activity is preferable. For example, short activities such as moving from one meeting room to another meeting room nearby will be excluded as noise.

In the second of the above techniques, the content and order of user activities are predefined as an activity flow, which is used during activity recognition to improve accuracy ⁽⁸⁻⁹⁾. Activity flows make it possible to anticipate activities occurring after certain activities, thus enabling exclusion of misrecognition. However, this technique poses issues such as increased costs arising from the need to predefine activity flows and the inability to cope with undefined activities.

In the third of the above techniques, user locations are recognized and location information used to improve the accuracy of activity recognition by inhibiting misrecognition. Locations can be used to filter user activities. Recognition of locations such as companies, homes and meeting rooms makes it possible to perform tasks such as narrowing down the number of candidate activities and inhibiting misrecognition. User location recognition techniques include approaches that set up beacons on users and in facilities ⁽¹⁰⁾ and approaches that combine, for example, Bluetooth, Wireless LAN or GPS addresses and signal strength and use feedback from users to recognize locations without the need for pre-learning ⁽¹¹⁾.

These techniques require the performance of tasks such as setting beacon setup information in advance and input of location information by users, thus presenting the problem of inability to realize location recognition in cases where there are no set information or specifications from users.

3. System Architecture and Algorithm

To solve the problem of degradation of activity partition recognition accuracy in cases where granularity acquired for sensors is rough, we developed a “multiphase activity recognition technology” that uses activity scenes that can be created from Bluetooth sensor logs to inhibit misrecognition of activity partition information, thus improving accuracy. A detailed description of the aims, configuration, definitions of recognized activity scenes and algorithms of this multiphase activity recognition technology is set out below.

3.1 Aims and Configuration of the Multiphase Activity Recognition Technology

Primitive activity grouping technology converts user activities into a form called “micro-activities” and this makes it possible to create activity partition information without the need for pre-learning. Since, although available as conventional methods of correcting misrecognition of activity partition information, the three techniques described earlier are all either unsuitable for improvement of accuracy or require pre-learning, thus lacking the advantage of the technology under consideration here, it was necessary to devise a technique that requires no pre-learning.

As a technique requiring no pre-learning, we devised an approach that defines and recognizes the period of user macro-granular activities as the activity scene. For example, activities such as “meetings”, “business trips” or “document creation” registered in a schedule are specified for activity scenes.

Activity scenes are created based on logs from sensing of the user’s peripheral environment, therefore differing from activity partition information created based on sources such as personal user sensor logs.

This paper proposes a multiphase activity recognition technology that uses activity scenes that show the peripheral state to improve the accuracy of user activity partition information. This technology aims to mutually supplement activity partition information and location information to realize detailed activity recognition.

Bluetooth sensors mounted in large numbers of smartphones and capable of logging terminal information in the surrounds at low power consumption levels were used for activity scene recognition. Other Bluetooth sensor information in the surrounds can be acquired from Bluetooth sensors in the form of advising messages and activity scene recognition is performed using this log information.

The multiphase activity recognition technology is configured with two functions: an “activity scene recognition function” that recognizes user activity scenes and activity scene shifts from Bluetooth sensor log information and outputs activity scene information and a “partition comparison function” that compares activity scene information with activity partition information created using conventional technologies and deletes misrecognition of activity partition information. A technological overview of the multiphase activity recognition technique is shown in Fig.4. These two functions will be described following the section on definition of activity scenes and activity scene shifts below.

3.2 Definition of Activity Scenes and Activity Scene Shifts

Firstly, we defined activity scenes handled by the activity scene creation function and activity scene shifts. Regarding definition, we defined activity scenes “in half or more of which the same terminal was in the surrounds and the signal strength of the terminal was equivalent to or more than the threshold value” compared to the activity scene used as reference as being the same activity scene. Using this definition makes it possible to define, for example, clerical work carried out in the same office or meetings ongoing in the same meeting room as the same activity scene. For activity scene shifts, we compared the preceding activity used as reference with the current status and defined activity scene shifts as occurring when deviation from the above defined reference occurred.

There are several types of activity shift. For the purposes of the current undertaking, we classified activity scene shifts into the following four categories.

Fig. 4. Overview of Multiphase Activity Recognition Technology

1. Leaving a group
2. Joining a group
3. Accompanying members
4. Change of members

The first category specifies leaving the location of a certain group and moving to a different location. For example, leaving a meeting while it is still ongoing, or leaving one’s own seat to visit the bathroom fall into this category.

The second category specifies moving to a place where a certain group is located. For example, arriving late a meeting, or visiting a customer’s facility fall into this category.

The third category specifies engaging in activities with one or more other persons. For example, visiting a customer’s facility together with one’s superior, or heading toward a meeting room with colleagues from the same group fall into this category.

The fourth category specifies partial changes in the group of which the user concerned is currently a member. For example, movement of some members to another location during a meeting, or new people entering the location where the user concerned is located fall into this category.

3.3 Activity Scene Recognition Function

The activity scene recognition function uses two elements: “addresses” of peripheral terminals included in advising messages obtainable from Bluetooth sensors and “signal strength” to determine activity scenes and activity scene shifts. Details of this function are set out below.

(a) Activity scene learning

A technique using a decision tree using a C4.5 algorithm is used for activity scene recognition. The decision tree is a model that performs categorization with the aid of a teacher and, when used, teacher data must be prepared and new data learned when it appears. To achieve this, this function uses activity partition information for the creation of teacher data and learning.

Continuation of an activity partition for a fixed period is assumed to represent the continuation of the same activity scene. When an activity partition continues for a fixed period (e.g. 5 minutes) in absence of teacher data, the Bluetooth sensor log is learned as teacher data.

In addition, because, rather than being fixed, activity scenes are added dynamically, relearning is required when a new activity scene occurs. When a judgment output indicates that an activity scene falls into a new class,

teacher data for the new class is created taking a fixed period from the point of judgement as teacher data.

(b) Activity scene judgment

During activity judgment, firstly a date set is created using peripheral terminal “addresses” and “signal strength” from Bluetooth sensor logs as variables. Next, based on the activity scene definition, formula (1) is applied to determine whether or not the current activity scene is the same as the immediately preceding activity scene. The combination of the address and signal strength output at a certain time t is defined as ar_{ti} , and the ar_{ti} set as AR_t . Then, the class to which the AR_t set belongs is defined as $C(AR_t) = \{s1, s2, s3, \dots\}$. The class is determined applying the decision tree using the C4.5 algorithm. The decision tree teacher data is defined as T_m , the class to which ar_{ti} belongs calculated using teacher data as $D(T_m, ar_{ti})$ and the class calculated with the largest output count from the class output result as $\text{MAX}(\sum_i D(T_m, ar_{ti}))$.

$$C(AR_t) = \text{MAX} \left(\sum_i D(T_m, ar_{ti}) \right) \quad (1)$$

From the formula, the class with the largest output count among the classes determined by the decision tree is obtained. Although the class to which a certain element ar_i belongs is output from the decision tree, sensor output fluctuation cause errors in the value. This formula absorbs errors and determines whether or not the activity scenes are the same. In this way, it is possible to recognize cases, for example, of temporary movement within the same meeting room, or returning to the same room after temporary movement as the same activity scene.

(c) Determination of activity scene shifts

Depending on the activity scene determination, an activity scene history that shows changes in the activity scene within a certain time window is created. The activity scene recognition function determines into which category each activity scene shift falls from the activity scene history. The four activity scene shift categories described earlier are used for this purpose. Determination is made using the

```

if (sensor_party_num > THRESHOLD_PARTY) &&
  (sensor_num_diff > 0)
  "Accompanying members"
elseif (sensor_num_diff < THRESHOLD_DIFF_DECREASE)
  "Leave from group"
elseif (sensor_num_diff > THRESHOLD_DIFF_INCREASE)
  "Joining a group"
elseif (sensor_num_diff > prev_sensor_num_sum / 2)
  "Change of members"
end

```

Fig. 5. Formula for Determination of Activity Scene Shifts

Fig. 6. Example of Activity Scene Information

conditional formula shown in Fig.5.

The determination formula makes determination using sensor_num_diff that expresses the number of sensor differences before and after the activity, scene sensor_party_num that expresses the number of accompanying sensors and $\text{prev_sensor_num_sum}$ that expresses the total number of sensors in the preceding activity scene. This formula shows conditions for the four scenes described in section 3.2. Shift types that have been determined are added to the activity scene history, which is then output as activity scene information.

A specific example of activity scene information is illustrated in Fig.6.

3.4 Activity Scene and Activity Partition Comparison Function

The activity scene and activity partition comparison function enhances activity recognition accuracy by comparing activity scene information created by the activity scene recognition function with activity partition information and correcting activity partition information. The comparison process is performed when a shift exists in either activity scene information or activity partition information. Shift methods can be divided into the following three types and corrections are carried out matched to each type.

1. An activity partition occurs even though the activity scene is the same.
2. The activity scene has shifted, even though the activity partition is the same.
3. Changes in both

The causes of the first of the above types are assumed to fall into one of the following two scenarios.

- Temporary changes in activity during a single activity, or changes in activity due to noise
- The start of a new activity

To distinguish between the above two scenarios, the

length of the activity partition after the shift is compared with the threshold value (e.g. 2 minutes): if the result falls below the threshold value, the activity partition determined to be the former pattern is linked and, if the results exceeds the threshold value, the activity partition determined to be the latter pattern is maintained.

The following two scenarios can be assumed as two causes.

- Failure to output sufficient micro-activities, resulting in recognition of activities as being the same, even though they are different
- Changes in the surrounding status

To distinguish between the above, determination is made by checking activity scene shift information. If activity scene shift information falls into one the categories “leaving a group”, “joining a group” or accompanying members”, the former is assumed and the activity partition is added. On the other hand, if the category “change of members” applies, the latter pattern is assumed and the activity partition maintained.

In the third pattern, it is assumed that both the activity scene and the activity partition information have been correctly recognized and the activity partition is maintained.

An example of activity partition information created using this function is shown in Fig.7.

4. Evaluation

Evaluation of the multiphase activity recognition technology was performed targeting activity recognition systems using primitive activity grouping technology. A white-collar environment was taken as the usage scene and evaluated items were classified into the percentage of activities before and after application of the developed technology. In addition, evaluation by comparison with noise reduction technology was also performed.

Fig. 7. Example of created Activity Partition Information

Table 1. Evaluation Environment and Sensors used

Terminal	Nexus 5X (Android 7.1.2)
Sensors used	For activity scene recognition: Bluetooth For activity partition recognition: Wi-Fi, Display, Step Counter, OS Activity API
Sensing frequency	Bluetooth: (4 times/minute) Wi-Fi, Display, Step Counter, OS Activity API: (Once/minute)

Table 2. Activity Scenario Classification Percentages and Evaluation Results

	Conventional Technologies	Developed Technology
Scenario A	54.1%	92.3%
Scenario B	56.9%	79.9%

The white-collar activity scenes represented below were taken as the usage scenes during evaluation.

A. In an office

Clerical work at the user’s own desk – movement – meeting – giving explanations at a meeting – walking in the meeting room – movement – clerical work at the user’s own desk

B. On a business trip

Clerical work at the user’s desk – movement from the office to the station – traveling by train – meeting at the customer’s facility – traveling by train – movement from the station to the office – clerical work at the user’s own desk

During user activities in accordance with the activity scenario, evaluation was performed by classification into percentages targeting sensor logs at the time of activities. Logs acquired three times from five users were used for evaluation. The evaluation environment is detailed in Table 1. The results are shown in Table 2.

In each activity scenario, comparison with non-application of the technology confirmed classification at an accuracy 10% or higher. We were able to realize improved recognition accuracy of activity partition shift segments during classification. Thus, greater usability when applied to services that recommend information triggered by changes in activities was increased. In cases that fell into the activity shift category especially, it was confirmed that accuracy was improved by appropriate processing of grouping that had previously been misrecognized due to

Fig. 8. Example of Activity Classification

Table 3. Results of Comparative Evaluation of Noise Reduction Techniques

	Precision Ratio	Recall Ratio
Conventional technology (maximum frequency value)	60%	95%
Conventional technology (activity flow)	90%	100%
Developed technology	90%	100%

noise. Furthermore, even in cases including multiphase activities in situations such as business trips, this technology confirmed that classification could be performed high accuracy with high grain size equivalent to schedule. An example of activity classification is shown in Fig.8.

Next, we performed evaluation by comparison with noise reduction techniques from the results of conventional activity recognition. As conventional technologies, we used a noise reduction technique that used maximum frequency values ⁽⁷⁾ within a specific window and a technique that used activity flows ⁽⁸⁻⁹⁾ as techniques capable of application to activity recognition of logs with rough granularity during the current project. During evaluation, the noise reduction relevance ratio and recall ratio were used as evaluated items. The presence and absence of noise in the evaluated data were defined as positive and negative, respectively. In addition, ten data of with and without noise were prepared for the evaluation. The results of evaluation are shown in Table 3.

The results of evaluation confirmed that the multiphase activity recognition technology demonstrated a high level of performance equivalent to that of techniques using activity flows in terms of both relevance and recall ratios. Techniques using the maximum frequency value exhibited a reduced recall ratio due to recognition of data with no noise

as noise. Moreover, techniques using activity flows require the preparation of activity flows in advance. The technology presented in this paper requires no such advanced preparation and was shown to be capable of achieving high relevance and recall ratios.

5. Conclusion

We have developed a multiphase activity recognition technology to resolve the problem posed by primitive activity grouping technologies capable of ease of actual site application of support services of degraded activity recognition accuracy that occurs when logs of rough granularity are used due to limitations posed by reduced power consumption. This technology defines macro-activity periods of users nearby as activity scenes, detects changes in the user's surrounds and environment and creates activity scene information from such changes. Activity scene information was then compared with activity partition information and corrections made matched to the partition status, thus correcting activity partitions to cope with temporary activity changes and reduce noise. Furthermore, use of Bluetooth sensors capable of logging at low power consumption levels for detection of activity scenes realized improved accuracy with relevance to the actual environment maintained. This technology was applied to a white-collar activity support system and evaluated assuming usage scenes envisaging actual environments. The results of evaluation verified an improvement of 20% or more in the ratio of recognition of activity partitions. Furthermore, as a comparative evaluation, the developed technology with conventional technologies, assessment of noise and non-noise reduction verified the possibility of achieving high relevance and recall ratios, and recognition of activity partitions with large granularity equivalent to user schedule was confirmed.

This technology enables greater accuracy of activity partition information and, as a result, realization of the timely provision of services to users using user activity partition information.

This paper has presented details of studies, prototyping and evaluation envisaging white-collar environments. Since the developed technology is capable of application if other users or sensors exist in the user's surrounds, we believe that it has the potential for expansion to a wide a range of fields such as collaborative work in stores and construction sites. In the future, undertakings in the area of activity recognition assuming logs of rough granularity are planned

to address techniques for improving activity recognition accuracy during collaborative work involving multiple users.

References

- (1) Google, "Google Now",
<https://www.google.com/intl/en-GB/landing/now/>,
(accessed 2017-11-09).
- (2) Apple, "iOS - Siri - Apple",
<https://www.apple.com/ios/siri/>, (accessed
2017-11-09).
- (3) Amazon.com, "Amazon Alexa",
<https://developer.amazon.com/alexa>, (accessed
2017-11-09).
- (4) O. Anya, H. Tawfik and A. Nagara, "ContextMorph: A Model of Context-Aware Cross-Boundary Decision Support in E-Health," *Developments in eSystems Engineering (DeSE)*, 2013 Sixth International Conference on, Abu Dhabi, 2013, pp. 361-366.
- (5) H. Wu, W. Pan, X. Xiong and S. Xu, "Human activity recognition based on the combined SVM&HMM," *Information and Automation (ICIA)*, 2014 IEEE International Conference on, Hailar, 2014, pp. 219-224.
- (6) T. Imai and K. Matsui, "Extensible Activity Recognition System for Behavior Support," 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), Taipei, 2017, pp. 606-611.
- (7) D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Personal and Ubiquitous Computing*, vol. 14, no. 7, pp. 645-662, 2010.
- (8) J. Wen, R. Kang, L. Ma and V. Wang, "Operation and maintenance support resources forecast model based on support activity flow," 2010 Prognostics and System Health Management Conference, Macao, 2010, pp. 1-5.
- (9) S. Bansal, S. Khandelwal, S. Gupta and D. Goyal, "Kitchen activity recognition based on scene context," *2013 IEEE International Conference on Image Processing*, Melbourne, VIC, 2013, pp. 3461-3465.
- (10) J. Hallberg, M. Nilsson and K. Synnes, "Positioning with Bluetooth," *10th International Conference on Telecommunications, 2003. ICT 2003.*, 2003, pp. 954-958 vol.2.
doi: 10.1109/ICTEL.2003.1191568
- (11) T. Schwartz, C. Stahl, C. Müller, V. Dimitrov and H. Ji, "UbiSpot - A user trained always best positioned engine for mobile phones," *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, Kirkkonummi, 2010, pp. 1-8.