Personal Activity Logger with Hierarchical Activity Representation

Yi-Ting Tsao and Jane Yung-jen Hsu
Department of Computer Science and Information Engineering
Nation Taiwan University
Taipei,Taiwan
Email: {d97027,yjhsu}@csie.ntu.edu.tw

Abstract

Activity recognition is a key function for many context-aware applications in a smart environment. However, data collection and annotation for activity recognition is both time-consuming and costly. This paper proposes the hierarchical activity representation to enhance data reusability and introduces Personal Activity Logger (PAL), a computer aided tool with it, to reduce annotation efforts. We experimented with PAL in annotating activities within a personal space from power meters and a webcam in the office. Preliminary results show that PAL is effective in reducing the annotation efforts with only a slight loss in quality. In addition, we indicate the potential possibility to identify users from the distribution of events in their activities through the data analysis.

Introduction

Energy conservation is more and more popular because of limited resources in the world. It is also directly related to our daily expenditure. If someone travels for a long time and forgets to turn off the air conditioner, he or she will receive a high electricity bill. Even though providing saving tips is useful for reducing power consumption, it is useless in this situation. Therefore, an active way is analyzing user's activities to detect the abnormal behaviors and then providing specific services.

In recent years, activity recognition has been applied to health care (Wilson and Atkeson 2005). It detected the activities of daily living for caregivers to let them focus their attention. To achieve this goal, some researchers deployed different kinds of sensors, such as contact switches and radio frequency identifications, in the home to collect data (Fishkin et al. 2003; Frank, Mannor, and Precup 2010; Intille et al. 2006; Wilson and Atkeson 2005). Besides, some researchers deployed their sensors in the office to understand other activities (Lin et al. 2010; Oliver, Horvitz, and Garg 2004). However, there are two common problems in these two different environments. First, annotating these data manually is usually a costly, time consuming, and repeated task. Second, most of researchers directly construct a model from sensor measurements and activities. That might limit the reusability and the extensibility. In addition, one model usually requires more data to learn activities than hierarchical models.

This paper proposes a four-layered hierarchy to represent activities for resuability enhancement. Moreover, we introduce Personal Activities Logger (PAL) to record the personal activities with lower efforts and without interruptions. To evaluate the PAL's performance, we deploy power meters and a webcam within a personal space in the office. The results show that PAL could reduce the annotation efforts with only slight loss in quality. Furthermore, we could find that different users have their own characteristics while performing activities.

Related Work

To annotate data, several approaches have been proposed (Intille et al. 2003). Interviews collect the annotations from users through conversation but the weaknesses are selective recall and selective reporting bias. Gathering the annotations from users by trained observers is called direct observation. Although it avoids selective recall, it is costly, time consuming, and disruptive. Self report obtains the annotations from users reporting actively and there are two branches, recall survey and time diary. It mitigates selective recall and selective reporting bias but requests users to carry around something all day. The experience sampling method (ESM) (Scollon, Kim-Prieto, and Diener 2003), which is so-called the ecological momentary assessment (EMA), requests users to carry a timing device, such as a mobile phone. When the timing device beeps, the user are asked for an annotation. The main weaknesses are interrupting users' activities and annoying. Some modified experience sampling methods, such as the context aware experience sampling method and the image-based experience sampling method, improve them but other issues are elicited. Since there is no perfect method to collect the annotations so far, researchers try to combine different methods to deal with the above issues.

Augmented recall survey (ARS) is an off-line review tool to collect the annotations (Wilson and Atkeson 2004). The sensor measurements are divided into some episodes and then the similar episodes cluster in a group. The representative episodes in each group are converted into English text by Narrator (Wilson and Atkeson 2003). Note that Narrator uses a finite state machine to parse the data. The set of states represent words and phrases and the input is a combination of sensor measurements and time. Finally, users look at the English text to annotate their activities through the multiple choice test. Context aware recall survey (CARS) is a tool to collect the annotations through the context awareness (Wilson, Long, and Atkeson 2005; Wilson, Wyatt, and Philipose 2005). It is similar to ARS but the representative episodes are converted into a series of descriptive images instead of text. Notice that the mapping between sensor measurements and images are hand coded in advance. At last, the episodes are annotated by users with a game-like recall survey.

The above tools are preformed well to collect the annotations for sensor measurements but they do not have enough flexibility to reuse. For example, if the sensors or the focused annotations are altered, they have to annotate again. However, the sensors and the focused annotations are always changed even if in the same environment.

Methodology

In essence, activities could be separated into several events and events include some objects. Researchers introduce to representing activities with objects (Philipose et al. 2004) but this representation ignores the possibility of events. Therefore, we propose a four-layered hierarchy for representing activities, including the physical layer, the object layer, the event layer, and the activity layer. The sensor measurements are represented in the physical layer, the status of objects, including people, are represented in the object layer, the interactions between people and objects are represented in the event layer, and the intentions of a series of events are represented in the activity layer. An example in the real world is shown in Figure 1. This representation reserves the flexibility for explaining the sensor measurements and the reusability.

Researchers mention that changes, such as object changes and tempo changes, are useful to mark the boundaries between two events (Zacks and Tversky 2001). Therefore, we combine the idea of potential boundaries and the hierarchical activity representation to construct PAL, which is an interactive off-line tool to help users with recording their daily life. The details are shown in Algorithm 1. The conversion

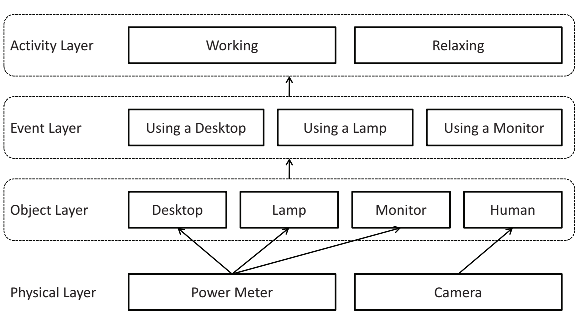

Figure 1: An example of the four-layered hierarchy for representing activities.

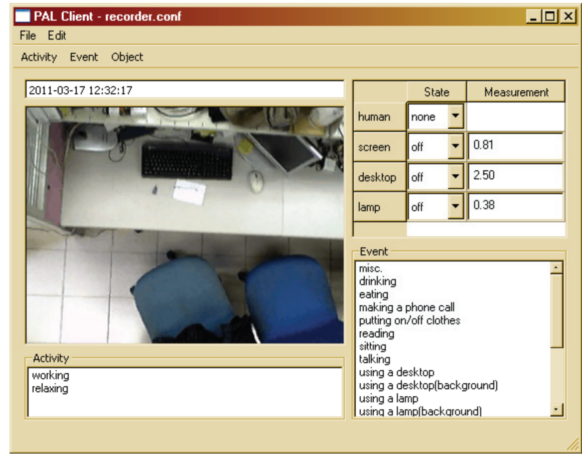

Figure 2: The user interface of PAL.

from the physical layer to the object layer is demonstrated in line 1. Collecting an event when the status of objects changes, which is shown in line 2 to line 8. Gathering an activity when the interaction between the user and objects changes, which is shown in line 9 to line 15. In addition, the user interface is shown in Figure 2. PAL displays the entire sensor measurements to request for the annotations at different layers because researchers found that the visualization of the collected data prompts users to respond (Hsieh et al. 2008). The advantage is two-fold. The first is that the focused activities might be changed by simply updating the annotations at the activity layer. The second is that the object states could be derived from different sensors in the different environments. In other words, the relation between the object layer and the event layer could be reused when the sensors or the focused activities are changed.

Algorithm 1 The algorithm of PAL

1: Convert sensor measurements into the object states O_t
2: **for all** t **do**
3:     **if** $O_t \neq O_{t-1}$ **then**
4:         Request for annotating the event E_t
5:     **else**
6:         $E_t = E_{t-1}$
7:     **end if**
8: **end for**
9: **for all** t **do**
10:     **if** $E_t \neq E_{t-1}$ **then**
11:         Request for annotating the activity A_t
12:     **else**
13:         $A_t = A_{t-1}$
14:     **end if**
15: **end for**

Experiment

The aim of the experiment is to evaluate the quality of annotations and the reduced efforts for annotating with PAL. We

Appliance	desktop*, lamp*, laptop*, monitor*
Event	drinking, eating, making a phone call, putting on/off clothes, reading, sitting, talking, using a desktop*, using a lamp*, using a laptop*, using a monitor*, using a phone, wiping off a desk, writing
Activity	working*, relaxing*

Table 1: The summary of appliances, events, and activities in the office, where * represents the selected items in our experiments.

ID	Duration	# of (used) appliances	# of events	# of activities
1	6:58:59	3(1)	156	4
2	6:45:45	3(2)	71	3
3	8:28:49	3(3)	308	4
4	3:50:42	3(2)	95	4

Table 2: The statistics of collected data.

Object	Precision	Recall
User	0.75	0.99
Desktop	1	0.99
Lamp	1	0.99
Laptop	0.99	0.99
Monitor	1	0.99

(a) At the object layer.

Action	Precision	Recall
Using a Desktop	0.96(1)	0.99(0.99)
Using a Desktop(bg)	- (1)	0(1)
Using a Lamp	0.98(1)	0.99(0.99)
Using a Lamp(bg)	- (1)	0(1)
Using a Laptop	0.96(0.99)	0.80(0.99)
Using a Laptop(bg)	0.70(1)	0.87(0.98)
Using a Monitor	0.99(1)	0.99(0.99)
Using a Monitor(bg)	- (1)	0(1)

(b) At the event layer, where bg = background; - = cannot detect.

Activity	Precision	Recall
Working	0.99(0.99)	0.96(1)
Relaxing	0.87(1)	0.98(0.99)

(c) At the activity layer.

Table 3: The results of quality evaluation at each layer, where with the perfect user's presence shown in brackets.

deploy some power meters and a webcam within a personal space in the office. Each appliance is attached to a power meter and the webcam records the videos from the top down view. Before starting to collect data, we record two videos with and without person in the same space. The use of these two videos is to construct a user's presence model. To convert the sensor measurements to the object states, we decide the user's presence from images by the pre-trained model and cluster the sensor measurements from the power meters individually. Then the means of groups are requested for annotating the states of appliances. We extend the annotations to the other sensor measurements by the nearest neighbor algorithm.

To realize the activities in the office, we ask 10 people which are between 20 and 30 years old for recording their appliances, events, and activities within their personal space in the office all day for a week. The summary is shown in Table 1. Then four studies are collected from different participants respectively and the related information is shown in Table 2. In these studies, we define that the states of objects, including the user's presence, are binary, the selected events are directly related to appliances, and the activities are working and relaxing. Furthermore, each event has its corresponding event in the background to represent that the event is triggered without the user in the space. The ground truth is manually annotated per second.

To evaluate the quality of annotations, we compare the results with the ground truth to compute the precision and recall, which are shown in Table 3. At the object layer, the precision of the user's presence is relative low because the method we adopt to detect it is naive and simple. A better algorithm could improve the performance significantly. In addition, the incorrect prediction of the state of appliances happens because of the delayed time and the failed measurements. At the action layer, the events in the background are influenced by the imperfect user's presence. At the activity layer, the imperfect user's presence also influences the precision and recall but the sustainability of activities alleviates these situations.

ID	object	event	activity
1	8	345(72)	11(70)
2	8	530(22)	14(19)
3	6	442(230)	52(59)
4	6	122(40)	36(11)

Table 4: The number of requested annotations in each layer, where with the perfect user's presence shown in brackets.

	User	Desktop	Lamp	Laptop	Monitor
UD	0.99	0.99	0.06	-	0.96
UDB	-	-	-	-	-
UL	0.99	0.99	1	1	1
ULB	-	-	-	-	-
ULT	0.99	0	0.14	0.99	0.97
ULTB	0.56	0	0	1	-
UM	0.99	0.99	0.05	1	0.99
UMB	-	-	-	-	-

(a) PAL.

	User	Desktop	Lamp	Laptop	Monitor
UD	1	0.99	0.07	-	0.99
UDB	0	1	0	-	0.03
UL	1	0.99	1	1	1
ULB	0	-	1	1	-
ULT	1	0	0.12	0.99	0.99
ULTB	0	0	0.01	1	0
UM	1	0.99	0.05	1	0.99
UMB	0	1	0	-	1

(b) PAL with the perfect user's presence.

	User	Desktop	Lamp	Laptop	Monitor
UD	1	1	0.07	-	0.99
UDB	0	1	0	-	0.03
UL	1	1	1	1	1
ULB	0	-	1	1	-
ULT	1	0	0.12	1	0.99
ULTB	0	0	0.01	1	0
UM	1	0.99	0.05	1	1
UMB	0	1	0	-	1

(c) The ground truth.

Table 5: The proportion of the objects in each event, where UD = using a desktop; UDB = using a desktop(bg); UL = using a lamp; ULB = using a lamp(bg); ULT = using a laptop; ULTB = using a laptop(bg); UM = using a monitor; UMB = using a monitor(bg);- = no instance.

To evaluate the reduced efforts for annotating, the proportion of the number of requested annotations to the number of ground truth is considered. The numbers of requested annotations are shown in Table 4. The results show that the imperfect user's presence increases the number of requested annotations at the action layer. Furthermore, some of changes at the event layer are only related to the user's presence so they might not be detected with the imperfect user's presence. Therefore, the number of requested annotations at the activity layer without the perfect user's presence is less than with it.

After the precise evaluation, we investigate the ability of the hierarchical activity representation. First, we compare the representation in PAL with and without the perfect user's presence with the ground truth to indicate the correctness of the representation. The proportion of the objects in each event is shown in Table 5. The results in Table 5b is similar to the ones in Table 5c and the results in Table 5a is also similar to Table 5c except the events in the background. The proportion of the actions in each activity is shown in Table 6. The results show that the trend is similar even though the values in Table 6a 6b 6c are not exactly the same. Furthermore, we discover that the users in these four studies usually keep their laptop running when they relax. That might be a waste behavior in the office. Second, the diversity of these four studies at the activity layer is shown in Table 7. It demonstrates that the hierarchical activity representation could express the individual difference while performing the same activity.

Conclusion

We introduce PAL, an interactive tool, to help record and annotate office activities using an object-event-activity hierarchy which enhances the data reusability. The preliminary results show that PAL could reduce the annotation effort efficiently and the quality of the user's presence influences the results directly. The reduced efforts facilitate the gathering of long term information. We plan to improve the precision of the user's presence detection and collect more data from the office. Besides, prior knowledge about the distribution of objects, events, and activities in the long term should be useful for detecting abnormal or wasteful behaviors and identifying users.

References

Fishkin, K. P.; Kautz, H.; Patterson, D.; Perkowitz, M.; and Philipose, M. 2003. Guide: Towards Understanding Daily Life via Auto-Identification and Statistical Analysis. In *Proceedings of the International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*.

Frank, J.; Mannor, S.; and Precup, D. 2010. Activity and Gait Recognition with Time-Delay Embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Hsieh, G.; Li, I.; Dey, A.; Forlizzi, J.; and Hudson, S. E. 2008. Using Visualizations to Increase Compliance in Experience Sampling. In *Proceedings of the International Conference on Ubiquitous Computing*.

Intille, S. S.; Tapia, E. M.; Rondoni, J.; Beaudin, J.; Kukla, C.; Agarwal, S.; Bao, L.; and Larson, K. 2003. Tools for Studying Behavior and Technology in Natural Settings. In *Proceedings of the International Conference on Ubiquitous Computing*.

Intille, S. S.; Larson, K.; Tapia, E. M.; Beaudin, J. S.; Kaushik, P.; Nawyn, J.; and Rockinson, R. 2006. Using a Live-In Laboratory for ccUbiquitous Computing Research. In *Proceedings of the International Conference on Pervasive Computing*.

Lin, G.; Lee, S.; Hsu, J. Y.; and Jih, W. 2010. Applying Power Meters for Appliance Recognition on the Electric Panel. In *Proceedings of the IEEE Conference on Industrial Electronics and Applications*.

Oliver, N.; Horvitz, E.; and Garg, A. 2004. Layered Representations for Learning and Inferring Office Activity from

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0.32	0	0.07	0	0.32	0.02	0.45	0
Relaxing	0.17	0	0	0	0	0.39	0.13	0

(a) PAL.

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0.40	0	0.09	0	0.48	0.04	0.56	0
Relaxing	0.26	0.07	0	0	0.05	0.44	0.25	0

(b) PAL with the perfect user's existence.

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0.40	0	0.09	0	0.48	0.04	0.56	0
Relaxing	0.26	0.07	0	0	0.05	0.44	0.25	0

(c) the ground truth

Table 6: The proportion of the events in each activity.

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0	0	0	0	0.74	0.08	0	0
Relaxing	0	0	0	0	0.08	0.74	0	0

(a) User # 1

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0	0	0	0	0.54	0.01	0.52	0
Relaxing	0	0	0	0	0	0	0	0

(b) User # 2

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0.87	0	0.07	0	0	0	0.87	0
Relaxing	0.74	0.21	0	0	0	0	0.73	0

(c) User # 3

	Using a desktop	Using a desktop(bg)	Using a lamp	Using a lamp(bg)	Using a laptop	Using a laptop(bg)	Using a monitor	Using a monitor(bg)
Working	0	0	0.26	0	0.37	0.08	0	0
Relaxing	0	0	0	0	0	0.01	0	0

(d) User # 4

Table 7: The diversity of the four studies at the activity layer.

Multiple Sensory Channels. *Computer Vision and Image Understanding* 96(2):163–180.

Philipose, M.; Fishkin, K.; Perkowitz, M.; Patterson, D.; Fox, D.; Kautz, H.; and Hahnel, D. 2004. Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing* 50–57.

Scollon, C. N.; Kim-Prieto, C.; and Diener, E. 2003. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies* 4(1):5–34.

Wilson, D., and Atkeson, C. 2003. The Narrator: A Daily Activity Summarizer Using Simple Sensors in an Instrumented Environment. In *Proceedings of the International Conference on Ubiquitous Computing*.

Wilson, D., and Atkeson, C. 2004. Automatic Health Monitoring Using Anonymous, Binary Sensors. In *Proceedings of the CHI workshop on Keeping Elders Connected*.

Wilson, D. H., and Atkeson, C. 2005. Simultaneous Tracking and Activity Recognition (STAR) Using Many Anonymous, Binary Sensors. In *Proceedings of the International Conference on Pervasive Computing*.

Wilson, D. H.; Long, A. C.; and Atkeson, C. 2005. A Context-Aware Recall Survey for Data Collection Using Ubiquitous Sensors in the Home. In *Proceedings of the International Conference on Human Factors in Computing Systems*.

Wilson, D. H.; Wyatt, D.; and Philipose, M. 2005. Using Context History for Data Collection in the Home. In *Proceedings of the International Workshop on Exploiting Context Histories in Smart Environments*.

Zacks, J. M., and Tversky, B. 2001. Event Structure in Perception and Conception. *Psychological Bulletin* 127(1):3–21.