Multimodal User Guide for Experience Sharing based on Activity Modeling

Hyoungnyoun Kim[1,2] and Ji-Hyung Park[2]

[1]Korea University of Science and Technology (UST)

[2]Intelligence and Interaction Research Center
Korea Institute of Science and Technology
Seoul, Republic of Korea
{nyoun, jhpark}@kist.re.kr

Abstract— **This paper presents a multimodal user guide system, comprised of wearable sensors, which provides interactive feedback to users for the purpose of sharing experiences. The system we propose captures and analyzes contextual information using wearable sensors, while a user is performing an activity. The information can subsequently be shared with other users and used to guide users in performing similar activities. Audiovisual feedback from the guiding system is provided to the users, in real time, in accordance with their activity state. To realize the system, we propose a probabilistic method for modeling of user experiences, and a multimodal feedback process based on the previous experience model. By applying the system to a short tour experience, we evaluated the feasibility and suggested uncovered potential future directions in regards to our proposed system.**

Keywords- experience sharing, multimodal wearable sensors, user guide system, multimodal feedback, activity modeling.

I. INTRODUCTION

Sharing personal experiences is necessary for providing or receiving the skills and knowledge of others. Previously, audiovisual information such as text, images, and videos were mainly used to transfer one's experiences. However, experience sharing would be more effective if multimodal personal information such as locations, actions, and physiological states can be utilized. This paper focuses on the synchronization of a live activity of one user with a recorded activity of another user, thereby allowing users to relive the experiences of another. The system we propose captures and analyzes contextual information while a user is performing an activity. Subsequently, this information is shared with other users, which can be used to guide them in performing the same activity.

An experience is composed of several unique and constant activity states. So if the meaning of an experience is the same, it should generate similar patterns within a certain time range, regardless of variation. For example, when people dine at a restaurant, every experience is similar despite idiosyncratic variations in typical activities such as ordering foods, drinking beverages, and paying the bill. We define the typical patterns as *activity states*. In this paper, we use an unsupervised modeling process to extract the activity states that represent a particular experience. Via multimodal wearable sensors, the raw data of each user activity is acquired and used to build a hierarchical structure of a particular experience.

After modeling a user's experience using a combination of activity patterns, the proposed guiding system predicts the next activity from a current user state by using modeled activity transition patterns. When the activity state of a subsequent user is different from the predicted state of the model, our system assists the user in correcting his/her state through guidance via audiovisual feedback. This system would be an example of 'human in the loop'.

Previous research work has already suggested representing information based on contextual sensor information [1]. In one study, user feedback was provided through multimodal sensors according to user context in a museum environment [2]. Recent location based systems have had the similar ideas and goals as our system [3][4]. Many researchers have attempted to model the daily life of a person [5][6] and discover their abnormal activities [7]. Despite these previous works, this paper presents two contributions. First, we propose a probabilistic and automatic modeling method of user experiences using a hierarchical activity structure. Second, we investigate the model with user activities in real-time to provide audiovisual feedback to the user depending on his/her current activity state and the previous experience model.

As an example, we experimented with a short peregrination experience. A user travelled along a specific route with several wearable sensors, and after modeling the experience, another user followed the same route, receiving guidance from a mobile system. The models used were based on GPS and acceleration data, recorded using wearable sensor devices.

This paper is organized as follows. In Section 2, the activity states modeling process from multimodal sensor data is described. Section 3 explains how the created activity model is applied to the system to analyze a user's activity and guide his/her behavior. Lastly, Section 4 presents an experiment, in which the proposed system models a user's peregrination experience and provides active feedback to guide another user in the same experience.

II. ACTIVITY STATE MODELING

An activity model that represents our daily lives has several hierarchical layers with their own scales. In order to model a user's activity, the hierarchical structure of an experience should be considered. Higher-level activities such as housework, shopping, eating, touring, and travelling contain contextually meaningful situations with varying degrees of continuity and consistency within an experience. Each higher-level activity consists of various temporal combinations of lower-level activities such as walking, running, conversation, or grasping. The lower-level activities are composed of multiple contextual primitives, which include categories of time, location, specific people, and an action. Each primitive is detected from sensors on a user's body. In a higher-level activity, combinations and the order of lower-level activities vary for each trial and are dependent on the users. In this confusing circumstance, however, we assume that the variation of activities is represented by the transition probability table. Thereupon, we describe how to model the activity transition probability and to find stable activity states in an experience.

User activity modeling begins with extracting contextual primitive information such as locations, speeds, actions, and environmental information from multimodal sensor data. A user wears several sensors, which include a GPS, a camera, a microphone, and accelerometers. The sensor data is analyzed by signal processing techniques to obtain multi-dimensional primitives. This multivariate primitive data is logged sequentially during an experience and is structured for a hierarchical activity model.

As we mentioned above, variations in a user's activity will have certain patterns, so the same primitives would be acquired several times if a user were in a similar state. In order to make a state transition graph and detect stable states in an experience, we use primitives that have temporal, categorical, and combinatorial characteristics.

A clustering process is applied to extract activity states that have common properties among the primitive data set. The entropy-based Quality Threshold clustering method (QT clustering) [5][8] is adaptable to cluster this type of data. Since activity modeling operates autonomously, the system has no information about the number of clusters and exact range of clusters. In this case, QT clustering can deal with the ambiguity in data sets and exclude noisy data that the system fails to cluster. QT clustering applied in this system is based on entropy calculation because the primitives acquired are categorical - primitives such as location and object are not labeled numbers but categories (e.g. store, toilet, bus, fork, pen, etc).

After this step, the multivariate primitives are determined with labeled state numbers. These labels are represented as a state array. The state array is basis for detecting ranges of stable states and building a state transition graph. By discerning the temporal relation between adjacent states in the array, the transition probability table is created and activity modeling is accomplished [6]. Fig. 1 shows an example of the state transition graph of a specific experience from the clustered activity state array. The generated model has both a probability distribution of each state and transition probabilities to other states. One the other hand, we also extract stable state ranges from the state array. If a certain state continues for a considerable amount of time, we assign the range for the state as stable. Since the stable states are

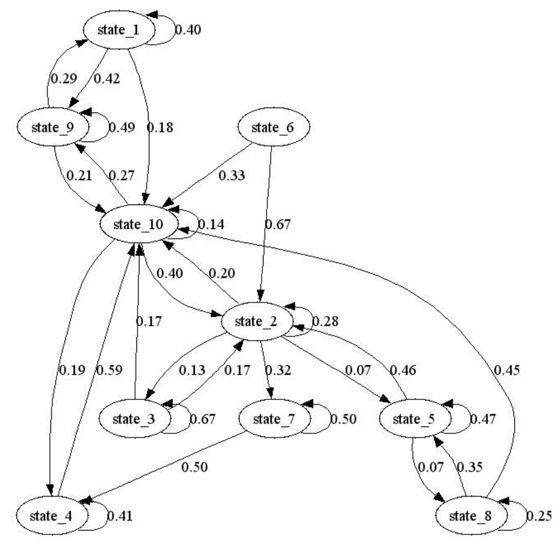


Figure 1. An example of state transition graph. The probabilities and nodes are calculated from temporal relation between adjacent labels of states in a state array

meaningful in an experience, multimodal user feedback is executed only in those ranges.

III. MULTIMODAL USER FEEDBACK

When sharing a created experience model, our system provides active multimodal feedback to allow users to have the same experience as others by comparing his/her current activity with that of a previous model. The overall feedback process is represented in Fig. 2.

During the process of observing a user's current activity, primitives such as user contexts are obtained from multimodal wearable sensors in real time. Based on a clustering threshold used in activity state modeling, the most

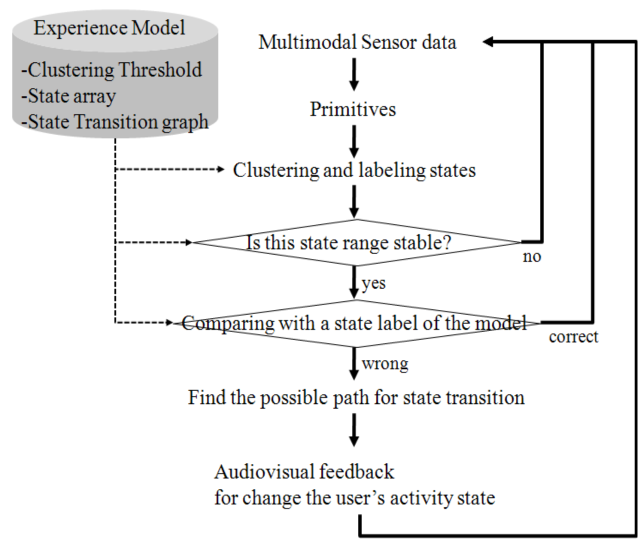


Figure 2. The real time feedback process based on the experience model for experience sharing.

probable state of a current activity is determined by the comparison of entropy variation as shown in equation (1).

$$i = \underset{i}{\operatorname{argmin}}\{H(C_i) - H(C_i + E)\} \tag{1}$$

Here, i denotes the ith index among the total states, Ci denotes the instance set of the ith state, E denotes the current instance, and H denotes the value of entropy.

After extracting a current state, the system validates whether the inferred state is in a stable state range. When the certified current state is different from the indicated state of the model, the system recognizes that the user is in an inadequate situation of the experience. Then, a guide process instructs the user to modify his/her state in order to be consistent with the corresponding state of the model. Using Dijkstra's algorithm [9], the optimal path to the target state is calculated from a state transition graph like Fig. 1. To consider the continuity of a user's activity and minimize behavioral resistance, the guide system finds the optimal transition path that allows a user to change his/her action smoothly. In the experience model, both Markov chain and transition probability are generated between activity states. In order to calculate the optimal path to transit, a cost function is used as shown in equation (2).

$$cost(u,v) = -\log\{p(u,v)\} \tag{2}$$

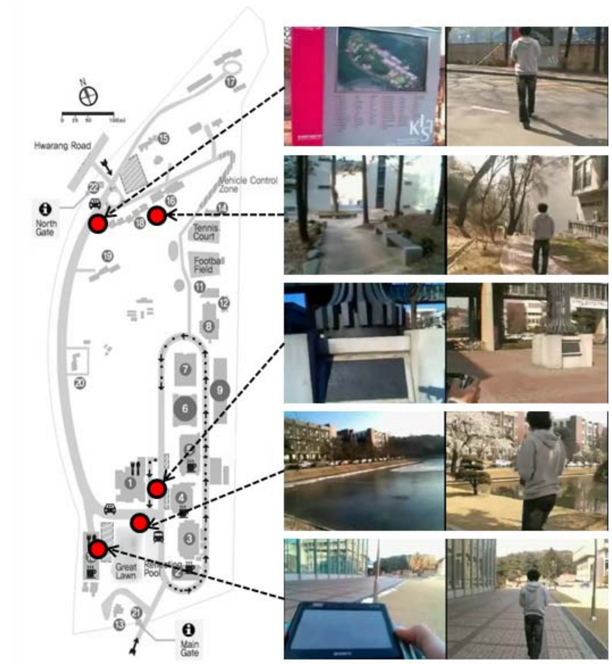
Here, u and v denote a source state and a destination state, respectively.

In order to prompt a user to change the current activity state, the guiding system provides feedback to him/her through verbal and visual information. Since the properties of two states can be represented by multivariate primitives, feedback is based on the difference between the primitives. For example, if the current user's state were walking instead of the relevant stable state, which is standing, the verbal feedback would be 'slow down your pace'.

IV. Experimental Result

To evaluate our multimodal user guiding system, we experimented with a short tour experience. The user wore a GPS, accelerometers, a microphone, and a camera and travelled along a given path as shown in Fig. 3. The sensors were connected to a mobile PC via Bluetooth. From these multimodal sensors, three primitives were logged every four seconds and processed into an activity model according to the method in section 2. The duration of the experience was 15 minutes. After the activity model was created, another user had the same experience. Our system guided the user by providing audiovisual feedback through the speaker and the monitor of a mobile PC.

The tour experience was analyzed via activity state modeling. Speed and direction information were acquired from the GPS, and action information such as walking, standing, and speed walking was extracted by analyzing data from the accelerometers. The audiovisual information of a camera and a microphone were not used because of the



(a) (b) (c)

Figure 3. An experimental tour scenario (a) a campus map. Red circles indicate major places in the scenario. (b) captured images from an experience model creator. (c) captured images from a user sharing the tour experience

processing complexity. They were mainly used to monitor a user's states. After clustering the multivariate primitives, a state array was generated. From the state array, stable state ranges were determined and a transition probability table of states was generated. Fig. 4 represents the result of the experience modeling. The tour experience had a total of 10 activity states. Four states and their stable ranges would be a basis for comparing with the next user's activities (referring Table 1).

The generated tour model was applied to allow another user to experience the exact same tour. In the model, the system checked the new user's routes by monitoring both the GPS data and total distance from the starting position to recognize whether the user's location was in a stable range. When the states between two users were different at a certain place during stable state ranges, appropriate verbal feedback appeared to warn the user and provide appropriate guidance to change the activity (see Table 1). In Table 1, the primitives of (ls, ms, w, d, s) denote (low speed, middle speed, walking, down stairs, standing), respectively. Verbal user feedback was used to change user's activity states, and visual user feedback was used to provide information of tour locations.

V. Discussion and Conclusion

This paper presented an interactive user guiding system comprised of wearable multimodal sensors, which provides multimodal feedback to users in order to improve experience sharing. The system uses an automatically inferred activity

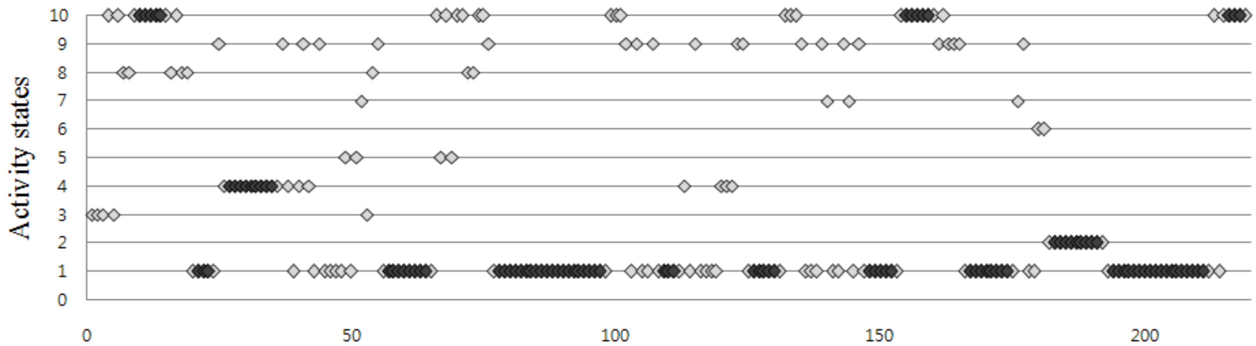


Figure 4. A result of the tour experience modeling. Row axis indicates the time of the tour, and column axis denotes activity states. Diamonds represent clustered labels according to the progress of the experience, and its temporal relation consists of a state array. Gray diamonds mean analyzed states of the user activity, and black diamonds mean the areas of stable states.

TABLE I. VERBAL USER FEEDBACKS ACCORDING TO THE DIFFERENCE OF USER STATES IN SHARING EXPERIENCE

Stable states	**Main primitives**	**Currently unmatched**		**Verbal user feedback**
		primitives	***state transition path***	
1	ls,w,ls,w	ms,w,ms,w	4	'slow down your pace'
			4–10–9–1	
2	ls,d,ls,d	ls,w,ls,w	1	'down stairs'
			1-10-2	
4	ms,w,ms,w	ls,w,ls,w	1	'quicken your pace'
			1-10-4	
10	ls,s,ls,s	ls,w,ls,w	1	'slow down your pace'
			1-10	

model of a user performing a task to guide a different user in imitating the same task. By applying the system to a short tour experience, we tested the feasibility of our proposed system.

This work is a prototype and needs additional considerations in terms of usability. The audio-based guiding messages can be irritating for some people. Audiovisual information of user experiences is not used yet. In addition, contextual information of sensors such as a camera or a microphone is hardly changed according to the will of the user because that information is related to the environment of the user, not the user's own behaviors. However, the system can provide feedback to users not only based on a location but also through other contextual information such as action, time, and objects. In the case of visiting strange places, cooking, or trying something new, our system will reduce trial-and-error behavior and help users to perform optimal activity routines.

REFERENCES

[1] J. Pascoe, "The stick-e note architecture: extending the interface beyond the user", In IUI Conference Proceedings, 1997.

[2] F. Bohnert, I. Zukerman, S. Berkovsky, T. Baldwin, and L. Sonenberg, "Using interest and transition models to predict visitor locations in museums", AI Communications, 21(2-3):195-202, 2008.

[3] K. Cheverst, N. Davies, K. Mitchell, and A. Friday, "Experiences of developing and deploying a context-aware tourist guide: the GUIDE project", Proceedings of the 6th annual intl. conf. on Mobile computing and networking, 2000.

[4] J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh, "Who's viewed you?: the impact of feedback in a mobile location-sharing application", In CHI'08 Conference Proceedings, 2008.

[5] H. Kim, I. J. Kim, H. G. Kim, and J. H. Park, "Adaptive modeling of a user's daily life with a wearable sensor network", In ISM'08 Conference Proceedings, 2008.

[6] A. Krause, D. P. Siewiorek, A. Smailagic, and J. Farringdon, "Unsupervised, dynamic identification of physiological and activity context in wearable computing", In ISWC'03 Conference Proceedings, 2003.

[7] J. Yin, Q. Yang, J. J. Pan, "Sensor-based abnormal human-activity detection", IEEE Trans. on Knowledge and Data Engineering, 20(8):1082-1090, 2008.

[8] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data: identification and analysis of coexpressed genes", Genome Res., 9: 1106-1115, 1999.

[9] E. W. Dijkstra, "A note on two problems in connexion with graphs", Numerische Mathematik, 1: 269-271, 1959.