ISBN: 978-972-8924-68-3 © 2008 IADIS

AN APPROACH FOR AMBIENT COMMUNICATION BY DETECTING REAL-WORLD ACTIVITIES FROM ENVIRONMENTAL SOUND CUES

Mostafa Al Masum Shaikh
Department of Information & Communication Engineering, University of Tokyo, Japan
7-3-1 Hongo, Bunkyu-ku, Tokyo 113-8656

Helmut Prendinger
Digital Content & Media Sci. Research National Institute of Informatics (NII), Japan
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430

Keikichi Hirose
Department of Information & Communication Engineering, University of Tokyo, Japan
7-3-1 Hongo, Bunkyu-ku, Tokyo 113-8656

Ishizuka Mitsuru
Department of Information & Communication Engineering, University of Tokyo, Japan
7-3-1 Hongo, Bunkyu-ku, Tokyo 113-8656

ABSTRACT

In the modern world where people are usually busier than ever, family members are geographically relocated due to globalization of companies and humans are inundated with more information than they can process, ambient communications through mobile media or internet based communication can provide rich social connections to friends and family. People can stay connected to their loving ones that they care about by sharing awareness information in a passive way and even simulate real-world living in virtual worlds. For users who wish to have a persistent existence in a virtual world – to let their friends know about their current activity or to inform their caretakers – new technology is needed. Research that aims to bridge real life and these virtual worlds to simulate virtual living, while challenging and promising, is currently rare. Most existing works focus on the dynamic representation of inanimate and passive objects (e.g., buildings, cars etc.) inside virtual worlds. Only very recently the mapping of real-world activities to virtual worlds has been attempted by processing multiple sensors data along with inference logic for real-world activities. Detecting or inferring human activity using such simple sensor data is often inaccurate and insufficient. Hence, this paper proposes to infer human activity from environmental sound cues and common sense knowledge, which is an inexpensive alternative to other sensors (e.g., accelerometers) based approach. Because of their ubiquity, we believe that mobile phones or hand-held devices (HHD) are ideal channels to achieve a seamless integration between the physical and virtual worlds. Therefore, we present the approach of a prototype integrating mobile phone based computing and Second Life by inferring activities from environmental sound cues. To the best of our knowledge, this system pioneers the use of environmental sound based activity recognition in mobile computing to reflect one's real-world activity in virtual worlds.

KEYWORDS

Ambient Communication, Life Logging, Virtual World, Activity Recognition, Sound Cues, Auditory Scene Analysis

1. MOTIVATION AND INTRODUCTION

We envision that with the proliferation of computing power of hand held devices (HHD), availability of internet connectivity and improvements in communication technologies ambient communication will find a universal place at our daily life and allow us to create a vivid and intelligent online social network. Let's consider the following two scenarios.

Scenario 1: Arvind family (Mr. and Mrs. Arvind) lives in a rural place of Delhi, India. They have three

sons living overseas, one in Texas, another in Ottawa and the youngest one in Bonn of Germany. The Arvind family is now at their age of over 50 and Mr. Arvind had a massive heart operation last year. Mrs. Arvind is also ailing from several sickness like diabetics, high blood pressure etc. The three sons are always worried regarding the well beings of their parents and consequently they often talk to their parents over the phones to know their whereabouts. Though calling to India from USA, Canada, and Germany is cheaper now-a-days, but having a phone conversation with their parents is not always possible due to various reasons, for example, due to inconvenience in time differences (i.e., when it is 10 am in Delhi it is 12:30 am in Texas, 1:00 am in Ottawa and 6:30 am in Bonn) when the sons have convenient time to call, their parents are usually sleeping or resting and vise versa. But they are often worried to know at least how their parents are doing everyday. Therefore, let's imagine that Arvind family has installed an inexpensive system capable of doing the following that may provide mental peace to the sons by providing ambient communications through the internet. The system makes life logging of daily activities by detecting and recognizing sound cues from the surrounding environments and sends email message(s) to their sons reporting their daily life-sketch. In this case, an example email message as following might be very relieving to the sons. *"Your parents woke up at 7:30 am today and they had breakfast around 8:15 in the morning. They watched TV several times in the day. Went out of home for two times and walked in the roads and parks. They took lunch and dinner at around 2 PM and 8 PM. Your mother went to toilet for 5 times and father went to toilet 6 times in a day. They had communicated with each other or other people by talking. It seems they are doing fine."*

Scenario 2: Sami, Anny, Harry, and Silvia have become friends on a social network but they live at different corners of the world. All of them have their own virtual worlds and they frequently update their status to let others know what they are doing just for fun. They are using "Second Life" [1] which is a very popular online 3D-graphical representations of real (and fictitious) places inhabited by real people in the form of personal avatars in a virtual world. They want to use iPhone to automate the status update. Let's assume that on the iPhone they have installed our system that can capture and allow processing of environmental sounds at some time interval. The processed sound cues are used together with common sense knowledge to infer the present activity and automatically reflects the real-world activity on the virtual world. For example, while Silvia is cooking in real-world (i.e., she generates sound cues as in Figure 1), her friends see her moving around the kitchen in the Second Life, as in Figure 2.

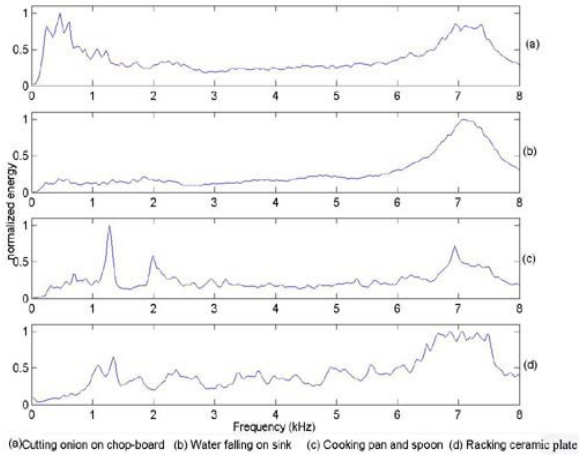

Figure 1. Example sound cues to infer cooking

Figure 2. Silvia is also 'cooking' in Second Life

Second Life (SL) is one of the most popular Internet-based virtual worlds in terms of subscribers with over 500,000 active users. Virtual world simulators represent one of the upcoming successful niches for online services for entertainment, advertisement, business, but also health and elderly care. However, in the usual interaction model, there is still a gap (between the virtual world and the real one) for users who wish to have a persistent existence in a virtual world to let their friends know about their current activity or to inform their caretakers. Specifically speaking, a user cannot automatically mirror/reflect his current movements, activities or surrounding environment (e.g., park, shopping mall, etc.) in his real life to the virtual life of his avatar. Only very recently [2] the mapping of real-world activities to virtual worlds has been attempted by processing multiple sensors data along with inference logic for real-world activities. Detecting or inferring

human activity using such simple sensor data is often inaccurate and insufficient. Moreover deploying a sophisticated ubiquitous sensor network at outdoor environment is often expensive and not feasible.

Our vision is to bridge the current separation between real worlds and virtual worlds by detecting real world activities (e.g., laughing, talking, traveling, cooking, sleeping, etc.) and situational aspects of the person's (e.g., inside a train, at a park, at home, at school, etc.). We focus how to perform daily life logging and to provide a virtual representation of humans in terms of activity and their surrounding by inferring human-activity from environmental and object-interaction related sound cues as well as common sense knowledge. We have collected 114 types of acoustic sounds that are usually produced during object interaction (e.g., cooking pan jingling sound while cooking) or by the environment itself (e.g., bus/car passing sound while on a road) or by a deliberate action of a person (e.g., laughing, speaking). These sample sounds serve as the underlying clues to infer a person's activity and his surroundings with the help of common sense knowledge (e.g., while the system identifies cooking pan's jingling and chopping board sound as consecutive cues and system's local time indicates evening then from common sense database the system infers this activity as 'cooking').

2. SYSTEM ARCHITECTURE

Figure 3 serves as the top-level pipelined architecture of the system. Because of their ubiquity we plan to use mobile phones (e.g., iPhone) to deploy this application that will capture environmental sound at some intervals to be processed. According to Figure 3, a mixed signal is passed through a robust signal processing and sound classes are detected by trained HMM classifiers. Based on the detected sounds and common sense knowledge regarding human activity, object interaction, ontology of human life (e.g., daily life of a student, or a salary man etc.) and temporal information (e.g., morning, noon etc.) are applied to infer both the activity and the surrounding of the person. This information is sent to the client application of SL as a text message and this message is parsed by SL client to map the activity of the person after rendering an appropriate virtual world for that activity.

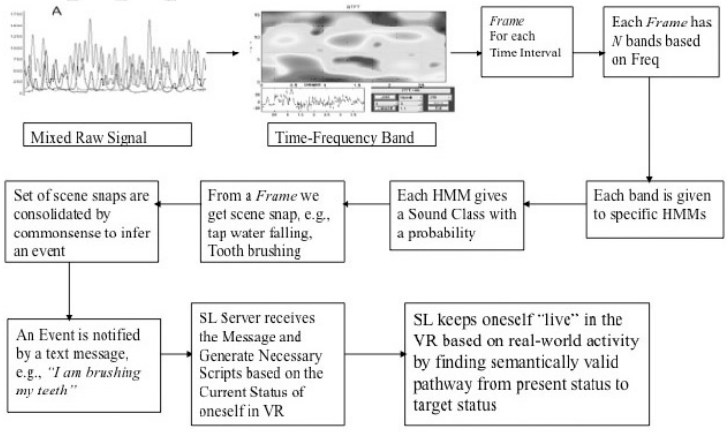

Figure 3. Brief System Architecture

2.1 Development Challenges

In order to develop this system we have the following challenges and concerns:

- Environmental sound corpus and features

For simplicity our collected sound corpus is limited to sound cues related to certain genres like, cooking, bathroom activity, human physiological action, outdoor, home, etc. Since the recognition rate is the only performance measure of such system, it is hard to judge how suitable the selected feature is. To solve this problem, it is essential to analyze the feature itself, and measure how good the feature itself is. At present we are considering mel-frequency cepstral coefficients (MFCC) feature set in this task.

- Accuracy of Acoustic Event Detection/Classification

Microphones that capture sounds with sufficient fidelity for automated processing are also much cheaper in

comparison with other sensors. Acoustics-based behavioral understanding systems work from a distance: they do not constrain subjects by requiring them to wear special devices; a prerequisite for individuals with dementia. Computational auditory scene analysis (CASA)[3], the understanding of events through processing environmental sounds and human vocalizations, has become an increasing important research area.

- Computing power of iPhone or Hand Held Devices

The computing potentials of HHDs are increasing in a rapid manner with respect to memory and incorporation of full-fledged operating system but they are still inferior to personal computer comparing to the speed of data processing. Running of the core signal processing task requires high-end computing and therefore the processing should be done by means of external devices connected to the HHDs through the Bluetooth radio. Therefore in this case a light process to capture environmental sounds after some intervals will be running on the HHDs to be transmitted for activity detection.

- Privacy and Social Issues

To manage their exposure, users should be able to disconnect from the virtual worlds at any time or be able to interrupt the sound capturing of their activities. The actual representation of users in the virtual world may be disclosed according to pre-configured user policies or users list. Additional privacy issues are related to the collection of environmental data by means of people carrying devices to different places and data collected about them. We consider important privacy and security challenges related to people-centric sensing.

- Common sense knowledge base

What this system try to create is a sense of presence that allows users to “see and feel” a remote place. From the analysis of captured signals the system gets aware of the objects available within the vicinity of the person carry the HHD. This object level information is mapped with a common sense knowledge base to infer the activity. For example, if the system detects that the sound cues represent frying pan, sink water, and chop board from consecutive input samples the developed common sense knowledge can infer a cooking activity. Moreover to minimize error in inference we incorporate temporal and ontology of daily life of an individual (e.g., student, salary man, old people etc.). For example, if system detects sound cues that initially inferred that the person is in the road, but according to the temporal information if it is at night 2:00 am on Monday and from the daily life ontology of a salary man (i.e., the user is a salary man in real world) the time frame indicates “resting”, the system may either show proactive behavior or ignore the initial inference.

- Scalability of the solution

Our default approach is to run activity recognition algorithms on the mobile device to decrease communication costs and also to reduce the computational burden on the server. However, we recognize that signal processing based classification algorithms may be too computationally intensive for handheld devices and propose to run the classifier on a back-end server in this case. This may be particularly appropriate during the training phase of a classification model.

3. CONCLUSION

Software like Google Earth and Microsoft Virtual Earth map the real world with great accuracy in terms of geographical locations and local features. We believe that the next step is to enable a user to represent real world activities in this kind of virtual world whereby personal avatars will be able to reflect what the real persons are doing in the real world. This type of application is a source of fun for young generation while it has lots of potential regarding virtual shopping mall in e-commerce context, easy monitoring or life logging for elderly people for the caregivers etc.

REFERENCES

LindenResearchInc. SecondLife. http://secondlife.com/

Mirco, M., Emiliano, M., Nicholas D. L., Shane B. E., Tanzeem, C., Andrew T. C., 2008. The Second Life of a Sensor: Integrating Real-world Experience in Virtual Worlds using Mobile Phones. In Proceedings of the Fifth Workshop on Embedded Networked Sensors (Charlottesville, Virginia, June 2-3, 2008). HotEmNets 2008.

Wang, D. Brown, G., 2006. Computational Auditory Scene Analysis: Principles, Algorithms and Applications, Wiley-IEEE Press.