MoMM 2012

David Taniar, Eric Pardede, Matthias Steinbauer, Ismail Khalil (eds.)

The 10th International Conference on Advances in Mobile Computing & Multimedia (MoMM2012)

Proceedings

3-5 December 2012
Sanur Paradise Plaza Convention Center
Bali, Indonesia

Association for Computing Machinery

International Organization for Information Integration and Web-based Applications & Services
www.iiwas.org

David Taniar, Eric Pardede, Matthias Steinbauer, Ismail Khalil (eds.)

MoMM2012

The 10th International Conference on Advances in Mobile Computing and Multimedia

December 3 - 5, 2012
Bali, Indonesia

Association for
Computing Machinery

A System for Visualizing Sound Source using Augmented Reality

Ruiwei SHEN
Kobe University
Kobe, Japan
nogamijoe@gmail.com

Tsutomu TERADA
Kobe University
PRESTO, JST
Kobe, Japan
tsutomu@eedept.kobe-u.ac.jp

Masahiko TSUKAMOTO
Kobe University
Kobe, Japan
tuka@kobe-u.ac.jp

ABSTRACT

In recent years, Augmented Reality(AR) has been widely used in various research. AR augments human sensation and offers information to support users in daily life. In this research, we propose an AR system that recognizes sound sources to augment human's vision. In our system, a sound source and its position are detected by acoustic processing. The system notifies a user through different visual markers, which are allocated on different kinds of sound sources. Our system detects different kinds of sound sources in daily life and notifies user of the information and the position of sound source. Using our system, the user can find out which object is making sound in the environment without hearing it.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous

General Terms

Sound Visualization

Keywords

Augmented Reality, sound recognition, hear-impaired,wearable computing

1. INTRODUCTION

Environmental sound is one of the most important sources from which we receive information[1]. Humans can grasp what is happening in the surrounding environment through environmental sound. For instance, we can realize whether there is an automobile approaching us through the sound of the engine. Nowadays, there is a lot of research on how to use computers to recognize sound sources automatically. Sound source recognition is one of the methods for distinguishing sound in the environment through analyzing and comparing sound data with learnt data that is created before hand, and the result of recognition is shown in a text or image format. However, for the hearing impaired, detection and presentation of the position of a sound source are important for showing a user the kind of sound source or the sound of danger that is detected. Because of the muffling of automobiles and other machines, environmental sound is hard to recognize, especially for eld people and the hearing impaired who may not aware of the sound of danger around them. The detection and visualization of sound can help them to find out the sound source in the surrounding environment. Therefore, an intuitive interface is required for the system used for recognizing sound sources.

In recent years, augmented reality (AR) has been used in various fields. AR is a kind of technique that can provide different kinds of information to a user in a real-world environment by augmenting the user's sensation. Accordingly, AR can be an ideal intuitive interface for showing a user the result of recognition. Therefore, in our research, we proposed a system that detects environmental sound by using the algorithm of sound recognition, recognizes the direction of a sound source by using a microphone array, identifies the sound source through an AR marker that is attached on the sound source object, and finally shows information on the sound source in the view of the user. Therefore, a user can find out the sound source without hearing it. Our system can aid in recognizing unusual sounds during work, notifying the hearing impaired of the position of the unusual sounds and so on.

This article is made up of six parts: reference research in section 2, introduction of system design in section 3, system implementation in section 4, evaluation and consideration in section 5, and finally, the conclusion of this article in section 6.

2. REFERENCES

Sound occurs by the oscillation of objects, and it is difficult for human to observe, but we know that sound can be described in waves on graphs, which are calculated through trigonometric functions. However, such visualization works for those professionals who have studied sound, but common users can hardly understand the meaning of such kinds of visualization.

Nowadays there is a lot of research on sound visualization. For instance, in Azar's research, in order to find out the most easy way to understand how to visualize sound, several methods of sound visualization are discussed[2]. In

MoMM2012, 3-5 December, 2012, Bali, Indonesia.

this research, they discuss the methods of sound visualization on environmental sound and speech, and the results of visualization were displayed in one form of the system. They considered seven kinds of methods to visualize environmental sound. Meanwhile, all of these results of visualization were displayed in graphs by using MATLAB. They also used text, icons, and American Sign Language (ASL) with text for three kinds of methods to visualize speech. According to the result of the examination, they figured out that ASL is the most easy way to understand to visualize sound. However, ASL is only for those who have learnt it; moreover, it is hard to use only one ASL picture to explain all words. In Bertram's research, they used particles to visualize sound source positions in Virtual Reality (VR)[3], though it was hard to show the user exactly where the sound source was.

In Kaper's research, they addressed digital sound synthesis in the context of Digital Instrument for Additive Sound Synthesis (DIASS) and sound visualization in a virtual-reality environment by means of M4CAVE[8]. In Kaper's project they visualized sound in a room-sized virtual-reality environment. But when there were no speech sounds, it was still hard for humans to recognize when sound transferred into particles in a VR environment.

In Valin's research, they presented a robust sound source localization method in three-dimensional space that uses an array of eight microphones. The method is based on time delay of arrival estimation[5]. As result, they found that not all sound types are equally well detected, and the system functioned properly up to a distance between 3 and 5 meters. However, we proposed a system that can not only recognize the direction of sound objects but also show a user exactly where the sound source is.

3. SYSTEM DESIGN

In our research, we aim to design a system that can notify a user of environmental sound without he or she listening. Furthermore, the user can avoid the dangers through the detection and notification of danger sounds. In this section, we will first introduce the construction of our system, and then we will explain how to recognize a sound source. Finally, we will introduce the sound source visualization system that uses AR.

3.1 System Construction

The construction of the proposed system is shown in figure 1. The system is made up with input, process, and output parts. In the input part, we use a web camera to capture video in real-time from the view of a user and use a microphone array to pick up environmental sounds and detect sound source positions. In the process part, we use a PC to process the data, which is gotten from the input part in real-time, and transmit the result of the calculation to the output part. In the output part, we use AR to show the user the visualization of sound of 3D objects in the real-world environment through a head-mounted display (HMD).

3.2 Sound Source Recognition Function

In our research, we use mel-frequency cepstral coefficient (MFCC) to calculate the features of environmental sound and recognize sound sources. MFCC is a common algorithm for calculating the features of sound clips, which approximates the human auditory system's response. Moreover,

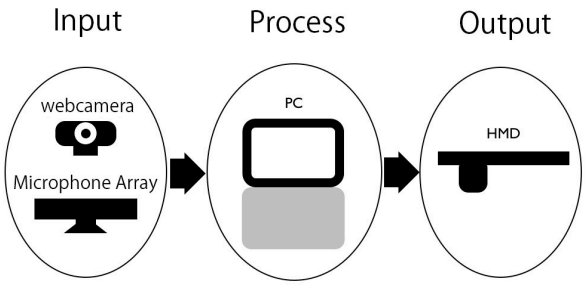

Figure 1: System construction

the recognition accuracy is higher than with mel-frequency cepstrum (MFC)[6]. It is widely used in speech recognition system and voiceprint recognition system. The MFCC calculates the sound features in the following four steps.

1. Use Fast Fourier transform (FFT) to calculate the spectrum of the sound in each window.
2. Map the spectrum onto the mel-scale by using triangular overlapping windows.
3. Take a log of the mapping result.
4. Take the discrete cosine transform after taking the log.

Through these steps, we calculate and label the features of environmental sound samples and build a database of features. When a user uses our system, our system uses the MFCC to calculate the features of environmental sound in real time and compare the features with the features in the database by using Euclidean distance. Finally, the nearest one is the result of recognition.

3.3 Augmented Reality Function

The system we proposed uses AR to show a user the a visualization of the sound source. AR is a kind of technology for displaying virtual object in the real world[7]. The user can use a screen or HMD to see not only the real world in real time but also virtual objects on AR markers as shown in figure 2. We make AR markers and objects of the same quantity beforehand. The objects we made are the same as the label or as the image of the sound for which it stand for. Showing objects on markers in the user's visual field provides the information of the sound source.

3.4 Sound Source Position Recognition Function

Since one kind of sound can have several different sources in the environment, such as different cellphones, copy machine, or people in the surrounding environment, our system uses a microphone array to recognize the position of sound sources. Human beings can find out the direction of a sound source through the difference in intensity between their two ears. For machines the direction of a sound source can be

Figure 2: Sample of AR

Figure 3: Image of implementation

recognized through interaural time difference (ITD) or interaural level difference (ILD), which are two common methods, but in our system, we use Microsoft's Kinect sensor to recognize the direction of sound sources[8]. The Kinect sensor's microphone array is made up of four microphones that are on the bottom of the Kinect sensor, and using APIs, which are released by Microsoft, the microphones can recognize the sound source direction in a range of 100 degrees in front of the Kinect sensor. In this way, our system can recognize a sound source even if the same kind of sound source objects exist in the surrounding environment.

4. IMPLEMENTATION

We implemented a prototype of the proposed system. Figure 3 shows an image of how we implemented the prototype. In this section, we will introduce the prototype system in the order of input, process, and output.

4.1 Input

In the prototype, we use Digital Cowboy's Net Cowboy web camera, Microsoft's Kinect for Xbox 360 as the microphone array, and the SDK, which is released by Microsoft for programming.

Table 1: Sample of matching labels and objects

Label	Object
guitar	node
clap	hand
copy machine	thunder
phone	ring
talk	dialog box

4.2 Process

To recognize environmental sound in real-time, we added the function of making learnt data to our system. Thus, we collected and labeled the features of sample sound beforehand. A screenshot of the sound source recognition function is shown in figure 4. When our system works, it collects environmental sound through microphones in real time and calculates the sound clips into features that have 13 dimensions. It then sets the candidate that has the nearest Euclidean Distance with the features calculated in one frame. After three frames, the result of recognition is decided by majority, and then it will be transmitted to the output part. In AR part, we use NyARToolkit to make pattern files of the same quantity with labeled learnt data, attach these AR markers to the matching sound source objects. We used NyARToolkit to detect the AR markers from the images, which were captured by a web camera mounted on the user, and send the screen position of the AR markers, which are detected through pattern files, to sound direction recognition part. The sound source direction microphone array part in the Kinect detects the direction of a sound source, and therefore, our system can figure out which AR marker to use to show a virtual object when several of the same AR markers are in the surrounding environment.

4.3 Output

To provide information on the surrounding environment, the user wears an HMD. An HMD is an ideal device for providing the position and other information of a sound source to a user because the HMD can display virtual objects in the real world environment in the user's field of view. When the AR markers are detected, the virtual objects will show on the AR markers, which are matched with the labels of the detected sound source. Table 1 is a sample of how labels are matched with objects.

5. VALIDATION

We implemented and verified the proposed system. The order of consideration is sound source recognition function, sound direction recognition function, AR function, and comprehensive evaluation.

5.1 Sound Source Recognition Function

The database of learnt data is made up of six kinds of sample sound features. Concretely, they are the sound of a guitar, clapping, ringing of a cellphone, the sound of a copy machine working, and silence as sample sound. Each learnt data contained 250 frames, and each frame was made up of 13 dimensional features, which were calculated by the MFCC. Afterwards, we operated our system in a real-world environment to confirm whether it can recognize a sound source correctly when we produce the sounds of a guitar and clapping.

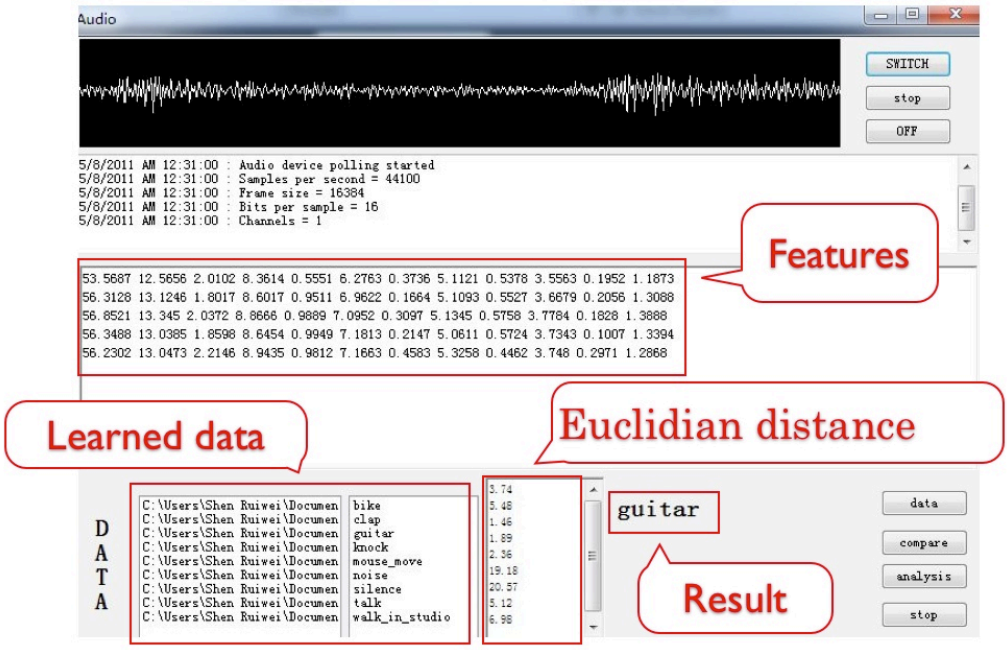

Figure 4: Sound recognition function

- Our system recognizes a sound source in one second, after three frames, the recognition result is decided by majority. Therefore, about 2 seconds of delay is coursed.
- In the situation that there is only one kind of object making a sound in the surrounding environment, the accuracy of recognition can be over 90%.
- In the situation that there are several kinds different kinds of objects making sound in the surrounding environment, the accuracy of recognition can be less than 60% because the specific sound source can be hard to separate form the environmental sound.

5.2 Sound Source Direction Recognition Function

We evaluated the sound source direction recognition part. In our system we use the microphone array, which is in the Kinect to detect the sound source direction. Figure 5 shows a sample of the microphone array being used to detect the direction of a sound source. This figure shows two of the same AR markers in different positions, and through the detection of the sound source direction, we can find out that the object making sound is the right one in the picture.

To detect the direction of a sound source, we fixed the web camera on the Kinect so that it could capture video in a range of 40 degrees. We used Kinect to detect the direction of a sound source in a range of 40 degrees in front of it at the same time and matched the direction with the result of the AR marker recognition to specify the position of the sound

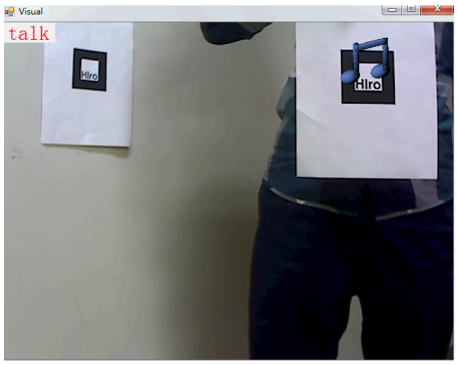

Figure 5: Detecting the sound source direction

source. As the result of validation, the position of the sound source, which was calculated by the Kinect, was in a range of 20 pixels on the x vector of the AR marker on the screen. We considered the reasons for this are as below.

- To match the result calculated by the Kinect with the position of AR marker, we needed to convert degrees to pixels, and this conversion could have caused the deviation.
- We used Microsoft's official SDK for Kinect to detect the direction of a sound source. Therefore, the Kinect may have a deviation that interferes with detecting the direction.

Figure 6: Visualization using AR

Furthermore, the Kinect can only detect the direction in a range of 100 degrees and 3 meters in front of it. Therefore, the user will hardly find sounds that are beside or far from him or her.

5.3 Augmented Reality Function

We evaluated the AR part of the proposed system. Figure 6 shows four samples of sound source visualization using our system: the ringing of a cellphone, the sound of a guitar, the sound of a shredder crushing paper, and the sound of a copy machine printing.

There are several parameters in NyARToolkit's library that are used in the AR part of our system, such as the size of the AR markers and the threshold of the recognition rate, and we set the side of each AR marker to be 80 millimeters and the threshold of AR markers's recognition rate to be 0.6. If several different kinds of AR markers exist in one picture, the recognition accuracy of AR markers can be about 80%. Virtual objects, which are prepared beforehand, can be displayed on the AR markers, but because of the light and the distance between web camera and AR marker, it can be miss recognized in these situations.

- If AR markers are under intense light, the binarization of AR markers may meet trouble. Because of the reflection of the light, our system cannot recognize the AR markers.

- If the marker captured is small, the system can hardly recognize it.

- If part of a marker is covered by an other object, the system cannot recognize it.

5.4 Comprehensive Evaluation

By confirming the effectiveness of sound source recognition, augmented reality, and sound source direction recognition mentioned above, we found that user can uses our system to recognize and visualize environmental sounds. In the situation that there is only single sound source in the surrounding environment such as at home or when doing some simple work, especially when sound source is near to a user, our system can provide information on the sound source and visualize the sound source to satisfy the user's need. However, in outdoor environments, it is difficult to recognize sound source. The recognition accuracy falls because of the complex sound source environment. Furthermore, if the AR marker is taken far away from the web camera that is mounted on the user, the recognition accuracy of sound source position can fall.

6. CONCLUSION

In this research, we proposed a sound source visualization system that uses augmented reality so that the hearing impaired and those who are working indoor for recognize environmental sounds, especially danger sounds. The prototype system, which was implemented, can recognize a sound source and the position of the sound source and visualize the sound source in user's field of vision by using AR.

As future work, we need to use questionnaires to find common images for sounds and make objects so that users can

realize the sound easily. We will apse discuss how to use the method to recognize several sound sources at the same time we need to raise the recognition accuracy in complex sound environments because the user needs to grasp the details of the surrounding environment. In addition, to improve usefulness of this system, we need more experience to create a new recognition method. Not only the confirmation of the prototype system, we also need to evaluate the usefulness of the system, the effect of visualization, and the level of satisfaction after using the system.

7. ACKNOWLEDGMENTS

This research was supported in part by a Grant in aid for Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology Agency and by a Grant-in-Aid for Scientific Research(A)(20240009) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

8. REFERENCES

[1] T. Matthews, J. Fong, and J. Mankoff: Visualizing Non-Speech Sounds for the Deaf,*Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pp. 52–59 (Oct. 2005)

[2] J. Azar, H. Abou Saleh, and M. A. Al-Alaoui: Sound Visualization for the Hearing Impaired, *International Journal of Emerging Technologies in Learning -iJET*, Vol.2 , No.1 (March 2007).

[3] M. Bertram, E. Deines, J. Mohring, J. Jegorovs, and H. Hagen: Phonon tracing for auralization and visualization of sound, *Proceedings of IEEE Visualization*, pp. 151–158 (Oct. 2005).

[4] Hans G. Kaper, Sever Tipei, Elizabeth Wiebel, Data sonification and sound visualization, *Computing in Science and Engineering*, 1(4), pp. 48–58, (1999).

[5] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1228–1233, vol.2, (2003).

[6] B. Clarkson, N. Sawhney, and A. Pentland: Auditory Context Awareness via Wearable Computing, *Workshop on Perceptual User Interfaces (PUI '98)*, San Francisco, CA (Nov. 1998).

[7] H. Kato, and M. Billinghurst: Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, *In Proc. IWAR '99*, pp.85-94 (Oct. 1999)

[8] B. Rakerd and W.M. Hartmann : Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise,*J. Acoust. Soc. Am.*, 128(5), pp. 3052–3063 (2010).