Finding Nursing in the Room from Accelerometers and Audio on Mobile Sensors

著者	Nakamura Masato, Inoue Sozo, Nohara Yasunobu, Nakashima Naoki
journal or publication title	Proceedings of the Third International Workshop on Location Awareness for Mixed and Dual Reality (LAMDa'13)
page range	17-20
year	2013-03-19
URL	http://hdl.handle.net/10228/00007654

Finding Nursing in the Room from Accelerometers and Audio on Mobile Sensors

Masato NAKAMURA and Sozo INOUE
Kyushu Institute of Technology, Japan
{masa1031to@gmail.com|sozo@mns.kyutech.ac.jp}

Yasunobu NOHARA and Naoki NAKASHIMA
Kyushu University Hospital, Japan
{y-nohara|nnaoki}@info.med.kyushu-u.ac.jp

ABSTRACT

In this paper, we propose a method for finding intervals of nursing activities from accelerometers and audio on mobile sensors which are attached to nurses in reality. If we can find the intervals of nursing activities correctly, it helps the data to be used for machine learning for activity recognition. We have extracted the times of nursing interactions between nurses and patients by A) recognize walking activity from accelerometers, B) recognize if s/he is in the patient's room or not at each time duration divided by walking activities, from the environmental noise levels of sounds, and, C) for the duration where s/he is assumed to be in the patient's room, apply voice activity detection by fundamental frequencies using Cepstrum method, and extract the duration in which a person speaks. As a result of the experience for 300sec of sensor data, we observed sufficient accuracy for each step of A)-C), and could reduce the time to 8%.

Author Keywords

Activity Recognition, Annotation, Speech Interval Estimation, Nursing Activity

INTRODUCTION

In this research, we aim at capturing nursing interactions with patients from mobile accelerometers attached to each nurse. Capturing nursing is important, since 1) it helps understanding what/when/how interactions should be performed for better health results of the patients, and 2) it can be utilized to improve the skills of nurses. If we have evidences of interactions and the health result, we can analyze the correlations between them, and find the key factors for better interaction.

However, very few datasets for "real" nursing activity has been published and shared among the research community so far. Although one of the reasons is the immaturity of sensing/network/storage technology, we claim the most major reason is the difficulty of annotation task. In developing activity recognition algorithms, machine-learning technique is applied using training dataset, which is the set of pairs of input sensor data and an activity class, which is the "answer" of activity recognition. To prepare the training dataset, large effort is required for adding activity class information by annotating sensor data, since it almost requires manual labor.

We have collected 7,400 hours of mobile sensor data in total from nurses after one-year trial in a hospital[1]. The sensor device has recorded audio together with the accelerometer. Listening all of them and annotate them requires huge amount of time. Moreover, we need experts of the nursing domain to understand/imagine professional terminologies/activities. Any methods of reducing the labor time are highly demanded.

In this paper, we propose a method automatically finding the interaction part of nursing, which leads to reducing the time of audio data to be listened for annotation task. From sampled part of the collected data, we have extracted the times of interactions between nurses and patients in the following steps: A) recognize walking activity from accelerometers using SVM (Support Vector Machine). B) During certain duration of walking, the nurse is considered to move across rooms. Therefore, from the environmental noise levels of sounds at each interval divided by walking, we recognize if s/he is in the patient's room or not. C) For the duration where s/he is assumed to be in the patient's room, apply voice activity detection by fundamental frequencies using Cepstrum method, and extract the duration in which a person speaks. Since we know by experience that nurses talk and declare what to do to a patient before they perform cares to him/her, knowing and analyzing the talking part will be important for knowing what s/he did and for segmenting accelerometer data to meaningful activities.

As a result of the experience for 300 seconds of sensor data, we observed sufficient accuracy for each step of A)-C), and could reduce the time to 8%, which means the cost of annotating the data will be reduced to almost 8%.

BACKGROUND

In our one-year trial in a cardiovascular center in a hospital, we have collected large-scale mobile sensor data from nurses and patients, along with the medical records of the patients[1]. We asked nurses to bring mobile devices (iPod touches), which records audio and accelerations, into their breast pockets with a roughly fixed direction. They also attached small 2 accelerometer devices on their right wrists and the back waists. Moreover, each of them attached a semi-passive RFID tag in the breast pocket to recognize entrees and exists from the patients' rooms.

Figure 1. Nurses with three accelerometers on the right wrists, the breast pockets, and the back hips. We only used the one on the breast in this paper.

We also asked 70 hospitalized patients who have been applied PCI (Percutaneous Coronary Intervention) or CABG (Coronary Artery Bypass Graft), and have consented to the experiment, to provide vital sensor data such as monitoring cardiogram, bed sensor to measure heart rate and breath, accelerometer, environmental sensor, and also medical information which were recorded in the electronic clinical pathways and indirectly in patients' sensor data. As for the nursing data, have collected 7,400 hours of accelerometers and sounds.

In the experiment, we have a requirement to know the nursing activity interval to know what kind of care are done to each patient. We can focus on the intervals when the nurses are in the patients' rooms, so the RFID system is thought to be useful. However, RFID system is not always available, since the readers and antennas should be placed many places, such as every entrance of the patients' rooms. Therefore, it is welcomed if we can know when nurses stayed in patient rooms without using environmental sensors such as RFID.

RELATED WORK

In the literature, some work utilizes accelerometer and audio data to recognize human context. Lukowicz et al.[3] recognizes activities in a wood shop using body-worn multiple microphones and accelerometers. Lester et al.[4] shows the performance of activity recognition for 8 activity classes using accelerometers, audio, and barometric pressure sensor in a single device. Choudhury et al.[5] developed to implement them on a mobile embedded system. In the device, audio is down-sampled as not to be able for humans to harm privacy of the owner.

One of the differences of our work from above is that these work assume simple activity classes to recognize such as, "walk", "stair up", but our research aims at recognizing more complex and more number of nursing activities. For complex and more number of activity classes, the recognition accuracy will be worse. Therefore, we need more effort to collect larger-scale dataset as well as sophisticated machine learning that can be used in higher dimensions with larger-scale training data.

Another difference is that, in our approach, we set the final goal to recognize nursing activities without using audio, but only with accelerometers. We only use audio data for annotation tasks, because of the following reasons: 1) audio data often includes outer sounds such as environmental ones and other persons' speaking, which could be noises for recognition, and 2) audio has a risk of invading privacy of the owner. Even if it is recorded in lower quality, end users might not believe the safety of privacy by knowing audio is recorded.

Overall, we tackle with more complex activity classes, and try to annotate them as effective as possible as a first challenge.

METHOD FOR FINDING NURSING ACTIVITIES

In this section, we propose a method to find the interval which corresponds to medical activities. This method uses three-axis acceleration data and audio data that are collected by the devices attached to the breast pockets. Upon the collected activity data, we use two characteristics in order to efficiently locate the intervals where nurses performed medical activities.

One is the characteristic that a nurse certainly speaks to a patient when s/he performs medical practice to a patient. Nurses always talk to the patients what to do for medical practice. Therefore, if we can find an interval where nurses are talking, we can guess that the interval of medical activities is being performed.

The other is that a nurse walks for a specific while when s/he moves into a patient 's room. If we can detect the walking of nurses to move into the patient's room from 3-axis accelerometer, we can segment the time to either of being inside or outside the room. In addition, we can estimate if s/he is in the patient's room by examining the noise level from the audio data after a walking period.

In order to utilize the above characteristics, we adopt mobile sensors which record three-axis acceleration and audio data. With the data collected by the devices, we apply walking detection method for the accelerometer, speech interval estimation for audio data, and location estimation for the environmental noise level of the audio data. We can find the duration of walk from three-axis acceleration data by walking detection, location estimation from the environmental noise level of the audio data after walking periods, and the durations where a nurse talks from audio data by speech interval estimation.

Walking detection

In order to detect the walk of nurses, we recognize the walk of nurses using the technology of activity recognition. We calculate the feature vectors to train an activity recognition model from the three-axis acceleration data. Feature vectors are calculated with the time window of 2 seconds being shifted by 0.5 seconds. A feature vector consists of the variance and the entropy of the intensity: the square root of the sum of squares of the three-axis values of acceleration data. The recognition model is trained by support vector machine (SVM). To smooth continuous walking, the duration of less than 15 seconds between detected walks are also assumed as walk.

Location estimation

We can estimate if s/he is in the patient's room by the environmental noise level from the audio data. If the audio

is recorded in 16-quantization bit rate, the amplitude bandwidth is from -32768 to +32767. From our experience, environmental noises of our target were found to be from -1500 to 1500. Therefore, at first, we remove the intervals of amplitudes outer than -1500 from 1500, which contains human voices and metal sounds. After that, we estimate the location by the median amplitude value of 30 seconds after the end of walking period.

Speech interval estimation

To find the nurses' speech interval, we estimate the speech interval by seeking fundamental frequency of the audio data. The fundamental frequency is one of the speech features used in speech recognition, and it represents the height of the voice. Calculation of the fundamental frequency is performed by the cepstrum method[2]. Although the cepstrum technique is weak for noises, there are advantages that the fundamental frequency can be correctly acquirable in any languages. *Cepstrum* $c(\tau)$ is defined as the inverse Fourier transform of the logarithm of the short amplitude spectrum $X(\omega)$ of the waveform. Cepstrum separates the original voice to the approximate spectral envelope and the fine structure. Also, cepstrum's horizontal axis is called the *quenfrency*. When the Fourier transform of the impulse response of the sound source (vocal tract) is represented by $G(\omega)$ ($H(\omega)$, respectively) , the following relation holds.

$$X(\omega) = G(\omega) \cdot H(\omega) \tag{1}$$

Taking the logarithmic of equation (1),

$$\log|X(\omega)| = \log|G(\omega)| + \log|H(\omega)| \tag{2}$$

The Cepstrum is the Fourier inverse transform of the expression (2) with the angular velocity ω as the variable. Equation (2) is denoted with Fourier transformation function F as,

$$c(\tau) = F^{-1}(\log|X(\omega)|)$$

$$= F^{-1}(\log|G(\omega)|) + F^{-1}(\log|H(\omega)|) \tag{3}$$

$G(\omega)$ of the expression (2) is a fine structure on the spectrum, and $H(\omega)$, is the spectrum envelope. Therefore, the 1st term of the formula (3) of right-hand side comes from the peak of a high quefrency part, and the 2nd term comes from the low quefrency part of about 0 to 4 milliseconds. As a result, we can obtain fundamental time period from the peak of the higher quefrency, and we can calculate the fundamental frequency.

In this study, using the Cepstrum method, fundamental frequency is calculated with the time window of 0.04 seconds being shifted by 0.02 seconds. By obtaining the time window with high peak quefrency, we can obtain the spoken interval.

EXPERIMENTS

We have conducted the experiments using real nursing data to evaluate the proposed method. The used data is activity data of one day of a nurse.

Walking Detection

In order to evaluate the walking detection, each of the training and test data with annotation for 300 seconds were prepared from a day of a nurse. Two kinds of annotations, "walk" and "others", are attached to the data. The data contained 100 seconds of "walk", and 200 seconds of "others". Recognition model was created by the modelusing the training data, and was evaluated by the test data. Tab. 1 shows the recognition result before smoothing.

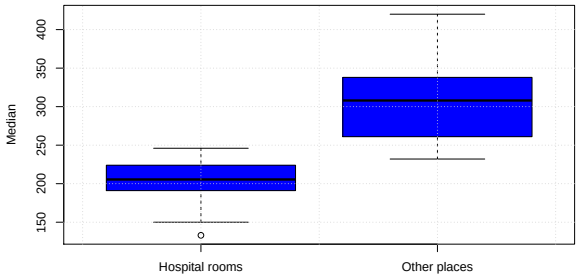

Figure 2. Distribution of median environmental noise levels of 43 data points of 30 seconds after a walking period. The left is in the hospital room and the right is in other places.

Table 1. Confusion matrix of the number of time windows for walking detection

→ Ground truth	Walk	Others
Walk	52	18
Others	19	492

Tab. 1 is the result of recognition. From the table, the whole recognition rate is 93.6%.

Location estimation

We picked up 43 data points from 4 audio data, and investigated the environmental noise level, which is put together in Fig. 2.

In Fig. 2, the left box is the distribution of the median environmental noises in the hospital room, and the right is in other places. Since the inter-quartile ranges (IQRs) do not overlap each other, we can estimate that we can differentiate the location at more than 75%. If we take priority on the recall rate, we can achieve at least 87.5%.

Speech interval estimation

We evaluated the speech interval estimationusing audio data of 300 seconds. The audio data was prepared from a day of a nurse. Fig. 3 shows the original sound and the result of the calculation from the proposed method of sampled 15 seconds.

In the figure, we can see that the noises are successfully eliminated around before 18 second and around 22 second. In addition, another person's voice around 19.5 second failed to be detected because the speaker seems to have been far from the microphone. However, it is acceptable, since our goal is to capture the nurses' voices only.

The confusion matrix which counts of the results are shown in Tab. 2. For comparison, the ground truths are classified as the nurses' speeches, patients' speeches, noises, and the silent

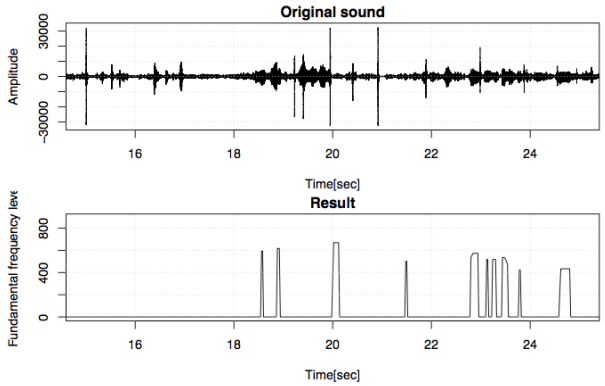

Figure 3. Result of speech interval estimation. The upper is the ground truth, and the lower is the recognition result. The Y-axis of the lower means the level of fundamental frequency.

Table 2. Confusion matrix of the durations for speech interval detection.

→ Ground truth	Nurse	Patient	Noise	Silence
Speech	27.62[s]	7.65[s]	0.14[s]	0.76[s]
None-speech	0.83[s]	2.6[s]	13.72[s]	246.68[s]
Total	28.45[s]	10.25[s]	13.86[s]	247.44[s]

intervals, whereas the proposed method only estimates speech or non-speech. The silent intervals of the ground truths were determined by whether the amplitude is greater than a specific threshold value, which resulted in that negligible small voices were included in the silence class. From the table, the method recognizes the speech intervals with the accuracy of 98.6%. However, the recognized speech includes patients' speeches. If we evaluate the rate of recognizing nurses' speeches only, it becomes 96.9%, which is still higher recognition rate.

Integration

We integrated the three method described above, and applied to 300 seconds which are obtained from a day of a nurse.

Fig. 4 shows the results of the speech interval estimation and walking detection. The above figure of the figure is the result of walking detection, in which three walking periods are detected. After applying location estimation method to the three intervals of 30 seconds after walking, only the first one of after 65.5 second was estimated to be in a hospital room. Then, applying speech interval estimation to that interval, the total time of speech interval were found to be 24 seconds. This means that we were able to reduce the time to listen for annotation tasks down to 8%.

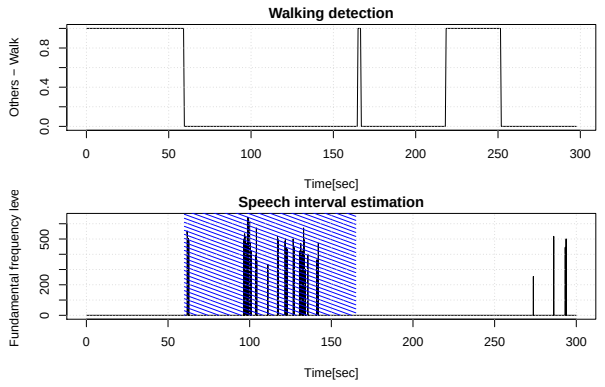

Figure 4. Result of integrated process. The upper is the result of walking detection, and the lower is the speech interval estimation. After applying 3 parts of 30 seconds after walking period detected by the upper part, the first 30 seconds were detected as in the patient's room, which could be applied by the speech interval estimation of the lower part.

CONCLUSION

In this paper, we proposed a method to find nursing activities from accelerometers and audio on mobile sensors which are attached to the nurses in a hospital. The proposed method firstly detects walking parts from accelerometer, infer whether s/he is in the patient's room or not by environmental noise levels of audio from the intermediate durations, and then, for the in-room durations, detect the voice activity from the audio. The extracted voice activities are to be utilized to reduce the cost of annotate nursing activities, and to open and share together with the whole sensor/medical data to the public for future research. From the experiment using 300sec of data, we could reduce the time to 8%, which means the cost of annotating the data will be reduced to almost 8%. As future work, we tackle with removing other people's voice from audio, and to apply and evaluate the already-collected large-scale nursing sensor data.

ACKNOWLEDGEMENTS

This work is supported by Grant-in-Aid for Young Scientists (A) (21680009) of JSPS and Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST). The authors would like to thank their support. We also appreciate the cooperation for experiment by the staff of Saiseikai Kumamoto Hospital, Japan.

REFERENCES

1. Yasunobu Nohara, Sozo Inoue, Naoki Nakashima, Naonori Ueda, Masaru Kitsuregawa, "Large-scale Sensor Dataset in a Hospital", International Workshop on Pattern Recognition for Healthcare Analytics, 4 pages, November 11, 2012, Tsukuba, Japan.
2. B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes: cepstrum, pseudo-auto covariance, cross-cepstrum, and shaft cracking". Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed), Chapter 15, pp. 209-243. New York: Wiley, 1963.
3. Paul Lukowicz, Jamie A Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash and Thad Starner, "Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers", Proc. Int'l Conf. Pervasive, vol. 3001, pp. 18-32, Springer LNCS, 2004.
4. Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello, "A Practical Approach to Recognizing Physical Activities", Proc. Int'l Conf. Pervasve, vol. 3968, pp. 1-16, Springer LNCS, 2006.
5. T. Choudhury et al., "The Mobile Sensing Platform: An Embedded Activity Recognition System", IEEE Pervasive Computing Magazine, vol. 7, issue 2, pp. 32-41, 2008.