ROC Analysis of Partitioning Method for Activity Recognition Using Two Microphones

Jame A Ward [*], Paul Lukowicz [†] and Gerhard Tröster [‡]

Abstract. *We present an analysis, using ROC curves, of a method for partitioning continuous activity data using two microphones. The algorithm is based on utilising the difference in sound intensity recorded by microphones placed on the upper arm and on the wrist. We show that the method is feasible for detecting activities involving the interaction of the hand with tools or machinery where noise is produced close to the hand. We also show that the method is relatively robust across multiple subjects.*

1. Introduction

Activity recognition for application in a maintenance or assembly scenario is currently a hot topic in the field of wearable computing. The development of such systems is the aim of the current European Union WearIT@Work project in which our group participates. To compliment this work we have developed a mock assembly scenario in a wood workshop which provides a useful dataset upon which techniques for activity recognition can be realised and evaluated. The first dataset, as reported in [2], was collected using two microphones placed at different parts of the body, together with a number of 3-axis accelerometers. It involved a single subject performing a series of pre-defined activities using typical woodwork tools (hammer, saw, file, sandpaper, screwdriver, drill, grinder, vise and drawer). One of the drawbacks of this work was in the use of a single subject. This has since been rectified by increasing the number of subjects to five. Whereas the broader work on development of activity recognition for this environment is ongoing, recent results from an analysis of the partitioning method using sound are presented here.

1.1. The partitioning problem

The main problem tackled in this work is that of partitioning, i.e. finding the beginning and end times of interesting events from a continuous stream of data. The assumption is that once this information is known, then recognition can then be performed by classifying such interesting events in isolation from the non interesting (garbage or null) events.

Our solution to this problem is to use sound analysis to identify relevant signal segments. In particular we use the intensity difference between wrist and upper arm mounted microphones in such a way that it allows us to detect sounds made close to the hand.

[*] Swiss Federal Institute of Technology (ETH), Wearable Computing Lab, 8092 Zürich, CH, email: ward@ife.ee.ethz.ch

[†] UMIT, Biomedical Informatics, Innsbruck, AT, email: paul.lukowicz@umit.at

[‡] Swiss Federal Institute of Technology (ETH), Wearable Computing Lab, 8092 Zürich, CH

The approach is based on the assumption that all of the activities we are interested in produce some kind of noise close to the hand. While this is certainly not true for arbitrary human activities, in our case it is a reasonable assumption, as most assembly tools and machines produce characteristic sounds. Any time where such characteristic sounds are not detected is defined as as null.

This algorithm was presented initially in [3] and implemented in [2]; this paper continues this work and applies the algorithm to a multiple subject scenario, however we limit this treatment to an analysis of the intensity analysis algorithm alone and make no attempt to smooth or filter the output as suggested in [2].

2. Intensity Analysis (IA) Using Two Microphones

Our partitioning cues are obtained from an analysis of the difference in sound intensity from two different microphone positions. It is based on the following ideas:

1. Most activities are likely to be associated with a characteristic sound which will originate in the proximity of the hand.

2. In general, two microphones 1 and 2 placed at different locations on the body will have a different distance from the source of any sound. Thus the signal intensities I_1 and I_2 will also be different. Since the intensity of a sound signal is inversely proportional to the square of the distance from its source, the ratio of the two intensities I_2/I_1 depends on the absolute distance of the source from the user. Assuming that the sound source is located at the distance d from the first microphone and $d + \delta$ from the second, the ratio of the intensities is proportional to:
$$\frac{I_1}{I_2} \simeq \frac{(d+\delta)^2}{d^2} = \frac{d^2 + d\delta + \delta^2}{d^2} = 1 + \frac{\delta}{d} + \frac{\delta^2}{d^2}$$
For sound sources located far from both microphones (and thus from the user), d will be much larger then δ (since δ can not be larger than the distance between the body parts on which the microphones are located). As a consequence the quotient will be close to one. If, on the other hand, the source is very close to the first microphone, then d will in general be smaller than δ. This in turn means that I_1/I_2 will get much larger then one. This is an indication that the source of a sound is located very close to the body part on which microphone 1 is placed. Putting the first microphone at the wrist, and the second one on the upper arm, we can use a large quotient as a sign that the sound was generated close to the user's hand.

Based on these ideas, the intensity analysis (IA) algorithm works as follows:

1. Slide a window w_{ia}, in increments of j_{ia}, over both channels of audio data, calculating the signal intensity ratio I_1/I_2 at each step

2. For each window, calculate the difference between the intensity ratio and its reciprocal, i.e.: $I_1/I_2 - I_2/I_1$. This provides a more convenient metric for thresholding: zero indicating a far off (or exactly equidistant) sound, while above or below zero indicates a sound closer to the microphones 1 and 2 respectively.

3. Compare this ratio difference with a suitable threshold T_{ia}, which if passed indicates a potentially interesting frame; and if not, classified as null or garbage.

2.1. Assembly Experiment with a Wood Workshop

In this experiment, mic. 1 was placed near the wrist and mic. 2 at a relatively fixed distance away on the upper arm. The audio streams were sampled at 2kHz. The window size w_{ia} chosen for this sample frequency was 100ms, with an increment j_{ia} of 25ms.

Five subjects were employed (1 female, 4 male), each performing a sequence of activities in repetition between 3 and 6 times producing a total of (3+3+4+4+6)=20 recordings. Each sequence lasted on average five minutes. (Some subjects performing more repetitions than others due to a combination of technical problems in recording and availability.)

Each of the activities were separated into four distinct sound categories: handheld - sounds produced by the use of some object (or tool) held in the user's hand; machine - sounds produced by an external machine with which the user might interact; quiet - activities which do not produce much noise, or produce noise away from the user's hand; and null - the 'garbage' class of background noises and silences.

The table below shows each of the four categories and gives the total duration (in seconds) represented in the collected data, together with the constituent activities.

category	activities	total time(s)
handheld	hammer, saw, sandpaper, file	1119
machine	drill, grinder, drawer	1178
quiet	screwdriver, vise	938
null	”	2778

3. Results

In order to gain a better idea of how the algorithm works over a range of threshold values T_{ia}, graphs plotting the Receiver Operator Characteristic (ROC) were used[1]. ROC curves plot true positive rate (or *recall*) against the false positive rate (*fp*), as defined below, for a sweep of some decision parameter (in this case T_{ia}):

$$TruePositiveRate = recall = \frac{TruePositiveTime}{TotalPositiveTime} \tag{1}$$

$$FalsePositiveRate = fp = \frac{FalsePositiveTime}{TotalNullTime} \tag{2}$$

Where Positive time refers to the duration of each non-null activity, and Null time refers to duration of null.

The recall for each activity category - handheld, machine, quiet, and an average over all activities - was plotted against the false positive rate for each T_{ia} (across the range 5 to -5). The results for each subject are shown in Figure 1. Points falling on the diagonal indicate a random result; the further above the diagonal (i.e. high recall, low fp), the better the result. Also shown is the graph of the mean

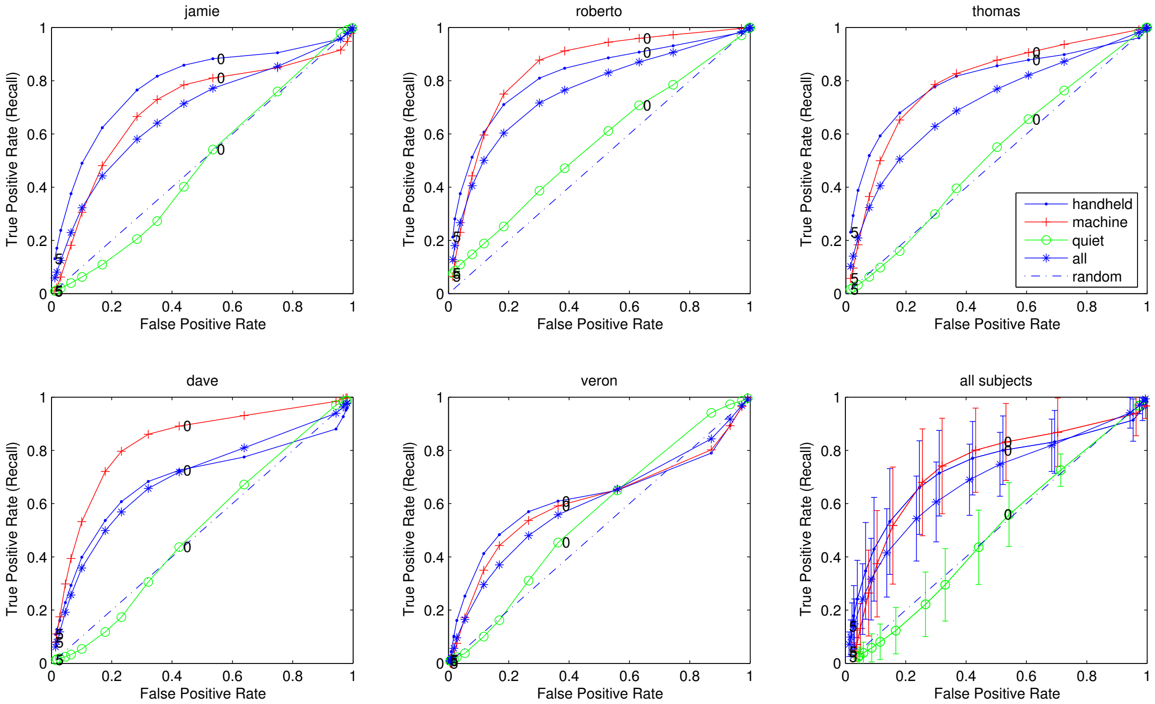

Figure 1. ROC graphs for the five subjects, and (bottom right) mean and standard deviation (of recall) over all subjects. Each graph plots the recall versus false positive rate over a sweep of T_{ia} for each activity set (handheld, machine, quiet and all activities.) The point where $T_{ia} = 0$ is shown on each curve (with positive thresholds to the left and negative to the right of this point.)

recall and mean fp (for a given threshold value) over all subjects. The standard deviation of the recall measure is shown by the error bars.

3.1. Analysis of results

As might be expected, the IA algorithm picks out both handheld and machine activities consistently better than quiet activities.

For most subjects, the machine sounds produce the best results of all. One reason for this is that machines tend to produce a constant sound when on. Additionally, one of the experimental constraints was that subjects kept their hand close to the machine at all times during operation. In reality, it might be expected that such strong performance will fall on relaxation of this constraint.

With the handheld tools, this constraint need not be enforced - it is inherent in the manner of use. It can be expected therefore that the performance reported here would be an accurate prediction of a real world application (as shown, to some extent, by a smaller standard deviation for handheld on the bottom right graph of Figure 1.)

The results do vary somewhat between users, in particular the only female subject (Veron) shows a relatively poor response. This might be explained by the somewhat awkward setup involved with placement of the microphones. A more comfortable arrangement of sensor placement - using smaller, more wearable apparatus- is a topic currently being investigated.

In general however, the graphs show that depending on the application requirements - recall vs. fp tradeoff - a suitable IA threshold parameter might be chosen independent of the user.

4. Summary

This paper presented an ROC evaluation of the two microphone Intensity Analysis algorithm. The following conclusions were made: this partitioning scheme is suitable for use in detection of hand-held tool activities (provided a sound is made during use); it is also suitable for detecting activities involving interaction with a (noisy) machine - provided that the user's hand comes into contact with machine; finally, the algorithm parameters can be set independent of any user.

References

[1] Tom Fawcett. *ROC Graphs: Notes and Practical Considerations for Researchers*. Kluwer Academic Publishers, 2004.

[2] Paul Lukowicz, Jamie A Ward, Holger Junker, Gerhard Tr/¿oster, Amin Atrash, and Thad Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*, 2004.

[3] Mathias Stäger, Paul Lukowicz, Niroshan Perera, Thomas von Büren, Gerhard Tröster, and Thad Starner. Soundbutton: Design of a low power wearable audio classification system. 7th Int'l Symposium on Wearable Computers, 2003.