EasyChair Preprint

№ 341

Object-based Activity Recognition with Heterogeneous Sensors on Wrist

Mohammad Albaida

July 14, 2018

Object-based Activity Recognition with Heterogeneous Sensors on Wrist

Mohammad Albaida

Abstract

Recent development of wearable technology has opened great opportunities for human performance evaluation applications in various domains. In order to measure the physical activities of an individual, wrist-worn sensors embedded in smart-watches, fitness bands, and clip-on devices can be used to collect various types of data while the subject performs regular daily activities.In this paper we are going to explain how to achieve activities of daily living(ADLs) using a sensor device attached to a users wrist. This device contains a camera, a microphone and an accelerometer. In this experience we will collect the data from the sensors in our device and try to analyse it, in order to recognise the type of the activity. In this way we will be able to recognize ADLs that contain manual use of objects such as making a drink or cooking. Finally, we will be able to say, that the camera plays the major rule in this experience and without it would be difficult to achieve our goal. We will also suggest a method that will protect the privacy of the user, as the camera and the microphone can records part of the users private life.

1 Introduction

The recognition of human activities by computers is a field of research that reaches back to the 1980s. It gained increased attention by the widespread possibility to observe sensors in the environment. For example, mobile phone motion sensors have been a popular choice for activity recognition in a trouser pocket or a similar position. Recently, wrist-worn motion sensors are also being used for human activity recognition. We can categorize activity recognition technologies into two main approaches:

-wearable sensing

-environment augmentation

The wearable sensing approach recognizes activities by using sensor data obtained from such body-worn sensors,as the accelerometer,the microphone or the camera. In many cases, the environment augmentation approach uses ubiquitous sensors such as RFID tags and/or switch sensors installed in the environment. Although the environment augmentation approach places a smaller burden on

the user than the wearable sensor approach, ubiquitous sensors are expensive to deploy because one must attach them to various indoor objects and maintain many of them. On the other hand, several studies have used sensor devices attached to a single point in the environment. The wearable sensing approach tries to recognize a users activities by employing the mentioned sensors to capture characteristic repetitive motions, postures, and sounds of activities. So many activities have been successfully recognized by using the wearable sensing such as: running, cooking and brushing teeth. One of the advantages of wearable sensing is that, it can be used indoors or outdoors. In this study we will try to achieve the ADLs using many kinds of sensors including (camera, microphone and accelerometer) attached to a single point on the body, which would be the wrist. At the beginning we will describe our device and build a prototype and after that we will send the data that has been collected wirelessly by the sensors to the host PC. After that we will use a supervised machine learning based ADL recognition method that uses the multi-modal sensor data. Our method will consider the private information of the user and because of that we will design a recognition method by which the sensor device does not send raw private information, but abstract information. Then we will capture ADLs that involve objects, such as taking medicine, making tea or even watering plants. Finally, we will evaluate our recognition method. The rest of the paper is structured as follows: (2) The design of the sensor device, (3) Proposed method (4) Recognition, (5) and Evolution.

2 The design of the sensor device

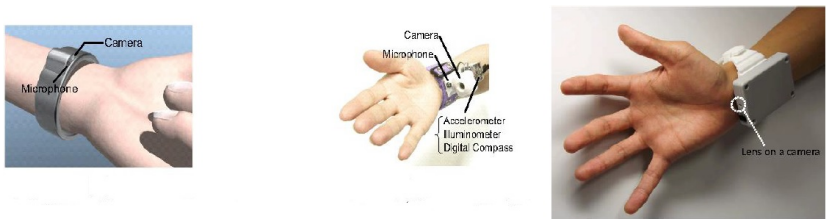

Figure 1: Conceptual image of wristband type sensor device

The device design is easy. We are going to use a camera, a microphone, an accelerometer, an illuminometer, and a digital compass and attach the device on the wrist. In other experiments, sensors are attached to different places such as one of the ankles, hips or even the head. In our experiment we are going to attach it to the wrist. All ADLs in our experiment involve objects are performed by hand, and the camera lens is placed on the inside of the wrist to capture the space around the persons hand. This is so that the camera can capture objects held by the user and objects around his hand, and it would be a perfect place to capture the objects. In addition, attaching the device on the wrist will be more comfortable for the user than the ankle, hips or even the head, which can

bother the user. Moreover, the size and costs of the sensors will be suitable. Fig.1 shows our wrist-worn sensor device, one can see how the camera is placed inside of the wrist. The sampling rates of the accelerometer and camera are about fifty and five Hz respectively, and the USB camera captures 352 by 288 pixel 24-bit color JPEG images at about 6 fps; this happens with an automatic focus and white balance function.

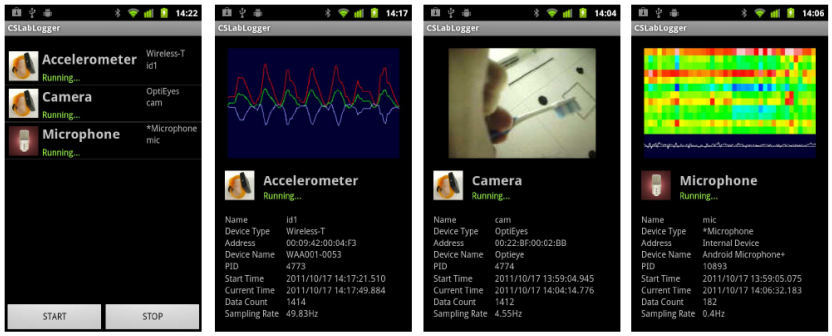

Figure 2: Screenshots of the sensor data logger application showing (b) acceleration data, (c) camera images, and (d) extracted sound features.

Fig.2 is a screenshot of the logger application. As we can see in Fig.2(a), it shows a list of sensors, which are being used in our experiment. And if we select any of the items in the list, then the logger will show the sensor data from the corresponding sensor as it was shown in Fig2.(b), (c) and (d). In Fig. 2 (b) we see the data obtained from the accelerometer in the device in real time. Using the accelerometer, we can capture hand postures and hand movements of the wearer. For example, when the wearer brushes his teeth, we can observe a characteristic frequency in the sensor data. Fig.2(c) shows the data that were sent from the camera. We can see that the wearer is using a toothbrush. Fig.2 (d) is a screenshot, which shows the data obtained from the microphone. Using the microphone is helpful to recognize activity such as vacuum cleaner, tooth brushing, and running water.

3 Proposed method

Collecting labelled sensor data

If we want to recognise the data with supervised machine learning approaches, the wearer must prepare his own labelled training data and specify its start and end points. To do that, there are many methods like embedded cameras in the environment or using a PDA device and everyone has advantages and disadvantages. But we found that our devices design is more helpful because we have a camera attached to the hand and using it we can make an accurate

label set by viewing image sequences recorded by the camera.

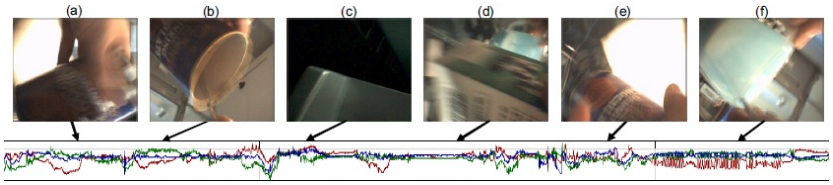

Figure 3: Example camera and acceleration data for making chocolate milk.

Problems with camera images

In our experiment there are two main problems. The first one is that the user will not feel comfortable to send the images to a host PC. Because some photos may contain private situations, such as those in the toilet. The second problem is that, our device is sending continuously data to the PC which requires about 90 KB/sec for raw image transmission and this is exhausting for the battery.

Summary of our approach to visual feature extraction

According to above we must send secured images with small volume. To do that we will extract rough visual features from an abstracted image sent from the device. Some studies also achieve fast object recognition [2, 3] by comparing histograms and object models prepared in advance.chieve fast object recognition [2,3] by comparing histograms and object models prepared in advance. With our method we will send a histogram from our device which contain the number of pixels in the image. For example, if the colour of a cocoa tin is blue, the number of pixels whose colour is like to blue in an image may be useful to achieve the activity. From noted training data we will get several characteristic colours in advance. And for each characteristic colour we got, we will count the number of pixels in the histogram that have a similar colour to the characteristic colour. Using the histograms and specific colours we will try to achieve a rough visual feature extraction with low communication and computation costs.

Finding characteristic colors of each ADL

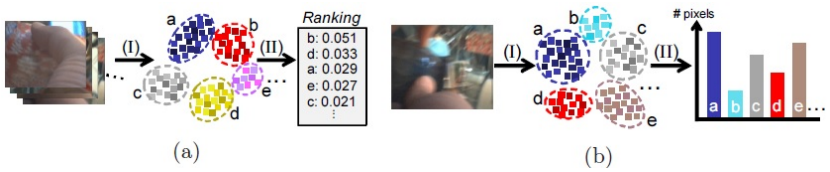

Figure 4: Clustering pixels and ranking them

Using the noted training data, we will have the characteristic colours of each ADL in advance. Fig.4 shows how the procedure woks.Fig.4(a) Clustering pixels of all images of an ADL class and ranking the clusters (features) by their computed information gains, and (b) clustering colour pixels of an image and building a histogram from the clusters. This procedure provides 64 representative colours of the ADL. From those 64 colours we extract the top-m colours as the characteristic colours of the ADL. We rank the 64 candidate colours in terms of information gained, which is then used to find distinguishable (features) of instances.

Visual feature extraction

Using the histogram and the characteristic colour we will be able to achieve the ADL. We then count the number of pixels in the histogram which have a colour similar to the characteristic color to model and recognize ADLs. The method is similar to that used for the characteristic color extraction. We employ this simple method because it requires low computational power.

Sound features

With the help of the sound we can achieve ADLs according to the objects which are used. For example, the sound of vacuum cleaner. In [4], the Mel-Frequency Cepstral Coefficient (MFCC) is reported to be the best transformation scheme for environmental sound recognition. In addition, the computing of MFCC is not expensive because it is based on Fast Fourier Transform (FFT).

Acceleration features

Information like postures and repeated hand movements can be helpful to achieve the ADL. For example, the major FFT frequency components of stirring chocolate milk were between about 2 and 4 Hz in our experiment. Those of brushing teeth were between about 4 and 6 Hz.

Illuminance and direction features

We use illuminometer and a 3-axis digital compass that are helpful to achieve the ADL. Let us suppose that the user usually brushes his teeth in front of a sink in his house. The direction during this ADL maybe the same and would be helpful.

Recognition

After collecting the data we should now use them, in order to recognize it. To do so, we classify each feature vector into an appropriate activity class by using machine-learning-techniques. We use a decision tree that is prepared for each activity type. Each decision tree type computes the probability ofeature vector

being classified into the corresponding class. The class with the highest probability is the classified class of the vector.

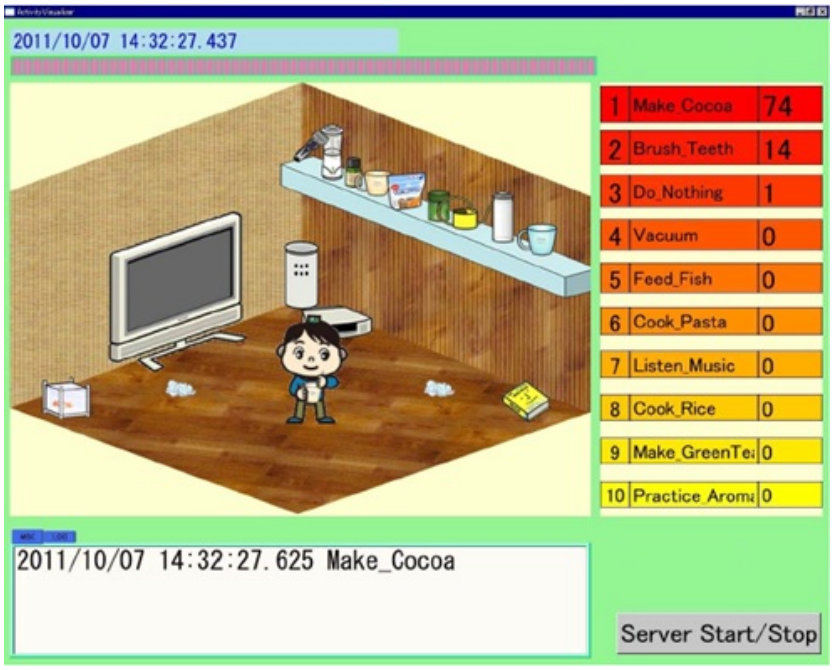

Figure 5: Activity recognition results in real time

Fig.5 is a Screenshot of the application, it shows activity recognition results in real time. On the right side, the ranking of the estimated activities, created based on the computed probabilities, are shown. This application is implemented for a demonstration of real time activity recognition.

4 Evolution

4.1 Privacy

Some users will feel uncomfortable to send images that contain private things wirelessly. The histogram solves the privacy problem because it is impossible to restore the original image from the histogram. In addition to that, it compresses an image into 64 pairs of 24 bit, which enable us to reduce the communication traffic of our device

4.2 The placement of the device

Placing the device on the hand would have advantages such as: It is a good place for the camera to capture the objects and the space around the hand.

In addition, it is comfortable and doesnt require so much money compared to other methods. Furthermore, there are also some disadvantages:
Fig.3 shows a sequence of images while a user made chocolate milk. The objects in the images have the following characteristics: (1) The objects in most of the images are blurred because of the continuous movement of the hand. (2) Most images show only a part of the object. (3) Objects have been captured from different angels. (4) The brightness is depending on lighting, camera and object positions. Many studies try to detect objects from images while taking occlusion, rotation, scale, and blur into account [5, 6]. But to do that we need to take so many pictures from different angels for every single object and this will be hard for the user. In addition, we should use algorithms which have to deal with huge data and this would be exhausting if our device is designed to achieve real time ADL recognition. According to that we can leverage only rough visual information.

4.3 Data transfe

4.3.1 Images

Continuously transmitting raw images in real time occupies a constant communication band. Our implemented device requires about 90 KB/sec for the transmission of raw image. This may also exhaust the devices batteries very quickly. As a result, we determined that the device should send images consisting of small quantities of abstracted data.

4.3.2 Sound

We extracted sound features on the sensor device and only sent these to the host PC. Also, the extraction of sound features chosen from all sound data, captured at a high sampling rate, is costly. Thus, we intermittently capture short periods of sound, and then compute a 13 order MFCC of each captured sound windowed by a Hamming window. In this implementation, we record 25 milliseconds of sound six times a second. From that sound data, we can obtain thirteen features.

5 Refrences

1. WristSense: Wrist-worn Sensor Device with Camera for Daily Activity Recognition Takuya Maekawa, Yasue Kishino, Yutaka Yanagisawa, Yasushi Sakurai NTT Communication Science Laboratories
2. D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking, IEEE Trans. on Pattern Analysis Machine Intelligence, 25(5), pp. 564577, 2003.
3. M.J. Swain and D.H. Ballard, Color indexing, Intl Journal of Computer Vision, 7, pp. 1132, 1991.
4. M. Cowling, Non-speech environmental sound recognition system for autonomous surveillance, Ph.D. Thesis, Griffith University, Gold Coast Campus,

2004.
5. B. Schiele and L.C. James, Object recognition using multidimensional receptive field histograms, Proc. European Conference on Computer Vision, pp. 610619, 1996.
6. D.G. Lowe, Distinctive image features from scaleinvariant keypoints, Intl Journal on Computer Vision, 60(2), pp. 91110, 2004.