Proc. of International Conference on Pervasive Computing (Pervasive 2010), May 2010.

Object-based Activity Recognition with Heterogeneous Sensors on Wrist

Takuya Maekawa, Yutaka Yanagisawa, Yasue Kishino, Katsuhiko Ishiguro, Koji Kamei, Yasushi Sakurai, and Takeshi Okadome

NTT Communication Science Laboratories
2-4 Hikaridai Seika-cho, Souraku-gun, Kyoto, Japan
{maekawa, yasue, ishiguro, yasushi}@cslab.kecl.ntt.co.jp,
yanagisawa.y@west.ntt.co.jp, kamei@atr.jp, tokadome@acm.org

Abstract. This paper describes how we recognize activities of daily living (ADLs) with our designed sensor device, which is equipped with heterogeneous sensors such as a camera, a microphone, and an accelerometer and attached to a user's wrist. Specifically, capturing a space around the user's hand by employing the camera on the wrist mounted device enables us to recognize ADLs that involve the manual use of objects such as making tea or coffee and watering plant. Existing wearable sensor devices equipped only with a microphone and an accelerometer cannot recognize these ADLs without object embedded sensors. We also propose an ADL recognition method that takes privacy issues into account because the camera and microphone can capture aspects of a user's private life. We confirmed experimentally that the incorporation of a camera could significantly improve the accuracy of ADL recognition.

Key words: Wearable sensors; Recognizing daily activities; Experiment

1 Introduction

Activity recognition is one of the most important tasks in pervasive computing applications. This task has a wide range of applications in, for example, context-aware systems, life logging and monitoring and has thus been the subject of a large amount of research. Two main approaches are used for activity recognition studies: environment augmentation and wearable sensing. The environment augmentation approach attempts to recognize users' activities by using sensors embedded in indoor environments. In the computer vision community, activity recognition tasks are accomplished by using cameras installed in a given environment. For example, hand washing and operating medical appliances can be recognized by domain specific solutions [23, 28]. However, the task has become dominated by various types of embedded small sensors. Recently, many researchers in the field of ubiquitous computing have tried to recognize activities based on dense object usage sensors such as RFID tags and switch sensors installed in indoor environments [25, 32, 14]. With this approach, many studies recognize activities of daily living (ADLs) such as using the toilet, making coffee,

washing dishes, and taking medicine by using object usage sensors that are embedded in or attached to such daily use indoor objects and appliances as toilets, coffee makers, sinks, and cups.

The wearable sensing approach tries to recognize a user's activities by employing such sensors as body-worn accelerometers and microphones to capture characteristic repetitive motions, postures, and sounds of activities [19, 20, 2, 3, 16]. Using these types of wearable sensors, sensing studies have successfully recognized such activities as walking, bicycling, brushing teeth, speaking and laughing, and workshop activities such as sawing and drilling that have characteristic motions and/or sounds. An advantage of this approach is that it does not require environment embedded sensors. That is, this approach incurs no cost in terms of money or time for embedding sensors in indoor objects and furniture. Also, users can easily turn off their wearable devices when they want to preserve their privacy. The ADL recognition method proposed in this paper also uses body-worn sensors. However, because most existing studies use only such sensors as accelerometers and microphones, they cannot recognize ADLs that have no characteristic motions or sounds. For example, recognizing such ADLs as making tea and taking medicine, which the environment augmentation approach can achieve by using object usage sensors, is difficult when using only accelerometers and microphones. This study tries to recognize ADLs that involve object use by employing many kinds of sensors including cameras, microphones, and accelerometers attached to a single point on the body. In particular, to recognize these ADLs, we leverage visual features of objects, obtained from a camera on a user's wrist with which we may also easily capture such other features as the motion and sound of the ADLs. One of the characteristics of this study is that it incorporates the visual features of object use into wearable sensing. This permits us to recognize various kinds of ADLs that involve object use without the need for environment embedded sensors. To our knowledge, no work has reported object based ADL recognition employing the vision, sound, and motion features of object use captured by wrist worn sensors.

First, we describe the design of our proposed practical wearable sensor device, which is attached to one point on the body to recognize ADLs that involve object use, and then we build a prototype of the device. We report our design of a wristband type sensor device equipped with such sensors as a camera and a microphone. The device captures sensor data such as images of used objects and the sound emitted when a user performs an ADL and sends them wirelessly to a host PC. Second, we propose a supervised machine learning based ADL recognition method that uses the multi-modal sensor data. Note that, because the raw data obtained from a camera and a microphone on the user's wrist include private information, we design a recognition method where the sensor device does not send raw private information but abstracted information. Third, we collect sensor data by using the implemented prototype device. We capture ADLs that involve object use such as making tea, making green tea, taking medicine, vacuuming, washing dishes, and feeding fish, and annotate the collected data. Finally, we evaluate our recognition method by using the collected data and investigate

(a) (b)

Fig. 1. (a) Conceptual image of wristband type sensor device and (b) prototype device.

the contributions of each sensor. In summary, our contributions reported in this paper are (1) the design of a wearable device that enables us to recognize ADLs that involve object use without environment embedded sensors, (2) the proposal of an ADL recognition method that can detect ADLs involving object use, and (3) an experimental evaluation of the proposed method.

2 Practical sensor device

Our goal is to recognize ADLs that involve the use of objects. Designing a sensor device to achieve this goal, we must choose which types of sensor the device should be equipped with and select which point on the body the device should be attached to. We selected a camera, a microphone, an accelerometer, an illuminometer, and a digital compass from the range of commonly used sensors. We can expect both the cost and size of such sensors to decrease. Specifically, a camera captures visual information about objects used in ADLs. For example, an image (frame) including a coffee maker that is captured when a user makes coffee can be useful for recognizing the ADL of making coffee. The other four types of sensors are usually used for wearable activity recognition [19, 16]. Also, we attach just one wristband type device equipped with the above five sensors to a user's dominant wrist. We attach the device to a single body location because wearing multiple devices on different parts of the body such as the waist, arms, and legs may place a large burden on the user in her daily life. Because almost all ADLs that involve object use are performed by hand, a sensor device attached near the hand can capture ADL characteristics well. Moreover, we can embed these sensors in a wristwatch.

Fig. 1 (a) shows our ideal wristband sensor device designed based on the above discussion. We assume that the device sends preprocessed data obtained from the five sensors wirelessly to a host PC. Feature extraction and ADL recognition are performed on the PC (as shown in Fig. 4). The camera lens is placed on the inside of the wrist to capture the space around the wearer's hand because then the camera can capture objects held by the user and objects around her hand. Based on these assumptions, we fabricated the prototype wristband type sensor device shown in Fig. 1 (b) for the experiment. We fixed together a USB camera, a wired microphone, and a USB cable wired sensor board with a 3-axis accelerometer, an illuminometer, and a 3-axis digital compass and attached them to the wristband. The USB camera captures 352 by 288 pixel 24-bit color JPEG

images at about 6 fps with an automatic focus and white balance function. We used a monaural omni-directional microphone with a sampling rate of 44.1 kHz. The sampling rates of the other three sensors on the sensor board were all about 30 Hz. The frequency of the accelerometer was sufficient compared with the 20 Hz frequency that is required to access daily activities [4]. We selected a small camera and a thin USB cable to avoid disturbing the user's activities. We also bound the sensor cables together with tape. The bundle was fixed in place with a band worn on the brachial region. These sensors are connected to a laptop carried in a backpack via the cables and they send their data to the laptop.

3 Proposed method

We model each ADL class trained with annotated training data and use the models to classify test data. To recognize ADLs by using sensor data, training data should be acquired in each user's environment because these sensor data are environment dependent. For example, the sound of vacuuming may depend on the type of vacuum cleaner used in the environment. That is, users should label each ADL collected in their environment during a certain period of time. Models of ADL classes for ADL recognition are then generated by using features extracted from the annotated training data.

3.1 Annotating training data in our approach

To label an ADL, users should specify its ADL class and its start and end points. Our sensor device is equipped with a camera thus making it superior to those without a camera as regards labeling tasks. Assume that the sensor data are acquired from a sensor device with a microphone and an accelerometer on a certain day. After the data acquisition, it is almost impossible for users to annotate the acquired data solely by listening to the recorded sound. Here, we introduce two approaches that deal with the problem. The first is a method where users annotate the data while watching video recordings captured by cameras embedded in the environment [17]. However, it is very expensive to install cameras in various rooms in the users' houses to track their activities. The second approach uses an experience sampling method that permits users to make annotations in real time [12]. In one example, users carry a PDA that is used as a timing device to trigger self-reported activity entries. Although this approach is inexpensive, users have to be continuously aware of the annotation process. This may result in biased or unrealistic data [14]. To solve these problems, [14] proposes a voice based annotation method, which permits users to make annotations easily in real time via a headset. However, real time annotation methods have another problem in that users cannot easily modify mistakenly created labels. During long periods of training data acquisition, users are certain to produce incorrect labels. However, in an environment with no video recording, it is almost impossible for users to review them solely by referring to captured non-visual sensor data. In contrast, because our device has a camera, users can make an accurate label set

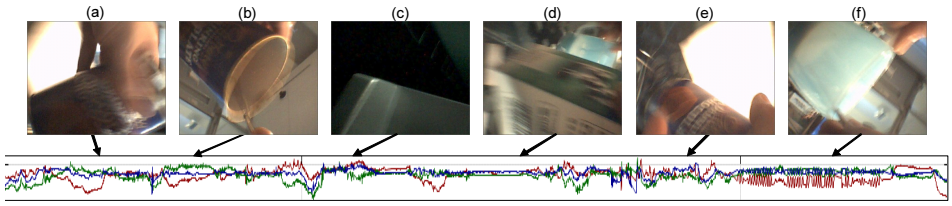


Fig. 2. Example camera and acceleration data for making cocoa.

by viewing image sequences recorded by the camera. Because the label set is used for training the models, it is very important to obtain training examples that are as accurate as possible [14].

As above, users annotate the sensor data while viewing an image sequence obtained by the camera and a chart of time-series acceleration data obtained from the device on our implemented annotation tool. By using the track bar of the tool, users can display an image captured at an arbitrary time on a panel component of the tool. Users can also play recorded image sequences and sound.

3.2 Classification features

We extract features from annotated training data that are used to model and recognize ADL classes. We deal with time-series data obtained from various types of sensors with different sampling rates. Thus, after extracting features from the sensor data for each sensor type in an appropriate size window, we combine them into one second windows with a 50% overlap and compute averages for each feature in each window. The 50% overlap has been employed successfully in past studies [2]. We perform ADL modeling and recognition by using a feature vector sequence generated by combining features extracted from all the sensors. Here, we describe how to extract features from each sensor data.

Visual features If we can detect which object the user is currently employing from the camera images, the information may be very useful for ADL recognition. In the following, after describing the characteristics of images captured by the camera and problems with the images such as privacy concerns, we use them as a basis for determining what kind of visual features are used to model ADL classes. Also, to achieve real time ADL recognition, we must extract the features from the image quickly. Note that we compute features for each captured image.

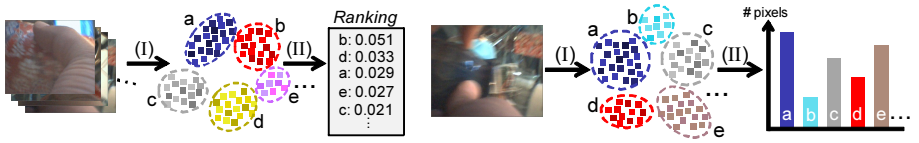
[**Characteristics of camera images**]
We introduce images captured by the camera in a data acquisition experiment. Fig. 2 shows a sequence of images and a chart of time-series acceleration data that were captured while a participant made cocoa. Fig. 2 (a) shows an image of when he took the cocoa tin from the cabinet, (b) shows an image of when he was spooning the cocoa powder, (c) shows an image of when he was moving toward the refrigerator, (d) shows an image of him holding a milk carton, (e) shows an image of him holding the cocoa tin prior to storing it in the cabinet,

and (f) shows an image of him stirring the cocoa. Images of objects captured by the wrist mounted camera have the following characteristics: (1) Objects are captured from various angles. (2) Most images show only a portion of the objects. (3) Objects seen in most images are blurred because of hand movement. (4) The brightness of an object can vary depending on the relationship between the lighting, camera, and object positions. Many studies try to detect objects from images while taking occlusion, rotation, scale, and blur into account [27, 18]. However, to detect an object from images captured from various angles, we must generate a model of the object from many images of objects. This may place a large burden on the end user because we must generate expensive models for each end user environment. Also, most existing object recognition algorithms are very costly if they are designed to achieve real time ADL recognition. On the basis of the above, we consider that we can leverage only rough visual information.

[**Problems with camera images**]
We describe two problems related to images captured by the wrist camera. The first concerns privacy. We assume that the sensor device sends such sensor data wirelessly as camera images to a host PC. Users may feel reluctant to send images related to their private lives wirelessly, e.g., those captured in a toilet. The second problem relates to communication traffic. Continuously transmitting raw images in real time occupies a communication band constantly. Our implemented device requires about 90 KB/sec for raw image transmission. This may also exhaust the device batteries very quickly. As a result, we determined that the device should send images consisting of small quantities of abstracted data.

[**Summary of our approach to visual feature extraction**]
Based on the above, we decided to extract rough visual features from an abstracted image sent from the device. The data volume of this image is small and the image is secure. Specifically, we use a color histogram of an image sent from the device. Some studies also achieve fast object recognition/tracking [30, 7] by comparing histograms and object models prepared in advance. In our approach, by using a histogram sent from the device, we simply count the number of pixels in the image (histogram) that are similar to a color characteristic of an ADL. For example, if a color of a cocoa tin is magenta, the number of pixels whose color is similar to magenta in an image may be useful for recognizing cocoa making activities. For each ADL, we obtain several characteristic colors from annotated training data in advance. For each characteristic color, we count the number of pixels in the histogram whose color is similar to the characteristic color. The result is used for the visual feature. Our purpose in using the histograms and characteristic colors is to achieve rough visual feature extraction with low communication and computation costs. In the following, we describe how to find the colors characteristic of each ADL, how to generate histograms, and how to compute features.

[**Finding characteristic colors of each ADL**]
We obtain the colors characteristic of each ADL in advance by using images of the annotated training data. Fig. 3 (a) shows the procedure. (I) We cluster all the color pixels in all the raw images labeled as the ADL into 64 clusters by



(a) (b)

Fig. 3. (a) Clustering pixels of all images of an ADL class and ranking the clusters (features) by their computed information gains, and (b) clustering color pixels of an image and constructing a histogram from the clusters.

using the k-means algorithm in the hue, saturation, and brightness (HSB) color space with a slight modification. Then, we compute the average color of each cluster. This procedure provides 64 representative colors of the ADL. Here, we focus on the HSB color space because it has a brightness axis. As mentioned above, the brightness of an object can change depending on the positional relationship of the lighting, camera, and object. Thus, we multiply the brightness values of the pixels by 0.5 to reduce the importance of the brightness axis. (II) From the obtained 64 candidate (representative) colors of the ADL, we extract the top-m candidate colors as the characteristic colors of the ADL. We rank the 64 candidate colors in terms of information gain. The information gain is usually used to find distinguishable attributes (features) of instances. The information gain of an attribute increases the better the attribute classifies the instances. We compute each attribute's information gain when distinguishing images (instances) of the ADL class from those of other ADL classes by using the attribute values of the images. In this case, each attribute corresponds to the number of pixels in an image whose colors are similar to each candidate color. (How to count the similar pixels is mentioned below.) We compute the information gain of each attribute by using the computed attribute values of the images and then rank the attributes by their information gains to obtain the top-m attributes as characteristic colors of the ADL. Note that, before obtaining the top-m attributes (colors), we remove colors that are similar to other higher ranked colors from the ranking.

Here, we provide an example. Assume that the color of a cocoa tin used in making cocoa is magenta and other ADLs do not include objects whose colors are similar to magenta. The number of magenta pixels in an image captured while making cocoa is large and so the information gain of the attribute (the number of magenta pixels) becomes high because the attribute contributes to distinguish the cocoa making images from the others. (An image in which the number of magenta pixels is above a certain threshold may correspond to cocoa making images.) See [34] for detailed explanation of computing the information gain. From the above procedure, we can obtain m characteristic colors for each ADL.

[Histogram generation]

Fig. 3 (b) shows the procedure. (I) The device reduces the color of an image to 64 colors simply by using the k-means algorithm to cluster the pixels in the image into 64 clusters. The representative color of each cluster corresponds to an aver-

age value for the colors in the cluster. (II) We compute a histogram of the image with 64 bins where each bin corresponds to one color of the 64 representative colors of the clusters. (The histogram is different from the commonly used color histogram.) The histogram includes only HSB data of each color (bin) and the number of pixels of the color in the reduced image. Comparing a characteristic color with colors of bins enables us to count the number of pixels in the image whose color is similar to the characteristic color. The histogram also permits us to compress an image into 64 pairs of 24 bit HSB data and 32 bit numeric data, i.e., (24+32)*64 = 3584 bit = 448 bytes. This enables us to reduce the communication traffic of our device, which captures images at 6 fps, to about 2.7 KB/sec. Moreover, we can solve the privacy problem because it is impossible to restore the original image from the histogram. Here, we use k-means clustering for color reduction in the device. We can process the algorithm at high speed by using a special purpose processing circuit [24]. We consider that all sensor data processing should be performed on special purpose circuits. Note that, in our prototype device shown in Fig. 1 (b), the host PC performs the color reduction and histogram generation offline. Also, to annotate training data, our approach requires raw captured images as described above. The device should be designed to store raw images in its flash memory card during training data acquisition periods. This enables users to safely transmit the data to the PC via the card.

[Visual feature extraction]
For each characteristic color, on the host PC, we count the number of pixels in the histogram whose color is similar to the characteristic color to model and recognize ADLs. The similarity is computed by using the Euclidean distances between the colors in the modified HSB color space. That is, we simply count the number of pixels whose similarity is smaller than a threshold th. The approach is identical to that used for the characteristic color extraction. Then, we normalize the result by dividing it by the dimensions of the image. The normalized result corresponds to the visual feature. That is, the number of visual features extracted from one image corresponds to the number of characteristic colors.

We employ this simple method because it requires low computational power. In fact, this method can extract visual features from a histogram in about 0.5 msec on a PC with a 2.4 GHz CPU by using 75 characteristic colors. We set $m = 5$ and $th = 15$ because they resulted in good performance in a preliminary experiment.

Sound features We extract features from sound that is emitted during ADLs that involve object use. For example, the sound of using a vacuum cleaner, tooth brushing, and running water may be useful for ADL recognition. We focus on the characteristic frequencies of such sounds. In [8], the Mel-Frequency Cepstral Coefficient (MFCC) is reported to be the best transformation scheme for environmental sound recognition. [5] achieves the highly accurate recognition of bathroom activities such as showering, flushing, and urination by using the MFCC. Thus, we decided to use the MFCC to recognize ADL related environmental sounds. Computing the MFCC is not expensive because it is based on

Fast Fourier Transform (FFT). Note that sound recorded by the microphone has problems related to data volume and privacy as well as the camera images. We thus extract sound features on the sensor device and send only them to the host PC. Also, the extraction of sound features 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 the sound data, we can obtain thirteen features.

Acceleration features We can extract features of postures and repetitive hand movements from acceleration data. For example, we can find a characteristic frequency in the acceleration data captured while the participant was stirring cocoa as shown in Fig. 2 (f). We extract features based on the FFT components of each 64 sample window acceleration data. We use the mean, energy, frequency-domain entropy, and dominant frequency as features. The mean can characterize the hand posture. For example, the hand posture during tooth brushing may have particular characteristics. The mean is the DC component of the FFT. The energy can be used to distinguish low intensity activities such as standing from high intensity activities such as walking [33]. The energy feature is calculated by summing the magnitudes of the squared discrete FFT components. For normalization, the sum was divided by the window length. Note that the DC component of the FFT is excluded from this summation. The frequency-domain entropy and dominant frequency can distinguish between repetitive motions with similar energy values. For example, the major FFT frequency components of stirring cocoa were between about 2 and 4 Hz in our experiment. Those of brushing teeth were between about 4 and 6 Hz. The frequency-domain entropy is calculated as the normalized information entropy of the discrete FFT component magnitudes [2]. The dominant frequency is the frequency that has the largest FFT component, and this component is three times larger than the average component of all the frequencies in this implementation. If there is no frequency that satisfies the conditions, we set the feature at zero. As above, we extract a total of twelve features from the 3-axis acceleration data.

Illuminance and direction features We use raw sensor data captured by the illuminometer and 3-axis digital compass directly as features. The digital compass captures the characteristic human orientation of each ADL. Assume that a user habitually brushes her teeth in front of a sink in her house. Her orientation during brushing may always be the same.

3.3 Classification methodology

We model ADL classes in advance by using annotated training data and the obtained feature vector sequence. We classify each feature vector in the test data into an estimated ADL class. The classification approaches used in machine learning are divided into two groups: one group uses discriminative techniques

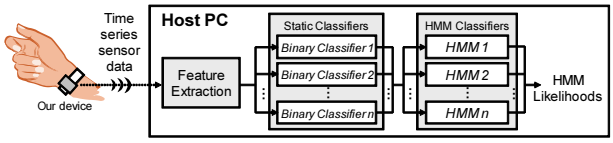


Fig. 4. Overview of the classification method.

that learn the class boundaries and the other uses generative techniques that model the conditional density functions of the data classes. The classification performance of the discriminative techniques, which find discriminant features of the classes, often outperform those of generative techniques. By contrast, handling missing data is often easier with the generative techniques. The ML community has shown increasing interest in a hybrid discriminative/generative approach that can combine the advantages of the two techniques [13, 26]. State-of-art activity recognition studies also achieve high accuracy by employing this approach [15, 16, 11]. In addition, because we deal with time-series data, incorporating the hidden Markov model (HMM), which is a generative model that can be used to model activities with temporal patterns [20, 14], into the hybrid approach, can improve the performance and smoothness of ADL recognition.

These facts provide our motivation for using the hybrid discriminative/generative approach with HMMs. Hybrid classification employs two main modules: static classifiers and HMM classifiers as shown in Fig. 4. The input of the first module is the extracted feature vector sequence. The first module consists of some discriminative binary classifiers trained with feature vector sequences. We build each binary classifier to recognize its corresponding ADL class. That is, the number of binary classifiers n corresponds to the number of ADLs the method learns. Each binary classifier computes the class probability for each feature vector in the feature vector sequence. That is, each binary classifier outputs the class probability sequence. The input of the second module consists of the class probability sequences computed by the n binary classifiers. The second module also comprises some HMM classifiers trained with a sequence of output class probabilities of the static classifiers. We also build each HMM to recognize its corresponding ADL class, that is, each HMM also outputs the likelihood of its corresponding ADL. The class with the highest likelihood is the classified class. We train the HMMs using the class probabilities of the static classifiers, which provide high levels of performance. The use of HMMs can smooth out sporadic errors of the static classifiers.

4 Data collection

For our experimental evaluation, we collected sensor data from participants by using our prototype device shown in Fig. 1 (b). Then, each participant annotated her own data using our annotation tool. In this study we learned the fifteen ADLs listed in Table 1. We selected these ADLs, which involve daily objects, by referring to ADLs dealt with by some reported ADL recognition studies.

Table 1. ADLs and their average duration (min).

	ADL	duration (min)		ADL	duration (min)
A	brush teeth	3.65	I	make juice	1.77
B	cook pasta	5.98	J	make tea	1.37
C	cook rice	4.33	K	practice aromatherapy	0.66
D	feed fish	0.40	L	take supplement	0.82
E	listen to music	1.69	M	vacuum	1.26
F	make cocoa	1.37	N	wash dishes	3.68
G	make coffee	1.63	O	water plants	0.27
H	make green tea	1.16			

4.1 Data set

The most natural data would be acquired from the normal daily lives of the participants. However, obtaining sufficient samples of such data is costly because researchers have to observe their normal daily lives. We collect sensor data by using a semi-naturalistic collection protocol [2] that permits greater variability in participant behavior than laboratory data. In the protocol, participants perform a random sequence of ADLs (obstacles) following instructions on a worksheet. The participants are relatively free as regards how they perform each ADL because the instructions on the the worksheet are not very strict, e.g., "vacuum the room" and "listen to an arbitrary track from a CD in the rack."

Data were collected from 10 participants who wore our prototype device in our experimental environments. The participants were workers (not researchers) in our laboratory. Because the features of the sensor data obtained from our device may vary depending on the environment, we collected sensor data in two environments: environment 1 and environment 2, and tested our method by using each data set. That is, we evaluated the test data obtained in environment 1 by using a classifier trained on training data also obtained in environment 1 and vice versa. Environment 1 is our home-like experimental environment [21]. We had equipped the environment with a cabinet, break time items, cooking utensils, etc. to emulate a home environment. In the experiment, we used objects originally installed in the environment. Also, four video cameras were fixed to the ceiling of the environment. The participants were familiar with environment 1 because they entered and left it many times every day. Because environment 2 is simply a room in our laboratory, we equipped it with new objects required for the ADLs for our experiment. We taught the participants the locations of the objects before undertaking data collection. Each data collection session included a random sequence of the fifteen ADLs listed in Table 1. We conducted fourteen sessions, which correspond to about two weeks data, in each environment. That is, each participant took part in a total of about three sessions in two environments. When performing the 'brush teeth,' 'cook rice,' and 'wash dishes' ADLs, they used sinks outside the environments. During data collection in environment 2, the participants used a timing device. The timing device is a PDA and can record the time at which its button is pushed. The participants can easily annotate collected data by referring to the recorded times. For example, they can push the button just before starting to vacuum.

The data obtained in this experiment were various and practical. Because the experiments were conducted from 9 a.m. to 6 p.m., images obtained under various light conditions are included. Also, because the experiment involved ten participants, their ways of performing the ADLs differed. For example, some participants made tea while standing and others while sitting. Of course, the participants' clothes, which were sometimes captured by the camera, were also different in different sessions. Furthermore, the experiment involved various kinds of objects such as those with complex textures, e.g., floral and arabesque patterns and translucent objects. Also, colors of some objects were similar with each other.

4.2 Labeling sensor data

The participants annotated their own collected data by using our tool. We asked them to select the start and end points of labels as they liked. After they had completed the task, we asked them to provide comments about the tasks. To enable us to compare our annotation method with conventional labeling in laboratory settings, the participants also annotated their data for environment 1 by watching video recordings captured by the cameras fixed to the ceiling. We call the label sets of environments 1 and 2 obtained by using the sensor data provided by our device *label sets 1A and 2*. We call the label set of environment 1 obtained using the video recordings provided by the fixed cameras *label set 1B*.

The average times needed to label the sensor data for one session were 44.1 and 36.4 min for set 1A and set 2, respectively. The participants annotated the data of environment 2 while referring to a printed list of times recorded by the timing device. Although we found no significant difference between two sets of results with a two-tail t-test ($p > 0.05$), all the participants commented that the timing device was useful. When end users annotate sensor data obtained during training data acquisition periods in their daily lives, they should determine their ADLs from sensor data obtained over long periods. Thus, the timing device may be useful to end users. The timing function should be embedded in our wristband device. The average labeling time for label set 1B was 27.0 min. While this approach is very costly, the average time was shorter than that of our device. Also, some participants commented that the images captured with our device can cause motion sickness. However, they also commented that labeling by using the images provided by our device was easier than they had thought because they could easily recognize routinely used objects in the images.

In label set 1A, there were three incorrect labels: a 'cook pasta' label did not include about half of a boil water activity in the ADL, a 'make juice' label ended while the participant was using a juicer, and a 'vacuum' label included part of another ADL. In label set 2, a participant forgot to label a 'feed fish' ADL. In label set 1B, there were two incorrect labels: a 'cook pasta' label did not include about half of a boil water activity and a 'listen to music' label did not cover the whole ADL. We could not find any significant differences between label sets 1A and 1B in terms of labeling accuracy. We asked the participants to correct these mistakes. In addition, each participant had a different labeling strategy. When labeling 'brush teeth' and 'wash dishes,' six participants selected

start and end points to include walking with related objects such as a dish rack from the environment to the sink. The labels of four other participants did not include this. In addition, some labels of the two participants did not include the ADL preparation time. For example, when making cocoa, the participants have to prepare a cup, a cocoa tin, and a milk carton. We should instruct end users in the same environment to establish a consensus on labeling strategy. We asked the participants in the minority to modify their labels in accordance with those of the majority.

5 Evaluation

We evaluated the performance of our method by using the annotated sensor data (label sets 1A and 2). We conducted a 'leave-one-session-out' cross validation evaluation. That is, we tested one session by using a classifier trained on thirteen other sessions. In this evaluation experiment, we used AdaBoost M1 and the C4.5 decision tree implemented on the Weka toolkit [34] as binary classifiers. AdaBoost is a boosting algorithm that combines weak classifiers to construct a strong classifier. We use a decision stump as a weak classifier.

5.1 Performance of our method

Table 2 lists the accuracies of the various recognition methods in some metrics. The AdaBoost+HMM (window) and C4.5+HMM (window) columns present precisions and recalls calculated based on feature windows (vectors). That is, precision is the ratio of the number of feature windows correctly classified into an ADL class to the number of all feature windows classified into the class. Recall is the ratio of the number of feature windows correctly classified into an ADL class to the number of actual feature windows of the class. C4.5+HMM, which uses C4.5 as a discriminative binary classifier and HMMs as a generative classifier, achieves relatively high accuracies for many ADLs and outperforms AdaBoost+HMM, which uses AdaBoost and HMMs. The accuracies of AdaBoost+HMM for certain ADLs such as 'feed fish,' 'take supplement,' and 'water plants' whose duration was short were zero. Because of the short duration of these ADLs, few feature windows were labeled as these ADLs. The AdaBoost algorithm combines weak classifiers, which usually ignore a minor class because they can achieve high accuracy by classifying all instances (windows) into a major class. This led to zero accuracies for these ADLs. [16] achieved the highly accurate recognition of primitive activities such as walking and sitting with a combination of AdaBoost and HMM. However, it was difficult to use this combination to recognize complex and/or brief ADLs.

The accuracies of C4.5+HMM for short duration ADLs such as 'feed fish,' 'practice aromatherapy,' 'take supplement,' and 'water plants' were also relatively low. This was caused by the head and foot margins of the labels. Handcrafted labels inevitably start and end with margins with no distinguishable feature window. For example, in 'take supplement,' the margin can correspond

Table 2. Averaged accuracies (precision / recall) of the recognition methods. The values are percentages.

	AdaBoost+HMM (window)		C4.5+HMM (window)		AdaBoost+HMM (instance)		C4.5+HMM (instance)	
	Env. 1	Env. 2	Env. 1	Env. 2	Env. 1	Env. 2	Env. 1	Env. 2
A: brush teeth	42.1/73.0	75.2/91.4	74.3/79.0	84.3/88.1	27.5/78.6	50.0/92.9	92.9/92.9	77.8/100
B: cook pasta	97.3/86.4	99.2/90.4	97.2/83.7	98.7/84.7	100/92.9	100/100	100/100	100/92.9
C: cook rice	76.2/93.1	79.1/96.0	88.3/85.1	88.3/87.5	54.2/92.9	66.7/100	81.2/92.9	87.5/100
D: feed fish	44.9/3.0	0.0/0.0	60.5/67.7	74.1/58.7	0.0/0.0	0.0/0.0	92.3/85.7	88.9/57.1
E: listen to music	86.7/81.2	50.2/65.3	84.7/90.1	58.4/82.4	80.0/85.7	45.0/64.3	93.3/100	72.2/92.9
F: make cocoa	0.0/0.0	87.9/72.0	74.6/64.4	85.2/76.4	0.0/0.0	84.6/78.6	91.7/78.6	92.9/92.9
G: make coffee	36.4/61.3	49.2/77.8	73.8/66.5	85.2/90.4	24.2/57.1	40.7/78.6	69.2/64.3	93.3/100
H: make green tea	16.4/16.6	69.9/7.0	50.1/13.8	34.5/72.9	18.8/21.4	100/7.1	40.0/14.3	45.8/84.6
I: make juice	86.1/72.9	27.0/53.1	79.7/78.2	76.4/70.4	92.3/85.7	17.9/50.0	93.3/100	92.3/85.7
J: make tea	0.0/0.0	72.1/47.8	24.5/70.3	72.7/42.3	0.0/0.0	60.0/42.9	47.6/71.4	75.0/42.9
K: practice aroma.	66.2/38.7	97.4/57.7	72.8/68.6	77.1/75.4	83.3/35.7	90.9/71.4	100/85.7	100/85.7
L: take supplement	0.0/0.0	0.0/0.0	50.8/69.2	73.7/62.4	0.0/0.0	0.0/0.0	70.6/85.7	90.9/71.4
M: vacuum	96.8/82.0	89.4/80.1	89.0/87.8	93.2/83.1	100/85.7	86.7/92.9	100/100	100/92.9
N: wash dishes	98.3/80.9	97.6/77.5	93.1/82.6	94.3/89.9	100/85.7	100/92.9	93.3/100	93.3/100
O: water plants	100/88.4	0.0/0.0	84.5/92.4	40.5/59.8	100/100	0.0/0.0	100/100	100/71.4
Average	56.5/51.8	59.6/54.4	73.2/73.3	75.8/75.0	52.0/54.8	56.2/58.1	84.4/84.8	87.3/84.7

to a time duration where a participant walks from a chair to a cabinet to get a pill case. Feature windows involved in these margins may be wrongly classified. For an ADL with a short duration, the ratio of the time duration of its margins to those of the whole label is large and so the accuracy becomes relatively low. However, in C4.5+HMM, most feature windows in each label were correctly classified. For ease of understanding, the AdaBoost+HMM (instance) and C4.5+HMM (instance) columns in Table 2 show instance based accuracies, which are computed based on majority voting. That is, we compute the accuracies based on the strategy: In an ADL instance, we regard the instance itself to be classified into the majority vote of the recognition results of each feature window included in the instance. While our recognition method depends on environmental conditions, C4.5+HMM achieved high accuracies in both environments.

Distinguishing between 'make green tea' and 'make tea' was difficult in both environments as also described in Table 3, which shows the confusion matrices of C4.5+HMM in environments 1 and 2. This is because the motions involved in making green tea and making tea are the same, and most of the objects used in these ADLs such as a kettle and a cup are also the same. In addition, each side of the tea caddy in environment 1 is a single color; red or gold. In many sessions, the camera on our device could capture only the red colored portion of the caddy depending on how the caddy was held. Because the green tea caddy is also red, it was difficult to distinguish these ADLs in environment 1. Also, the recognition of such ADLs as 'feed fish,' 'take supplement,' and 'water plants,' which involve small numbers of objects and do not have characteristic sound or hand activities sometimes failed. In particular, when the colors of objects involved in the ADLs were similar to those of objects used in other ADLs, it was difficult to distinguish between these ADLs. In environment 2, for example, the color of the fish food tin was similar to that of a kettle used for making tea.

Table 3. Instance based confusion matrices of C4.5+HMM in environments 1 and 2.

Env. 1	A: brush teeth	B: cook pasta	C: cook rice	D: feed fish	E: listen to music	F: make cocoa	G: make coffee	H: make green tea	I: make juice	J: make tea	K: practice aroma.	L: take supplement	M: vacuum	N: wash dishes	O: water plants
A	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0
B	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	0	13	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	12	0	0	1	0	0	0	0	0	0	0	0
E	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	11	1	0	1	0	0	1	0	0	0
G	0	0	0	1	0	0	9	1	0	1	0	2	0	0	0
H	0	0	0	0	0	0	2	2	0	9	0	1	0	0	0
I	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0
J	0	0	0	0	0	1	0	2	0	10	0	1	0	0	0
K	0	0	1	0	1	0	0	0	0	0	12	0	0	0	0
L	0	0	1	0	0	0	0	0	0	1	0	12	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14

Env. 2	A: brush teeth	B: cook pasta	C: cook rice	D: feed fish	E: listen to music	F: make cocoa	G: make coffee	H: make green tea	I: make juice	J: make tea	K: practice aroma.	L: take supplement	M: vacuum	N: wash dishes	O: water plants
A	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	13	0	0	1	0	0	0	0	0	0	0	0	0	0
C	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0
D	1	0	0	8	2	0	0	3	0	0	0	0	0	0	0
E	0	0	1	0	13	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	13	0	1	0	0	0	0	0	0	0
G	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	11	0	2	0	0	0	0	0
I	1	0	0	0	1	0	0	0	12	0	0	0	0	0	0
J	0	0	0	0	0	1	1	5	0	6	0	1	0	0	0
K	1	0	0	0	1	0	0	0	0	0	12	0	0	0	0
L	1	0	0	0	0	0	0	2	1	0	0	10	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	13	1	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
O	0	0	1	1	0	0	0	2	0	0	0	0	0	0	10

5.2 Contributions of each sensor

Table 4 (a) lists the accuracies of the C4.5+HMM recognition results in various sensor combinations. For example, the 'only camera' row shows instance based accuracies under a condition where the accuracies were computed on the basis of only features extracted from the camera sensor data. Also, the 'w/o camera' row shows accuracies under a condition where the accuracies were computed without features extracted from the camera sensor data. Surprisingly, we could achieve very high accuracies (about 75%) with just the camera. We also found that using only a camera could achieve almost the same accuracies as when combining an accelerometer, a microphone, a direction sensor, and an illuminometer. The camera played a significant role in ADL recognition when using our device. Sensors with a high contribution were the camera, accelerometer, and microphone in that order. The illuminometer and digital compass barely contributed to the recognition and sometimes even decreased the accuracy.

Table 4 (b) lists the accuracies of each ADL when we use only the camera and only the accelerometer. The accuracies of most ADLs when using only the camera were high. However, it was difficult to distinguish between 'make tea' and 'make green tea' in environment 1 because the colors of the objects involved in these ADLs were similar. Also, the accuracies for 'cook pasta' and 'listen to music,' which were characterized by their sound features, were not very high. With only the accelerometer, the accuracies of 'brush teeth,' 'cook rice,' and 'wash dishes' were relatively high. However, without the camera, it was difficult to distinguish between these ADLs with high accuracy because all three ADLs, which involved long periods of walking (and the sound of running water), were similar. Moreover, without a camera, it is very difficult to distinguish such ADLs as 'feed fish,' 'practice aromatherapy,' 'take supplement,' and 'water plants.'

Table 4. (a) instance based average accuracies (precision/recall) of C4.5+HMM in various sensor combinations and (b) instance based average accuracies of C4.5+HMM for each ADL with only camera features and with only accelerometer features.

(a)

Sensor	Condition	Env.	Accuracy
camera	only	1	76.7/73.2
		2	75.1/71.8
	w/o	1	77.7/75.2
		2	71.8/67.6
micro-phone	only	1	28.3/32.9
		2	21.8/28.6
	w/o	1	84.9/83.3
		2	83.8/81.0
accele-rometer	only	1	48.5/44.3
		2	47.3/43.8
	w/o	1	82.1/80.5
		2	84.9/79.5
illumi-nometer	only	1	0.1/6.7
		2	0.4/6.7
	w/o	1	81.7/82.4
		2	89.6/88.0
digital compass	only	1	23.1/21.9
		2	10.8/10.0
	w/o	1	85.9/84.8
		2	89.9/87.0

(b)

	only camera		only accelerometer	
	Env. 1	Env. 2	Env. 1	Env. 2
A: brush teeth	75.0/85.7	41.2/50.0	71.4/71.4	59.1/92.9
B: cook pasta	88.9/57.1	31.2/35.7	39.3/78.6	47.1/57.1
C: cook rice	70.0/100	72.2/92.9	50.0/78.6	54.5/42.9
D: feed fish	90.9/76.9	88.9/57.1	100/7.1	0.0/0.0
E: listen to music	72.2/92.9	62.5/71.4	53.8/50.0	21.2/50.0
F: make cocoa	66.7/42.9	84.6/78.6	23.5/28.6	61.5/57.1
G: make coffee	84.6/78.6	73.3/78.6	18.8/42.9	23.6/92.9
H: make green tea	35.0/50.0	52.4/84.6	7.7/7.1	0.0/0.0
I: make juice	76.5/92.9	81.8/64.3	100/78.6	60.0/42.9
J: make tea	27.8/35.7	83.3/71.4	9.5/14.3	28.6/28.6
K: practice aroma.	100/71.4	100/92.3	0.0/0.0	100/14.3
L: take supplement	77.8/50.0	81.8/64.3	0.0/0.0	0.0/0.0
M: vacuum	100/78.6	85.7/85.7	86.7/92.9	78.6/78.6
N: wash dishes	85.7/85.7	87.5/100	66.7/71.4	75.0/85.7
O: water plants	100/100	100/50.0	100/42.9	100/14.3
Average	76.7/73.2	75.1/71.8	48.5/44.3	47.3/43.8

These ADLs have few distinguishable features other than visual features. We consider that, without the camera, it is difficult to recognize the complex ADLs studied here.

From the above results, we consider that the wrist is a good place on the body to attach a single sensor device designed to capture ADLs that involve object use. Hand posture (mean) contributed to the recognition of many ADLs such as 'brush teeth' and 'make juice'. However, it is difficult to capture the features when using body locations other than the wrist. Without the mean features, instance based precision and recall decreased to 82.3 and 79.5 in environment 1. In addition, the wrist worn camera, which was the best contributor, can easily capture hand manipulated objects. [22] uses a shoulder mounted robot with a camera to recognize and record hand activities such as operating a keyboard and operating a calculator. However, the robot has to control the direction of its camera to track the hand.

6 Related work

We introduce related work that relates to vision based wearable sensing. [9] achieves gait analysis and floor recognition by using a shoe mounted camera and accelerometers. Floor recognition permits us to know the user's location. [35] recognizes kitchen activities by using RFID tags attached to kitchen objects and a camera that overlooks the kitchen counter. An RFID reader bracelet worn on a user's wrist and the camera detect the use of the objects. [6] uses a camera and a microphone attached to a chest strap to detect location related events such as entering an office, kitchen, or courtyard. [29] uses two hat mounted cameras to determine user's actions in a game, e.g., aiming a gun. On the other hand, we

use a wrist mounted camera, a microphone, an accelerometer, an illuminometer, and a digital compass to recognize ADLs that involve object use by capturing various characteristic features of manually used objects. Also, [1] uses a wrist mounted camera to realize a virtual keyboard by tracking the fingers with the camera.

There exists some works that recognize activities by using a single sensor device embedded in a home. These works also attempt to reduce costs of sensor deployment. For example, HydroSense [10] employs a water pressure sensor to understand activities that involve water use.

7 Conclusion and future work

We implemented a prototype wristband sensor device to recognize ADLs that involve the manual use of objects. The device is equipped with a camera, a microphone, an accelerometer, an illuminometer, and a digital compass to capture various characteristic features of object use. This device enables us to recognize various kinds of ADLs that existing wearable sensor devices cannot recognize without environment embedded sensors. In the experiments, we confirmed that the incorporation of a camera could achieve highly accurate ADL recognition.

As a part of our future work, we plan to solve the problems thrown up by the experiment. In both the environments, it was difficult to distinguish between such ADLs as 'make green tea' from 'make tea' that involve the same hand activities and the similar colored objects. To cope with such scalability problems, we should extract more detailed features such as SIFT features [18] from 'good' images, e.g., those including logos, while taking account of privacy concerns and communication costs. Furthermore, our ML-based approach cannot deal with situations where residents replace objects, e.g., residents frequently replace milk cartons. Because the types of milk that a family regularly purchases may be limited, we should instruct end users to prepare ADL training data that include various product types of such objects or we should realize an object replacement detection technique to induce users to prepare new training data.

We also plan to develop a new wristband sensor device that works without a laptop. The device permits us to capture sensor data in real environments and evaluate the performance of our method by using the data.

Acknowledgments. The authors would like to thank Dr. Takuya Yoshioka for the helpful comments and discussions.

References

1. F. Ahmad and P. Musilek, "A keystroke and pointer control input interface for wearable computers," *Proc. PerCom 2006*, pp. 2–11, 2006.
2. L. Bao and S.S. Intille, "Activity recognition from user-annotated acceleration data," *Proc. Pervasive 2004*, pp. 1–17, 2004.
3. M. Blum, A.S. Pentland, and G. Troster, "Insense: Interest-based life logging," *IEEE Multimedia*, 13(4), pp. 40–48, 2006.

4. C.V. Bouten, et al., "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Trans. on Bio-Medical Engineering*, 44(3), pp. 136–147, 1997.
5. J. Chen, A.H. Kam, J. Zhang, N. Liu, and L. Shue, "Bathroom activity monitoring based on sound," *Proc. Pervasive 2005*, pp. 47–61, 2005.
6. B. Clarkson, K. Mase, and A. Pentland, "Recognizing user context via wearable sensors," *Proc. ISWC 2000*, pp. 69–75, 2000.
7. D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. on Pattern Analysis Machine Intelligence*, 25(5), pp. 564–577, 2003.
8. M. Cowling, "Non-speech environmental sound recognition system for autonomous surveillance," Ph.D. Thesis, Griffith University, Gold Coast Campus, 2004.
9. P. Fitzpatrick and C.C. Kemp, "Shoes as a platform for vision," *Proc. ISWC 2003*, pp. 231–234, 2003.
10. J.E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, and S.N. Patel, "HydroSense: Infrastructure-mediated single-point sensing of whole-home water activity," *Proc. Ubicomp 2009*, pp. 235–244, 2009.
11. T. Huynh and B. Schiele, "Towards less supervision in activity recognition from wearable sensors," *Proc. ISWC 2006*, pp. 3–10, 2006.
12. S.S. Intille, E.M. Tapia, J. Rondoni, J. Beaudin, C. Kukla, S. Agarwal, L. Bao, and K. Larson, "Tools for studying behavior and technology in natural settings," *Proc. UbiComp 2003*, pp. 157–174, 2003.
13. T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," *Proc. Advances in Neural Information Processing Systems 11*, pp. 487–493, 1999.
14. T.V. Kasteren, A. Noulas, G. Englebienne, and B. Krose, "Accurate activity recognition in a home setting," *Proc. UbiComp 2008*, pp. 1–9, 2008.
15. J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," *Proc. IJCAI 2005*, pp. 766–772, 2005.
16. J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," *Proc. Pervasive 2006*, pp 1–16, 2006.
17. B. Logan, J. Healey, M. Philipose, E.M. Tapia, and S.S. Intille, "A long-term evaluation of sensing modalities for activity recognition," *Proc. UbiComp 2007*, pp. 483–500, 2007.
18. D.G. Lowe, "Distinctive image features from scale–invariant keypoints," *Int'l Journal on Computer Vision*, 60(2), pp. 91–110, 2004.
19. P. Lukowicz, H. Junker, et al., "WearNET: a distributed multi-sensor system for context aware wearables," *Proc. UbiComp 2002*, pp. 361–370, 2002.
20. P. Lukowicz, J. Ward, H. Junker, M. Stager, G. Troster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," *Proc. Pervasive 2004*, pp. 18–32, 2004.
21. T. Maekawa, Y. Yanagisawa, Y. Kishino, K. Kamei, Y. Sakurai, and T. Okadome, "Object-blog system for environment-generated content," *IEEE Pervasive Computing*, 7(4), pp. 20–27, 2008.
22. W.W. Mayol and D.W. Murray, "Wearable hand activity recognition for event summarization," *Proc. ISWC 2005*, pp. 122–129, 2005.
23. A. Mihailidis, B. Carmichael, and J. Boger, "The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home," *IEEE Trans. on Info. Tech. in BioMedicine*, 8(3), pp. 238–247, 2004.
24. S. Morikawa, K. Ito, and T. Shibata, "A k-means VLSI processor and its application to autonomous area segmentation in images," *IEIC Technical Report*, 106(342), pp. 19–24, 2006.
25. M. Philipose, K.P. Fishkin, and M. Perkowitz, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, 3(4), pp. 50–57, 2004.
26. R. Raina, Y. Shen, A.Y. Ng, and A. McCallum,"Classification with hybrid generative/discriminative models," *Proc. Advances in Neural Information Processing Systems 16*, 2003.
27. B. Schiele and L.C. James, "Object recognition using multidimensional receptive field histograms," *Proc. European Conference on Computer Vision*, pp. 610–619, 1996.
28. Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa, "Propagation networks for recognition of partially ordered sequential action," *Proc. CVPR 2004*, 2, pp. 862–869, 2004.
29. T. Starner, B. Schiele, and A. Pentland, "Visual contextual awareness in wearable computing," *Proc. ISWC 1998*, pp. 50–57, 1998.
30. M.J. Swain and D.H. Ballard, "Color indexing," *Int'l Journal of Computer Vision*, 7, pp. 11–32, 1991.
31. E.M. Tapia, S.S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," *Proc. Pervasive 2004*, pp. 158–175, 2004.
32. E.M. Tapia, S.S. Intille, and K. Larson, "Portable wireless sensors for object usage sensing in the home: challenges and practicalities," *Proc. AmI 2007*, pp. 19–37, 2007.
33. G. Welk and J. Differding, "The utility of the Digi-Walker step counter to assess daily physical activity patterns," *Medicine & Science in Sports & Exercise*, 32(9), S481–S488, 2000.
34. I.H. Witten and E. Frank, "Data Mining: Practical machine learning tools and techniques," 2nd Edition, Morgan Kaufmann, 2005.
35. J. Wu, A. Osuntogun, et al., "A scalable approach to activity recognition based on object use," *Proc. ICCV 2007*, pp. 1–8, 2007.