Pervasive and Mobile Computing 7 (2011) 287–298

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Recognizing multi-user activities using wearable sensors in a smart home

Liang Wang [a,b], Tao Gu [b,*], Xianping Tao [a], Hanhua Chen [c,b], Jian Lu [a]

[a] State Key Laboratory for Novel Software Technology, Nanjing University, China
[b] Department of Mathematics and Computer Science, University of Southern Denmark, Denmark
[c] School of Computer Science and Technology, Huazhong University of Science and Technology, China

ARTICLE INFO

Article history:
Available online 5 December 2010

Keywords:
Sensor-based human activity recognition
Multi-user activity recognition
Probabilistic model
Wireless sensor network

ABSTRACT

The advances of wearable sensors and wireless networks offer many opportunities to recognize human activities from sensor readings in pervasive computing. Existing work so far focuses mainly on recognizing activities of a single user in a home environment. However, there are typically multiple inhabitants in a real home and they often perform activities together. In this paper, we investigate the problem of recognizing multi-user activities using wearable sensors in a home setting. We develop a multi-modal, wearable sensor platform to collect sensor data for multiple users, and study two temporal probabilistic models—Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF)—to model interacting processes in a sensor-based, multi-user scenario. We conduct a real-world trace collection done by two subjects over two weeks, and evaluate these two models through our experimental studies. Our experimental results show that we achieve an accuracy of 96.41% with CHMM and an accuracy of 87.93% with FCRF, respectively, for recognizing multi-user activities.

1. Introduction

The problem of recognizing human actions and activities using a video camera has been studied in computer vision since a decade ago [1,2]. With the availability of low-cost sensors and the advancement in wireless sensor networks, researchers in pervasive computing have recently become interested in deploying various sensors to collect observations, and recognizing activities based on these observations. This in turn supports many potential applications such as monitoring activities of daily living (ADLs) [3] for the elderly or people with cognitive impairments [4].

By capturing useful low-level features, such as human motion, living environment and human-to-environment interactions, sensors show great potential to aid with human activity recognition. However, recognizing human activities using sensors is challenging because sensor data are inherently noisy and human activities are complex in nature. Most of the existing work focuses on recognizing single-user activities [5–12]. However, humans are social beings, and they often form a group to complete specific tasks. Activities that involve multiple users collaboratively or concurrently are common in our daily lives, especially in a home setting. For example, family members often watch TV together, and prepare meals in a collaborative way. From a social psychology point of view, people often form groups to perform certain activities collectively not only because they share socially relevant features but also because they interact and rely on each other to achieve specific goals. Among others, two distinctive features – social interdependence and task interdependence – are salient [13]. Foremost, they are socially interdependent because they rely on one another for feelings of connectedness and positive

* Corresponding author.
E-mail addresses: wangliang@ics.nju.edu.cn (L. Wang), gu@imada.sdu.dk (T. Gu), txp@nju.edu.cn (X. Tao), hhchen@imada.sdu.dk (H. Chen), lj@nju.edu.cn (J. Lu).

1574-1192/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2010.11.008

emotional outcomes; then, they are task interdependent because their mastery of material outcomes depend on working together to perform some collective task. This is especially demonstrated in a home environment among family members whether affiliated by consanguinity, affinity, or co-residence.

Recognizing multi-user activities using wearable sensors is more challenging than recognizing single-user activities. The main challenges are how to design appropriate wearable sensors to capture user interactions, and how to model interacting processes and perform inferences. In this work, we develop a wearable sensor platform to capture the observations of each user and the interactions among multiple users. Using this platform, we conduct a real-world activity trace collection done by two subjects over a period of two weeks in a smart home. We then study two temporal probabilistic models – Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF) – to model interacting processes which involve multiple users and recognize multi-user activities. Both models are multi-chained variants of their basic models; CHMM couples HMM with temporal, asymmetric influences while FCRF couples CRF with probabilistic dependencies between co-temporal label sequences. We evaluate and compare both models through real system experiments, and analyze their effectiveness in modeling multi-user activities from sensor data.

In summary, the paper makes the following contributions.

- To the best of our knowledge, this work is the first formal study of two temporal probabilistic models (CHMM and FCRF) in recognizing multi-user activities based on wearable sensors in a smart home environment.
- We develop a multi-modal, wearable sensor platform to capture the observations of each user and the interactions between users, and conduct a real-world trace collection in a smart home.
- We conduct extensive experiments to evaluate our models and analyze their effectiveness in a wearable sensor based setting.

The rest of the paper is organized as follows. Section 2 discusses the related work. In Section 3, we present the design of our wearable sensor platform. Section 4 describes our proposed activity models, and Section 5 reports our empirical studies. Finally, Section 6 concludes the paper.

2. Related work

Recently, researchers have become interested in recognizing activities based on sensor readings. Recognition models are typically probabilistic based, and they can be categorized into static and temporal classification. Typical static classifiers include naïve Bayes used in [6,7], decision tree used in [6,7], and k-nearest neighbor (k-NN) used in [6]. In temporal classification, state–space models are typically used to enable the inference of hidden states (i.e., activity labels) given the observations. We name a few examples here: Hidden Markov Model (HMM) used in [8,9], Dynamic Bayesian Network (DBN) used in [10] and Conditional Random Field (CRF) used in [14,12]. The variants of CRF have been used to model complex activities of a single user. For example, Wu et al. [15] applied Factorial Conditional Random Field (FCRF) to model concurrent activities.

There is some existing work on recognizing group activities and modeling interacting processes. Gong et al. [16] developed a dynamically multi-linked HMM model to interpret group activities. They also compared their methods with Multi-Observation HMM, Parallel HMM, and Coupled HMM. Nguyen et al. [17] employed hierarchical HMM for modeling the behavior of each person and the joint probabilistic data association filters for data association. Park et al. [18] presented a synergistic track- and body-level analysis framework for multi-person interaction and activity analysis in the context of video surveillance. An integrated visual interface for gestures and behavior was designed in [19] as a platform for investigating visually mediated interaction with a video camera. However, their system only tackled simple gestures like waving and pointing. Du et al. [20] proposed a new DBN model structure with state duration to model human interacting activities (involving two users) using video cameras, combining the global features with local ones. Choudhury et al. [21] modeled the joint turn-taking behavior as a mixed-memory Markov model that combines the statistics of the individual subjects' self-transitions and the partners' cross-transitions. Lian et al. [22] used FCRF to conduct inference and learning from patterns of multiple concurrent chatting activities based on audio streams. Oliver et al. [23] proposed CHMM to model user interactions in video. Oliver et al. [24] also proposed Layered Hidden Markov Models (LHMMs) to diagnose states of a user's activity based on data streams from video, audio, and computer (keyboard and mouse) interactions.

With regard to modality, most of the work mentioned above employed video data only. Audio and other modal data are less frequently used together. One possible reason is that it is rather hard to determine hidden parameters of HMMs in the case of multi-modal group action or activity recognition, where features from each modal are concatenated to define the observation model [25]. Wyatt et al. [20] presented a privacy-sensitive DBN-based unsupervised approach to separating speakers and their turns in a multi-person conversation. They addressed the problem of recognizing sequences of human interaction patterns in meetings with two-layer HMM using both audio and video data. Unlike other work, their framework explicitly modeled actions at different semantic levels from individual to group level at the same time scale.

However, to the best of our knowledge, there is no formal study on recognizing multi-user activities using wearable sensors in a smart home environment. Although in [7], the data collection was done by two subjects (i.e., a couple) in an instrumented home, the dataset contains annotated data for the male user only and the work reported in the paper focuses on this single user. An interesting work was done by Lin et al. [26], where they deployed various kinds of sensors in a home environment and proposed a layered model to learn multiple users' preferences based on sensor readings. However, their focus is on learning of preference models of multiple users, i.e., relationships among users as well as dependency

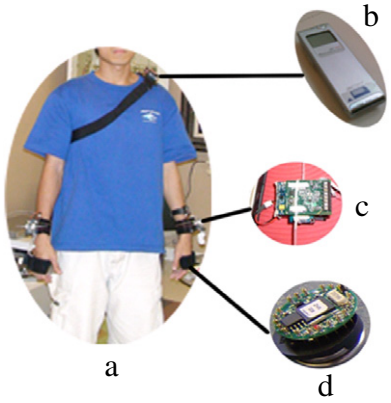

Fig. 1. (a) Wearable sensor set, (b) audio recorder, (c) iMote2 with ITS400, (d) RFID wristband reader.

between services and sensor observations, not on recognizing their activities. A series of research work has been done in the CASAS smart home project at WSU [27] to serve the residents in a smart home. Singla et al. [28] addressed the problem of recognizing the independent and joint activities among multiple residents in smart environments using a single HMM model. However, they only explored the use of infrastructure sensors deployed in a smart apartment which limits the information they could get on the activities of the residents. Besides, the association of the sensor data and the people who generated it is done manually. Data association is trivial for wearable sensors but difficult for infrastructure sensors, which indicates the advantage of our sensor platform. In our earlier work, we have explored a pattern mining approach [29] to multi-user activity recognition. In this paper, we use a machine learning approach and study two temporal models to model interacting processes involving multiple users.

3. Multi-modal wearable sensor platform

Multi-modal wearable sensors have been successfully applied in recognizing the activities of daily living by many researchers. Accelerometers have been used to capture the bodily movements of the user [6,9]. Recent advances in wearable RFID readers have enabled the recognition of activities by capturing object use [8]. Wearable audio devices have also been used in activity recognition [21].

We built a multi-modal wearable sensor platform using off-the-shelf sensors, as shown in Fig. 1. This platform measures user movement (i.e., both hands), user location, human–object interaction (i.e., objects touched and sound), human-to-human interaction (i.e., voice) and environmental information (i.e., temperature, humidity and light). To capture acceleration data, we used a Crossbow iMote2 IPR2400 processor/radio board with an ITS400 sensor board, as shown in Fig. 1(d). The ITS400 sensor board also measures environmental temperature, humidity and light. To capture object use, we built a customized RFID wristband reader which incorporates a Crossbow Mica2Dot wireless mote, a Skyetek M1-mini RFID reader and a Li-polymer rechargeable battery. The wristband is able to detect the presence of a tagged object within the range of 6 to 8 cm. The RFID wristband reader is also able to capture object interaction, i.e., objects passing from one user to another. To capture vocal interaction among users, initially, we used the built-in microphone sensor on an ITS400 sensor board. However, it fails to support a high sampling rate due to the bandwidth limitation of a wireless link. It turns out that we use a commercial audio recorder with a maximum sampling rate of 44.1 kHz to record audio data, as shown in Fig. 1(b). In addition, user location is detected in a simple way that a UHF RFID reader is located in each room to sense the proximity of a user wearing a UHF tag. To determine user identity, the device IDs of each iMote2 set and RFID reader are logged and bound to a specific user. However, it is not possible to determine audio identity precisely as the audio information may come from other users who live in the same physical space.

The sampling rate of the RFID readers is set to 2 Hz, the sampling rate of the 3-axis accelerometer in each iMote2 is set to 128 Hz, and the sampling rate of the audio recorder is set to 16 kHz. When a user performs activities, the acceleration readings from each iMote2 set are transmitted wirelessly to a local server which runs on a laptop PC with an iMote2 IPR2400 board connected through its USB port. When a user handles a tagged object, the RFID wristband reader scans the tag ID and sends it wirelessly to another server that can map the ID to an object name. This server runs on a Linux-based laptop PC with a MIB510CA serial interface board and a Mica2Dot module connected through its serial port. In addition, human voice and environmental sound are recorded by the audio recorder. All the sensor data with timestamps are logged separately, and will be merged into a single text file as the activity trace for each user.

4. Multi-chained temporal probabilistic models

In this section, we first describe our problem statement, then present two multi-chained temporal probabilistic activity models to model and recognize multi-user activities.

4.1. Problem statement

We formulate our multi-user activity recognition problem as follows. We assume that there are a number of training datasets, where each training dataset corresponds to each user. Each training dataset O consists of T observations $O = \{o_1, o_2, \ldots, o_T\}$ associated with activity labels $\{A_1, A_2, \ldots, A_m\}$, where there are m multi-user activities. For a new sequence of observations corresponding to a user, our objective is to train an appropriate activity model that can assign each new observation with the correct activity label.

4.2. Feature extraction

After obtaining sensor readings, we first need to extract appropriate sensor features. We convert all the sensor readings to a series of *observation vectors* by concatenating all of the data observed in a fixed time interval which is set to one second in our experiments. Different types of sensors require different processing to compute various features.

For acceleration data, we compute five features—DC mean, variance, energy, frequency-domain entropy, and correlation. The DC mean is the mean acceleration value in a time interval. Variance is used to characterize the stability of a signal. Energy captures data periodicity, and is computed as the sum of the squared discrete FFT component magnitudes of a signal. Frequency-domain entropy helps to discriminate activities with similar energy values, and is computed as the normalized information entropy of the discrete FFT component magnitudes of a signal. Correlation is computed for every two axes of each accelerometer and all pair-wise axis combinations of two different accelerometers. This feature aims to find out the correlation among different axes of any two accelerometers. We compute four features (DC mean, variance, energy, and frequency-domain entropy) for each axis of the two three-axis acceleromters. The correlation is computed for each of the fifteen pairs of axes. In total, there are 39 features extracted from acceleration data. To improve the reliability of our feature extraction, a cubic spline approach can be used to fill out the missing values, and a low-pass filter can be used to remove the outliers in the data.

For audio data, we compute both time-domain and frequency-domain features. The time-domain features measure the temporal variation of an audio signal, and consist of three features. The first one is the standard deviation of a reading in a time interval, normalized by the maximum reading in the interval. The second one is the dynamic range defined as $(max - min)/max$, where min and max represent the minimum and maximum readings in the interval. The third is Zero-Crossing Rate (ZCR) which measures the frequency content of a signal and is defined as the number of time-domain zero crossings in a time interval. In the frequency domain, we compute two features—centroid (the midpoint of the spectral power distribution) and bandwidth (the width of the range of frequencies that a signal occupies). In total, there are five features extracted from the audio data. A similar technique used for acceleration data can be applied to audio data to improve the reliability of feature extraction.

For RFID reading or location information, we use object name or location name directly as features. For each RFID wristband reader, we choose the first object in a one-second time interval since a user is unlikely to touch two or more objects in such a short interval. If no RFID reading is observed or in the presence of a corrupted tag ID, the value will be set to NULL. There are three features in total—two features for object usage of both hands and one feature for the user's location.

The above process generates a 47-dimensional *observation vector* every second. We then transform these observation vectors into feature vectors. A feature vector consists of many feature items, where a feature item refers to a feature name–value pair in which a feature can be numeric or nominal. We denote a numeric feature as $numfeature_i$. Suppose its range is $[x, y]$ and an interval $[a, b]$ (or in other forms, $(a, b]$, $[a, b)$, or (a, b)) is contained in $[x, y]$. We call $numfeature_i@[a, b]$ a numeric feature item, meaning that the value of $numfeature_i$ is limited inclusively between a and b. We denote a nominal attribute as $nomfeature_j$. Suppose its range is $\{v_1, v_2, \ldots, v_n\}$, we call $nomfeature_j@v_k$ a nominal feature item, meaning the value of $nomfeature_j$ is v_k.

The key step of transformation is to discretize numeric features. We follow the entropy-based discretization method which partitions a range of continuous values into a number of disjoint intervals such that the entropy of the partition is minimal [30]. The class information entropy of candidate partitions is used to select binary boundaries, and the minimal entropy criterion is then used to find multi-level cuts for each attribute.

The discretization method partitions 44 numeric feature values into a total of 484 disjoint intervals. Then we can directly combine the feature name and its interval into a numeric feature item. For the nominal feature, the feature name and its value are combined as a nominal feature item. For the *LEFTOBJ* and *RIGHTOBJ* features, we merge them into one feature by computing *LEFTOBJ* ∪ *RIGHTOBJ* without losing any object during the user–object interaction due to the user's handedness. In our current sensor setting, we have a total of 574 feature items. They are indexed by a simple encoding scheme and will be used as the inputs of our probabilistic models described in the next section.

4.3. Coupled hidden Markov model

After feature extraction, we obtain a sequence of feature vectors for each user, where a feature vector $f = \{f_1, f_2, \ldots, f_T\}$ is associated with activity labels $\{A_1, A_2, \ldots, A_m\}$. To model a single-user sequence, HMM is often used, and it consists of a hidden variable and an observable variable at each time step. In this case, the hidden variable is an activity label, and the observable variable is a feature vector. For multiple sequences of observations corresponding to multiple users, we can

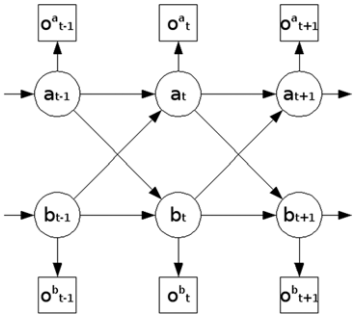

Fig. 2. Structure of CHMM.

factorize the basic HMM into multiple channels and couple HMMs with temporal influences to model interacting processes. The coupling bridges hidden variables with conditional probability of transition. CHMM was originally introduced in [31] for modeling interacting processes. The CHMM models the causal influences between the hidden state variables of different chains. The semantics of the model is clear and easy to understand. We explore CHMM in this work for modeling multi-user sensor sequences and capturing inter-user influences across time. The advantage of using CHMM is that it can recognize activities of both single and multiple users in a unified framework. To the best of our knowledge, this work is among the first using CHMM to recognize multi-user activities from sensor readings.

To illustrate, as shown in Fig. 2, there are two sequences of states A and B with observations O^a and O^b, respectively, at each time slice t. A two-chain CHMM can be constructed by bridging hidden states of its two component HMMs at each time slice with the crosswork of conditional probabilities $P_{a_t|b_{t-1}}$ and $P_{b_t|a_{t-1}}$.

The posterior of a state sequence through fully coupled two-chain CHMM is defined as follows:

$$P(S|O) = \frac{\pi_{a_1} P(o_1^a|a_1)\pi_{b_1} P(o_1^b|b_1)}{P(O)} \prod_{t=2}^{T} [P_{a_t|a_{t-1}} P_{b_t|b_{t-1}} P_{a_t|b_{t-1}} P_{b_t|a_{t-1}} P(o_t^a|a_t) P(o_t^b|b_t)] \tag{1}$$

where π_{a_1} and π_{b_1} are the initial probabilities of states, $P_{a_t|a_{t-1}}$ and $P_{b_t|b_{t-1}}$ are the inner-chain state transition probabilities, $P_{a_t|b_{t-1}}$ and $P_{a_t|b_{t-1}}$ are the inter-chain state transition probabilities modeling the interactions, $P(o_t^a|a_t)$ and $P(o_t^b|b_t)$ are the output probabilities of the states, we employ the Gaussian distribution in this case.

The CHMM inference problem is formulated as follows. Given an observation sequence O, we need to find a state sequence S which maximizes $P(S|O)$. The inference algorithm – Viterbi – for HMM could be applied to CHMM as well with some modifications. The key point is, for each step, we need to compute both the inner-chain and inter-chain state transition probabilities, i.e., $P_{a_t|a_{t-1}} P_{b_t|b_{t-1}}$ and $P_{a_t|b_{t-1}} P_{b_t|a_{t-1}}$. The algorithm outputs the best state sequence S which involves two state sequences S_a and S_b corresponding to the recognized activity sequences for the two users.

There are many existing algorithms for training HMM such as Baum–Welch. Since a two-chain CHMM, C, can be constructed by joining two component HMMs, A and B, and taking the Cartesian product of their states, we define our training method as follows. We first train A and B following the maximum likelihood method, and then, we couple A and B with inter-chain transition probabilities which can be learnt from training datasets. This method is efficient since we do not need to re-train the CHMM.

4.4. Factorial conditional random field

We also study another temporal probabilistic model – FCRF – which was first introduced in [32]. Unlike generative models such as HMM, CRF is an undirected, discriminative model that relaxes the independence assumption of observations and avoids enumerating all possible observation sequences. FCRF factorizes the basic linear-chained CRF by introducing co-temporal connections and, hence, it can be used to model interactions among multi-user activities. Fig. 3 shows an example of FCRFs where there are two CRF chains, and the inter-state connections are conditionally trained. States at both time slices are joined at the same time step. Given the input instance O at a time slice t, we can unroll the CRF chains, A and B, to get a full undirected model, similar to a DBN. It is reported that discriminative models often outperform generative models in classification tasks [33]. We explore the use of FCRF for multi-user activity recognition and compare the performance of FCRF and CHMM.

We define FCRF to model multi-user activities as follows. Let $\mathbf{s} = \{\mathbf{s}_1 \ldots \mathbf{s}_T\}$ be a sequence of random vectors $\mathbf{s}_i = (s_{i1} \ldots s_{im})$, where $\mathbf{s}_i$ is the state vector at time i, and s_{ij} is the value of variable j at time i. Let C be a set of clique indices, $F = \{f_k(\mathbf{s}_{t,c}, \mathbf{o}, t)\}$ be a set of feature functions and $\Lambda = \{\lambda_k\}$ be a set of real valued weights. FCRF (C, F, Λ) is then determined as follows:

$$p(\mathbf{s}|\mathbf{o}) = \frac{1}{Z(\mathbf{o})} \prod_t \prod_{c \in C} \exp\left(\sum_k \lambda_k f_k(\mathbf{s}_{t,c}, \mathbf{o}, t)\right) \tag{2}$$

where $Z(\mathbf{o}) = \sum_s \prod_s \prod_{c \in C} \exp(\sum_k \lambda_k f_k(\mathbf{s}_{t,c}, \mathbf{o}, t))$ is normalization constant, which ensures the final result is a probability.

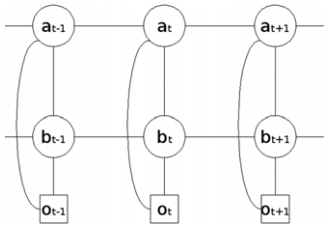

Fig. 3. Structure of FCRF.

Given an observation sequence O, we wish to solve two inference problems: (a) computing the marginals $p(\mathbf{s}_{t,c}|\mathbf{o})$ over all cliques $\mathbf{s}_{t,c}$, and (b) computing the Viterbi decoding $\mathbf{s}^* = \arg\max_s p(\mathbf{s}|\mathbf{o})$. The Viterbi decoding can be used to label a new sequence, and marginal computation is used for parameter estimation.

There are many inference algorithms for CRF. Exact inference can be very computationally intensive. Loopy belief propagation is one of the most popular algorithms for inference in CRF [32]. In general, belief propagation involves iteratively updating a vector $\mathbf{m} = (m_u(o_v))$ of messages between pairs of vertices o_u and o_v. The update from o_u to o_v is given by:

$$m_u(o_v) \leftarrow \sum_{o_u} \Phi(o_u, o_v) \prod_{o_t \neq o_v} m_t(o_u) \tag{3}$$

where $\Phi(o_u, o_v)$ is the potential on the edge (o_u, o_v). Performing this update for an edge (o_u, o_v) in one direction is also called *Sending a message* from o_u to o_v. Given a message vector $\mathbf{o}$, approximate marginals are computed as follows:

$$p(o_u, o_v) \leftarrow \Phi\Phi(o_u, o_v) \prod_{o_t \neq o_v} m_t(o_u) \prod_{o_w \neq o_u} m_w(o_v) \tag{4}$$

where Φ is a normalization constant.

Parameter estimation is an approximate method used for training of FCRF. Given training data O, we aim to find a set of parameters Λ by optimizing the conditional log-likelihood $\mathcal{L}(\Lambda)$ defined as follows:

$$\mathcal{L}(\Lambda) = \sum_i \log p_\Lambda(\mathbf{s}^{(i)}|\mathbf{o}^{(i)}) \tag{5}$$

where $O = \{\mathbf{o}^{(i)}, \mathbf{s}^{(i)}\}_{i=1}^N$ and $\Lambda = \{\lambda_k\}$.

To reduce overfitting, we define a prior $p(\Lambda)$ over parameters, and optimize $\log p(\Lambda|O) = \mathcal{L}(\Lambda) + \log p(\Lambda)$. By using a spherical Gaussian prior with mean $\mu = 0$ and covariance matrix $\Sigma = \sigma^2 I$, the gradient becomes

$$\frac{\partial p(\Lambda|O)}{\partial \lambda_k} = \frac{\partial \mathcal{L}}{\partial \lambda_k} - \frac{\lambda_k}{\sigma^2}. \tag{6}$$

The function $p(\Lambda|O)$ is convex and can be optimized using the L-BFGS [34] technique.

4.5. Activity models in CHMM and FCRF

The dataset we collected consists of a sequence of sensor observations for each user. We first preprocess observation sequences to feature vectors as we described in Section 4.2. Each sequence of feature vectors will be divided into a training sequence and a testing sequence, and input to a CHMM (or FCRF) model. The activity model will be first trained from multiple training sequences corresponding to multiple users, and the trained model is then used to infer activities given multiple testing sequences.

In the training process, we build a CHMM (or FCRF) model for multiple users where each single-chain HMM (or CRF) is used for each user and each hidden state in the HMM (or CRF) represents one activity for the user. We train the CHMM model with multiple training sequences using the parameter estimation method described in Section 4.3. When testing, the testing sequences are fed into the trained CHMM (or FCRF) model, the inference algorithm then outputs multiple state sequences for each single-chain HMM (or CRF) as the labeled sequences.

5. Experimental studies

We now move to evaluate these two models. Our first task is to select an appropriate sensor dataset with proper annotations, consisting of a reasonable number of multi-user activity instances. In this section, we first describe our data collection methodology and experimental setup, then present and discuss the evaluation results obtained from a series of experiments.

Fig. 4. Snapshots showing our smart home, tagged objects, and various activities being performed in our sensor data collection. (a) Our smart home, (b) brushing teeth, (c) making tea, (d) tagged objects, (e) watching TV, (f) making coffee, (g) having a meal.

5.1. Trace collection

Collecting sensor data for multiple users in a real home is a difficult and time-consuming task. Aiming for a realistic data collection, we first conduct a survey among 30 university students. In this survey, each participant was asked to report on what daily activities (both single- and multi-user activities) she/he performed at her/his home, and how each activity is performed (i.e., where each activity is performed, number of persons involved, number of times each activity is performed each day, duration of each activity, major steps and objects used in each activity, etc.). They were asked to report not only their own experiences, but also the experiences from their family members (e.g., parents, siblings, etc.). In return, each participant was awarded with a small amount of remuneration. From the survey reports, we have a number of findings as follows. In a home environment, although many daily activities can be performed by multiple users, single-user activity is still the majority. Second, there are typically fewer than four people involved in a multi-user activity. Third, most of the multi-user activities are performed in the same room. Fourth, interactions among users occur in a number of ways including voice conversation, objects passing, etc.

To have a reasonable and realistic data collection, we randomly select 21 activities (shown in Table 1) from the list in our survey. The ratio of number of single-user activities to number of multi-user activities is close to the result in our survey. We limit the number of users to two to reduce annotation efforts. Our data collection was done in a smart home (i.e., a living lab environment), as shown in Fig. 4 (top left). The smart home consists of a living room, a kitchen, two bedrooms, a study room, a bathroom and a store room. Each room, except the bathroom, is equipped with a video camera for recording the ground truth. We tagged over 100 day-to-day objects such as tablespoons, cups and computer mouse using HF RFID tags in three different sizes—coin, clip and card. Fig. 4 (top right) shows some tagged objects in the kitchen. We have two male subjects, and both are student volunteers from a local university. During data collection, each subject wore a set of wearable sensors we developed and performed these activities following the typical steps for each activity. These activities are performed in an order which is close to daily practice. Fig. 4 shows some snapshots of various activities being performed in the bathroom, the living room, the kitchen and the study room during our data collection. A set of servers was set up in the living room to log the trace. All the servers and sensors were synchronized before data collection. Before the collection, one server broadcasts a message containing a timestamp of the current system and the other servers and sensors synchronize their local clocks with the timestamp on receiving this message. For each user, a trace was logged and annotated by an annotator who is also a student volunteer from a local university. The ground truth was also recorded by video cameras. Data collection was done over a period of ten days across two weeks, and we collected a total number of 420 annotated instances for both subjects. For each user, there are 150 single-user activity instances and 60 multi-user activity instances, and the ratio of number of multi-user instances to number of single-user instances is higher than the result in our survey in order to have more multi-user activity instances.

Compared to the dataset collected in [28], our dataset involves more activities (21 versus 15) and a much larger amount of sensor data (wearable sensors with high sampling rate versus infrastructure sensors).

5.2. Evaluation methodology

We use ten-fold cross-validation to generate our training and test datasets. We evaluate the performance using the time-slice accuracy which is a typical technique in time-series analysis. The time-slice accuracy represents the percentage of correctly labeled time slices. The length of time slice Δt is set to 1 s. This time-slice duration is short enough to provide highly accurate labeling of activities as well as precise measurements for most of the activity recognition applications and

Table 1
Activities performed in our trace collection.

Single-user activities				Multi-user activities	
0	Brushing teeth	8	Vacuuming	15	Making pasta
1	Washing face	9	Using phone	16	Cleaning a dining table
2	Brushing hair	10	Using computer	17	Making coffee
3	Making pasta	11	Reading book/magazine	18	Toileting (with conflict)[a]
4	Making coffee	12	Watching TV	19	Watching TV
5	Toileting	13	Having a meal	20	Using computer
6	Ironing	14	Drinking		
7	Making tea				

[a] This is a case when one person tried to use the toilet which was occupied.

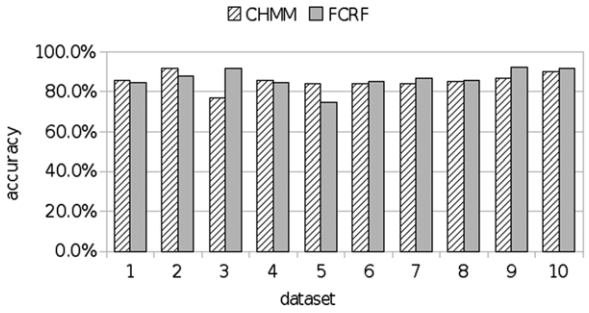

Fig. 5. Accuracy results breakdown in datasets.

Table 2
Accuracy results breakdown in type of activities and users.

Users	CHMM accuracies			FCRF accuracies		
	Single-user activity (%)	Multi-user activity (%)	Overall (%)	Single-user activity (%)	Multi-user activity (%)	Overall (%)
User 1	74.79	96.91	82.22	85.75	87.02	86.70
User 2	85.11	95.91	88.71	82.56	88.84	86.37
Overall	79.95	96.41	85.46	84.16	87.93	86.54

is commonly used in previous work [9,15]. The metric of the time-slice accuracy is defined as follows:

$$Accuracy = \frac{\sum_{n=1}^{N} [predicted(n) = ground\_truth(n)]}{N} \tag{7}$$

where $N = \frac{T}{\Delta t}$.

5.3. Accuracy performance

In the first experiment, we evaluate and compare the accuracies of the two multi-chained probabilistic models. Fig. 5 shows the accuracies of both CHMM and FCRF in all the ten datasets. Table 2 shows the result breakdowns in types of activity and users, respectively. Both models achieve an acceptable performance with similar accuracies—85.46% for CHMM and 86.54% for FCRF. One observation is that CHMM outperforms FCRF in the case of multi-user activity for both users. To analyze, CHMM couples HMMs with temporal, asymmetric influences while FCRF couples CRFs with probabilistic dependencies between co-temporal links. When modeling user interactions, the coupling method which bridges time slices seems to offer a better model of inter-process influences. A similar observation can be found in [31]. Another observation from Table 2 is that the recognition accuracy for multi-user activities is higher than single-user activities for both models. The results show that both models tend to recognize an activity as performed by multiple users. For example, over half of the instances of *watching TV* (single-user activity) were recognized as *watching TV* (multi-user activity) by CHMM and one-third of the instances of *toileting* (single-user activity) were recognized as *toileting (with conflict)* by FCRF. Currently this phenomenon is not properly explained. To have a clear understanding of this phenomenon, extensive experiments with more activities and more subjects involved are necessary, and hence we leave this for our future work.

We also present the confusion matrix for each model as shown in Tables 3 and 4, respectively. The columns in each table show the ground-truth labels and the rows show the predicted labels. The activity serial numbers in both tables are identical to the numbers in Table 1. The values in each table are percentages, indicating that the percentage of the entire observation

Table 3
Confusion matrix of CHMM (percentages).

		Ground truth activities[a]																			
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Predicted activities	0	27.8	2.4												0.1							
	1	72.2	97.6												5.0							
	2			100																		
	3				92.9	25.3																
	4				7.1	74.7																
	5						100															
	6							93.7	1.1													
	7							1.7	93.9	3.5												
	8									95.6												
	9										100											
	10											100										
	11												68.0	1.7								
	12							0.4					4.8	43.3								
	13							0.4					0.4		84.9							
	14							0.3					0.4	0.1		90.2						
	15							2.5					0.6	0.1			90.0	4.9				
	16							0.3					0.4		5.5		10.0	95.1	8.8			
	17							0.3	5.0				0.6	0.1	0.1				91.2			
	18							0.2					0.4	0.1	4.4					100		
	19									0.1			24.4	54.4		0.2					100	
	20									0.8				0.1		9.7						100

[a] Numbers in the first row and the first column are keys to the activities shown in Table 1.

Table 4
Confusion matrix of FCRF (percentages).

		Ground truth activities[a]																			
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Predicted activities	0	95.6	26.3	12.4									0.2							4.2		
	1	1.4	72.4	9.4		0.2	1.9															
	2		1.3	74.2	0.1																	
	3			3.0	78.1	3.3			1.8							0.2	10.4		1.3			
	4				1.9	73.7			2.3									0.2	9.6			
	5					0.1	63.7	0.2														
	6						0.6	82.2						1.9								
	7					6.2		0.2	93.9	0.2								0.1	5.2			
	8								0.3	99.6	0.4											
	9									0.2	94.7											
	10										0.2	80.9	0.6									7.3
	11	0.1									1.5	0.7	93.6	0.9							29.5	
	12							7.7			2.9		0.3	68.4						0.1		
	13													0.3	94.5	0.9		1.3				
	14														0.1	96.7						0.6
	15			0.4	9.9	0.6										0.4	89.3	0.2	1.2			
	16														5.3		0.3	98.0	0.1			
	17			0.4	10.0	15.6			1.6									0.3	82.2	0.1		
	18	2.9				0.3	33.4												0.3	95.5		
	19						0.3	9.7					5.2	28.4						0.2	70.5	
	20										0.2	18.4				1.8						92.1

[a] Numbers in the first row and the first column are keys to the activities shown in Table 1.

sequence for each activity is predicted correctly, and the percentages are predicted as other labels. For CHMM, three single-user activities (*brushing hair*, *toileting* and *using computer*) and two multi-user activities (*watching TV* and *toileting (with conflict)*) give the highest accuracies, and three single-user activities (*brushing teeth*, *reading book/magazine*, and *watching TV*) perform the worst. For FCRF, two single-user activities (*vacuuming* and *drinking*), and one multi-user activity (*cleaning a dining table*) perform the best, and two single-user activities (*toileting* and *watching TV*), and one multi-user activity (*watching TV*) perform the worst. Most confusion takes place in the following four cases:

Case 1: A single-user activity is predicted as another single-user activity. For example, the result of CHMM shows that, for the *making coffee* activity, while 74.7% of its entire observation sequence is predicted correctly, 25.3% of it is predicted as another activity *making pasta*. The result of FCRF shows that, for the *washing face* activity, while 72.4% of its entire observation sequence is predicted correctly, 26.3% of it is predicted as another activity *brushing teeth*. Most of these recognition errors occur between activities within the same place and involve similar user movements and object usage. *Making coffee* and *making pasta* both take place in the kitchen while *washing face* and *brushing teeth* both take place in the bathroom. It is also possible that the RFID reader worn on the users' hands may sense objects that are not supposed to be used. For example the

RFID reader may read the tag attached on the coffee pot or the toothbrush while the user is making pasta or washing face, respectively.

Case 2: A single-user activity is predicted as a multi-user activity. For example, the result of CHMM shows that 24.4% of the observation sequence of *reading book/magazine* is predicted as *watching TV* (multi-user), 54.4% of the observation sequence of *watching TV* (single user) is predicted as *watching TV* (multi-user), and the result of FCRF shows that 33.4% of the observation sequence of *toileting* is predicted as *toileting (with conflict)*.

Case 3: A multi-user activity is predicted as another multi-user activity. For example, for the *making pasta* activity, the result of CHMM shows that 10.0% of its observation sequence is predicted as another activity *cleaning a dining table*. These errors can occur when activities share some objects. Spoons and plates are touched when the user is *making pasta* or *cleaning a dining table*.

Case 4: A multi-user activity is predicted as a single-user activity. For example, the result of FCRF shows that 29.5% of the observation sequence of *watching TV* (multi-user) is predicted as *reading book/magazine*. Both activities involve sitting and both take place in the living room.

Possible solutions for improving multi-user activity recognition include deploying more sensors which are potentially useful to capture user interactions. For example, sensors such as gyro, 3D compass or tilt switch sensors can be potentially used to keep track of a user's head movement. We observe that two users usually face each other when they are talking. Detecting that they face each other can be used as an additional feature to capture their interactions other than voice conversation, which we leave for our future work.

5.4. Feature selection

In this experiment, we study the role of feature selection in the task of multi-user activity recognition. Feature selection has shown to be very useful in improving the performance of an activity recognition model [35]. Different from the existing work, we use a simple filter algorithm to rank sensor feature subsets and demonstrate its effectiveness.

A good sensor feature set is defined as one that contains sensor features highly correlated with the activity class, yet uncorrelated with each other. In other words, a sensor feature is useful if it is correlated with or predictive of the activity class; otherwise it is irrelevant. A feature is said to be redundant if one or more features are highly correlated with it. The goal of feature selection is to eliminate both irrelevant features and redundant features. We perform feature selection using the Correlation-based Feature Selection (CFS) algorithm [36] which is a simple filter algorithm that ranks sensor feature subsets according to a correlation-based heuristic function. The bias of this algorithm is toward subsets that contain sensor features that are highly correlated with the class and uncorrelated with each other. The evaluation of the CFS algorithm is defined as follows:

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{8}$$

where M_s is the heuristic merit of a sensor feature subset S containing k features, $\overline{r_{cf}}$ is the mean feature-class correlation ($f \in S$), and $\overline{r_{ff}}$ is the average feature–feature inter-correlation. The numerator of Eq. (8) can be thought of as providing an indication of how predictive of the activity class a set of sensor features is; the denominator how much redundancy there is among the sensor features.

By applying the CFS algorithm in our datasets, it turns out that there are seven sensor features being selected, listed as follows:

(*ACCEL_LEFT_X_MEAN*, *ACCEL_RIGHT_X_VARIANCE*, *ACCEL_RIGHT_Z_MEAN*,
ACCEL_CORRELATION_7, *AUDIO_ZCR*, *LEFTOBJ*, *LOCATION*).

While most of these sensor features are self-explanatory, feature *ACCEL_CORRELATION_7* represents the correlation between the acceleration data in the *y*-axis for the left hand and the acceleration data in the *x*-axis for the right hand. The result suggests that the four types of sensors such as *x*-axis accelerometer, audio, RFID and location play key roles in a sensor-based, multi-user activity recognition system, while other sensors such as temperature, humidity and light are less important. This result was expected since many daily activities are associated with their unique motion patterns and audio signals. RFID object and location are also very useful when combined with accelerometer and audio sensors. In addition, the interaction between users when performing a multi-user activity can be typically captured by the audio sensor (i.e., voice conversation between two users) and the RFID reader (i.e., passing objects from one user to another).

Based on the above sensor features, we re-train the two models and perform recognition based on the same sets of training and test datasets. The results for both models are given in Table 5 (the results before feature selection are also presented for comparison). We observe 6.23% of improvement for CHMM and 4.98% of improvement for FCRF. Table 5 also suggests that the improvement for single-user activity is more significant than that for multi-user activity, i.e., for single-user activity, we observe 18.64% of improvement using CHMM, and 7.81% of improvement using FCRF.

We end this evaluation with a conclusion that feature selection indeed plays an important role in activity recognition, and we figure out there are seven sensor features which are important for recognizing both single-user and multi-user activities.

Table 5
Effect of feature selection.

Feature selection	CHMM			FCRF		
	Single-user activity (%)	Multi-user activity (%)	Overall (%)	Single-user activity (%)	Multi-user activity (%)	Overall (%)
Without	79.95	96.41	85.46	84.16	87.93	86.54
With	98.59	95.91	91.69	91.97	90.49	91.52

6. Conclusions and future work

In this paper, we study the fundamental problem of recognizing activities of multiple users using wearable sensors in a home setting. We develop our wearable sensor platform and conduct a real-world trace collection in a smart home. We then investigate a challenging problem of how to model and classify multi-user activities. We study two multi-chained temporal probabilistic models—CHMM and FCRF—and our evaluation results demonstrate the effectiveness of both models.

There are a number of limitations in this work. Although the datasets we collected in this paper contain various cases of both single- and multi-user activities, they are still done in a "mock" scenario. A more natural collection should be conducted in a real home and done by real users. We plan to have such data collection in our future work, and evaluate these two multi-chained models further based on a real dataset. We perceive that such a dataset typically contains much background noise, hence it will be challenging to handle noise in these activity models. In addition, although we introduced in Section 3 how we designed our wearable sensor platforms, they are still obtrusive and uncomfortable for people to wear. This issue could be addressed by the future advancement of embedded sensor technology. Moreover, as we discussed in Section 5.3, we will further develop our sensing platform by introducing more sensors such as gyro, 3D compass, etc. These sensors will be very useful in capturing additional sensor features related to user interactions. We plan to deploy our system in a trial to study its performance. The temporal models we use in this paper are general models for modeling interactive signals of different sources. They are also expected to have good performance with these new sensors. We will also investigate better annotation methods that can accurately record the activities performed without affecting the user's behavior.

Another limitation of our work is that, despite the fact that it is effective to recognize multi-user activities using CHMM and FCRF, the scalability of these models has not been investigated. The complexity of inferencing CHMM is $O(TN^{2C})$ for an algorithm which takes the Cartesian product of C chains each with N hidden states observing T data points [31]. Although an approximate algorithm is proposed in [31] which reduces the complexity to $O(T(CN)^2)$, this solution remains questionable in term of scalability. For FCRF, the algorithms used for inferencing from the model are also computationally expensive. One possible way of reducing the complexity is using the location information to reduce the search space of the models by assuming that activities involving multiple users can only be performed when all the users related are at the same location. We plan to investigate the scalability issue of both models in our future work.

We will also investigate a possible future research direction—to recognize activities in a more complex scenario where single- and multi-user activities are mixed with *interleaved* (i.e., switching between the steps of two or more activities) or *concurrent* (i.e., performing two or more activities simultaneously) activities. For example, while two users are preparing a meal, one of the users turns on the TV to watch today's headline news. Similar cases may exist in our daily lives. Recognizing activities in such a complex situation can be very challenging while we consider both single- and multi-user activities at the same time and, hence, an in-depth study is required. Our final goal is to develop an efficient, real-time, sensor-based activity recognition system capable of recognizing various activities for multiple users under different real-life scenarios and deploy the system for real-life trials.

Acknowledgements

This work was supported by the Danish Council for Independent Research/Natural Science under Grant 09-073281, National 973 program of China under Grant 2009CB320702, Natural Science Foundation of China under Grants 60736015, 60721002 and 61073031 and Jiangsu PanDeng Program under Grant BK2008017.

References

[1] Y. Yacoob, M.J. Black, Parameterized modeling and recognition of activities, in: Proc. of Computer Vision and Image Understanding, vol. 73, 1999, pp. 232–247.
[2] D. Moore, I. Essa, M. Hayes, Exploiting human actions and object context for recognition tasks, Proc. of Int'l Conf. on Computer Vision, ICCV'99, vol. 1 1999, pp. 80–86.
[3] S. Katz, A.B. Ford, R.W. Moskowitz, B.A. Jackson, M.W. Jaffe, Studies of illness in the aged, the index of adl: a standardized measure of biological and psychological function, J. Amer. Med. Assoc. (1963) 914–919.
[4] M. Pollack, L. Brown, D. Colbry, C. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, I. Tsamardinos, Autominder: an intelligent cognitive orthotic system for people with memory impairment, Robot. Auton. Syst. 44 (2003) 273–282.
[5] T. Gu, L. Wang, Z. Wu, X. Tao, J. Lu, A pattern mining approach to sensor-based human activity recognition, in: IEEE Transactions on Knowledge and Data Engineering, TKDE, 23 Sept, 2010.
[6] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: Proc. of PERVASIVE 2004, Vienna, Austria, vol. 3001, 2004, pp. 1–17.

[7] B. Logan, J. Healey, M. Philipose, E. Munguia-Tapia, S. Intille, A long-term evaluation of sensing modalities for activity recognition, in: Proc. of Int'l Conf. on Ubiquitous Computing, Ubicomp'07, Austria, vol. 4717, 2007, pp. 483–500.
[8] D.J. Patterson, D. Fox, H. Kautz, M. Philipose, Fine-grained activity recognition by aggregating abstract object usage, in: Proc. IEEE Int'l Symp. Wearable Computers, Osaka, 2005, pp. 44–51.
[9] J.A. Ward, P. Lukowicz, G. Troester, T. Starner, Activity recognition of assembly tasks using body-worn microphones and accelerometers, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1553–1567.
[10] M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, D. Hähnel, Inferring activities from interactions with objects, IEEE Pervasive Comput. 3 (4) (2004) 50–57.
[11] R. Hamid, S. Maddi, A. Johnson, A. Bobick, I. Essa, C. Isbell, A novel sequence representation for unsupervised analysis of human activities, J. Artif. Intell. 173 (14) (2008) 1221–1244.
[12] T.L.M. van Kasteren, A.K. Noulas, G. Englebienne, B. Kröse, Accurate activity recognition in a home setting, in: Proc. of Int'l Conf. on Ubiquitous Computing, Ubicomp'08, Seoul, Korea, 2008, pp. 1–9.
[13] E.R. Smith, D.M. Mackie, Social Psychology, Routledge, London, 1999.
[14] D.L. Vail, M.M. Veloso, J.D. Lafferty, Conditional random fields for activity recognition, in: Proc. of Int'l Conf. on Autonomous Agents and Multi-agent Systems, AAMAS, 2007, pp. 1–8.
[15] T.Y. Wu, C.C. Lian, J.Y. Hsu, Joint recognition of multiple concurrent activities using factorial conditional random fields, in: Proc. AAAI Workshop Plan, Activity, and Intent Recognition, California, 2007.
[16] S. Gong, T. Xiang, Recognition of group activities using dynamic probabilistic networks, in: Proc. of the 9th Int'l Conf. on Computer Vision, ICCV'03, Nice, France, 2003, pp. 742–749.
[17] N. Nguyen, H. Bui, S. Venkatesh, Recognising behaviour of multpile people with hierarchical probabilistic and statistical data association, in: Proc of the 17th British Machine Vision Conference, BMVC'06, Edinburgh, Scotland, 2006, pp. 1239–1248.
[18] S. Park, M.M. Trivedi, Multi-person interaction and activity analysis: a synergistic track- and body- level analysis framework, in: Machine Vision and Applications: Special Issue on Novel Concepts and Challenges for the Generation of Video Surveillance Systems, vol. 18, Springer Verlag Inc., 2007, pp. 151–166.
[19] Y. Du, F. Chen, W. Xu, Y. Li, Recognizing interaction activities using dynamic bayesian network, in: Proc. of Int'l Conf. on Pattern Recognition, ICPR'06, Hong Kong, China, vol. 1, 2006, pp. 618–621.
[20] D. Wyatt, T. Choudhury, J. Bilmes, H. Kautz, A privacy sensitive approach to modeling multi-person conversations, in: Proc. of Int'l Joint Conf. On Artificial Intelligence, IJCAI'07, India, 2007, pp. 1769–1775.
[21] T. Choudhury, S. Basu, Modeling conversational dynamics as a mixed-memory markov process, in: Advances in Neural Information Processing Systems, NIPS'04, MIT Press, Cambridge, MA, 2005, pp. 281–288.
[22] C. chun Lian, J.Y. jen Hsu, Chatting activity recognition in social occasions using factorial conditional random fields with iterative classification, in: Proceedings of the Twenty-Third AAAI Conf. on Artificial Intelligence, 2008, Chicago, Illinois, 2008, pp. 1814–1815.
[23] N. Oliver, B. Rosario, A. Pentland, A bayesian computer vision system for modeling human interactions, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 831–843.
[24] N. Oliver, A. Gargb, E. Horvitz, Layered representations for learning and inferring office activity from multiple sensory channels, Comput. Vis. Image Understanding 96 (November) (2004) 163–180.
[25] D. Zhang, D. Gatica-Perez, S. Bengio, I. McCowan, G. Lathoud, Modeling individual and group actions in meetings: a two-layer hmm framework, in: Proc. of Computer Vision and Pattern Recognition Workshop, CVPRW'04, Washington, DC, USA, vol. 7, 2004, p. 117.
[26] Z. Lin, L. Fu, Multi-user preference model and service provision in a smart home environment, in: Proc. of IEEE Int'l Conf. on Automation Science and Engineering, CASE'07, 2007, pp. 759–764.
[27] P. Rashidi, G. Youngblood, D. Cook, S. Das, Inhabitant guidance of smart environments, in: Proc. of the Int'l Conf. on Human–Computer Interaction, 2007, pp. 910–919.
[28] G. Singla, D. Cook, M. Schmitter-Edgecombe, Recognizing independent and joint activities among multiple residents in smart environments, J. Ambient Intell. Human. Comput. 1 (2009) 57–63.
[29] T. Gu, L. Wang, Z. Wu, X. Tao, J. Lu, Mining emerging patterns for recognizing activities of multiple users in pervasive computing, in: Proc. of the 6th Int'l Conf. on Mobile and Ubiquitous Systems: Computing, Networking and Services, MobiQuitous '09, Toronto, Canada, 2009, pp. 1–10.
[30] U. Fayyad, K. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, in: Proc. Int'l Joint Conf. on Artificial Intelligence, San Francisco, 1993, pp. 1022–1027.
[31] M. Brand, Coupled hidden markov models for modeling interacting processes, Technical Report, vol. 1721, 1997, pp. 30–42.
[32] M. Brand, Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data, J. Mach. Learn. Res. 8 (2007) 693–723.
[33] A. Ng, M. Jordan, On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes, Adv. Neural Inf. Process. Syst. 10 (1.46) (2002) 6208.
[34] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.
[35] J. Lester, T. Choudhury, N. Kern, G. Borriello, B. Hannaford, A hybrid discriminative/generative approach for modeling human activities, in: Proc. of the Int'l Joint Conf. on Artificial Intelligence, IJCAI'05, 2005, pp. 766–772.
[36] M.A. Hall, Correlation-based feature subset selection for machine learning, Ph.D. Thesis, The University of Waikato, 1999.