Continuous Recognition of Daily Activities from Multiple Heterogeneous Sensors

Tsu-yu Wu and **Jane Yung-jen Hsu** and **Yi-ting Chiang**
Department of Computer Science and Information Engineering
National Taiwan University
yjhsu@csie.ntu.edu.tw

Abstract

Recognition of daily activities is the key to providing context-aware services in an intelligent home. This research explores the problem of activity recognition, given diverse data from multiple heterogeneous sensors and without prior knowledge about the start and end of each activity. This paper presents our approaching to *continuous recognition* of daily activities as a *sequence labeling problem*. To evaluate the capability of activity models in handling heterogeneous sensors, we compare several state-of-the-art sequence labeling algorithms including hidden Markov model (HMM), linear-chain conditional random field (LCRF) and SVM^{hmm}. The experimental results show that the two discriminative models, LCRF and SVM^{hmm}, significantly outperform HMM. In particular, SVM^{hmm} shows robustness in dealing with all sensors we used, and its recognition accuracy can be further improved by incorporating carefully designed *overlapping features*.

Introduction

Collecting information about daily activities plays an important role in health assessment. As has been pointed out by healthcare professionals, one of the best way to detect an emerging medical condition before it becomes critical is to look for changes in the activities of daily living (ADLs) (Katz *et al.* 1963), instrumental ADLs (IADLs) (Lawton & Brody 1969), and enhanced ADLs (EADLs) (Rogers *et al.* 1998). According to the definitions from MedicineNet.com, ADLs are "the things we normally do in daily living including any daily activity we perform for self-care such as feeding ourselves, bathing, dressing, grooming, work, homemaking, and leisure". The ability or inability to perform ADLs can be used as a very practical measure of ability and disability in many disorders.

However, self-reporting of ADLs may be unreliable due to dishonesty or inability of the subjects. With a computational system keeping track of daily activities, it becomes easier for family members or care-givers to identify potential health problem at home. Besides, recognizing daily activities enables the computer to be an active service provider. For example, an intelligent housekeeper can control the lighting, heating, and air conditioning level, or provide prompt help to family members according to their current activities.

The goal of our research is to recognize when and which activities take place given a sensor trace. Daily activities are usually performed without explicit cues available for us to know the boundary between activities. The number of activities in any data trace is unknown. Moreover, people have different patterns of activity execution. For example, a person may take a bath for one hour while another completes it in three minutes. A recognition system is practical only when it is able to handle such individual differences. We want to build a robust model that is able to properly deal with different types of sensor readings as well as ways to fuse them. It should also be robust to learn the variation of different individuals.

In this paper, we compare state-of-the-art sequence models for recognition of daily activities from multiple heterogeneous sensors. Our experiments include three types of senosrs: microphones, a wireless location tracking system, and an RFID-based object tracking system. In addition, issues on overlapping features, on-line recognition, and segment errors are discussed.

Related Work

The problem of sequence labeling seeks a mapping from the input to the output sequences. However, the input space and output space are exponentially large. Here we introduce some of the state-of-the-art sequence labeling models.

HMM

Hidden Markov Model (HMM) has been widely used in the area of speech recognition and natural language processing. It can be used to learn sequential characteristics in time series data. Although HMM is a generative model, it is widely used in discrimination problems. By defining the output sequence as the hidden random variable sequence, the states of random variables are corresponding to labels of output sequence. The state sequence that has highest joint probability with the observation is determined as the output sequence. While being an efficient model, HMM has some problems like the difficulty in dealing with complex temporal dependency, the limitation in the number of hidden states, and the label bias problem. Some extensions of HMM have been proposed. Ghahramani et al. (Ghahramani & Jordan

1997) proposed the factorial HMM which allows large cardinality of hidden state in HMMs. A HMM that can deal with countably infinite number of hidden states can be found in (Beal, Ghahramani, & Rasmussen 2003). Oliver et al. (Oliver, Garg, & Horvitz 2004) developed a HMM with layered structure (Layered HMM) that enables the HMM to learn in different temporal granularities. The LHMM can be used to decompose heterogeneous signals and lower the retraining cost when the environment changes.

DBN

Dynamic Bayesian network (DBN) is a more general model that extends BN to represent the temporal relationship between time slices. Rather than a single latent random variable every frame in HMM, DBN allows complex dependency structure of different random variables in and between time frames. Patterson et al. (Patterson *et al.* 2003) and Liao et al. (Liao, Fox, & Kautz 2005) propose different DBNs that model the complex dependency of the transportation mode, speed, location and GPS readings. Some filtering algorithm such as Rao-Blackwellized particle filter (RBPF) favors this kind of factored structured since it allows different filtering strategies for different parts. Wilson (Wilson & Atkeson 2005) propose an RBPF that estimates the association using the particle filter but estimates the distribution of the location and activity using Bayes filter. Although DBN is powerful for modeling complex relationship of multiple random variables, it usually pay for more computational effort. Exact inference algorithm such as junction tree algorithm (Smyth, Heckerman, & Jordan 1997) needs exponential inference time if the structure is too complex in DBN. Approximation techniques such as Gibbs sampling and loopy belief propagation try to solve this problem.

MEMM and CRF

Generative models select the sequence with maximum likelihood of observation. However, sequence labeling is a discrimination problem that predicts the sequence given the observation. Directly modeling the conditional probabilities of the label sequence given the observation seems more natural for this problem. Maximum entropy Markov model (MEMM) (Mccallum, Freitag, & Pereira 2000) is a conditional model that uses similar structure of HMM except the relationship of the observation is inverted. In this way, we can model dependent observation such as overlapping features and long-term observation without making improper independence assumption. However, like HMM, MEMM suffers from the label bias problem due to the per-state normalization.

Conditional random field (Lafferty, McCallum, & Pereira 2001) (CRF) achieves great success in this field by solving the label bias problem in MEMM using the global normalization. CRF has also been applied in activity recognition. Chieu et al. (Chieu, Lee, & Kaelbling 2006) and Sminchisescu et al. (Sminchisescu, Kanaujia, & Metaxas 2006) utilize a linear-chain CRF (LCRF) for the activity recognition problem and show the superiority over HMM. Liao et al. propose using a hierarchical CRF and iteratively inferring activities and important locations simultaneously (Liao, Fox, & Kautz 2007). Shimosaka et al. (Shimosaka, Mori, & Sato 2007) and we (Wu, Lian, & Hsu 2007) propose using a factored structure of CRF for solving the multi-tasking activity recognition. Benson et al. (Limketkai, Liao, & Fox 2007) propose a CRF-Filter that is adapted from the particle filter to solve the on-line recognition problem in localization.

Structural SVM

SVM is an effective approach for the classification problem. Altun et al. (Altun, Tsochantaridis, & Hofmann 2003) propose an extension of SVM, hidden markov support vector machine, that handles the sequence labeling problem. A general model for arbitrary output space, structural SVM, is proposed by Tsochantaridis et al. (Tsochantaridis *et al.* 2005). In an experiment conducted by Nguyen and Guo (Nguyen & Guo 2007), structural SVM outperforms 5 other state-of-the-art models including HMM, CRF, Averaged perceptron (AP), Maximum margin Markov networks(M^3N), and an integration of search and learning algorithm (SEARN) in two sequence labeling tasks, part-of-speech (POS) tagging and optical character recognition (OCR). The experiments show the superiority of structural SVM in two problems. But in a later technical report (Keerthi & Sundararajan 2007), CRF is shown to be comparable with structural SVM when appropriate features are used. In this paper, we also address this issue with the activity recognition problem.

Activity Modeling

We formulate the problem of recognizing daily activities as a sequence labeling problem, which is to assign a single label to each element in an observation sequence. In our problem, the activity is labeled every a fixed interval given the sensor readings. We compare state-of-the-art models including HMM, linear chain CRF and SVM^{hmm} in this problem. In this section, we describe briefly how these models are used in our work. Experimental details will be described in the next section.

We define the problem as following: Given an observation sequence $O = (O_1, O_2, ..., O_T)$ where O_t is the feature vector combining data of multiple heterogeneous sensors at time t, the goal is to generate a sequence of predicted activities $A = (A_1, A_2, ..., A_T)$ where A_t occurs at time t.

HMM

In our work, each activity is modeled as one state of the hidden node in HMM. Each dimension of the observation is assumed conditionally independent of each other given the activity. The conditional probability of each discrete feature given the activity is modeled as a multinomial distribution; the conditional probability of each real value feature given the activity is modeled as a mixture of 5 Gaussian distributions. We use the Graphical Modeling Toolkit (GMTK) (Bilmes 2007) for the implementation of HMM.

Linear Chain CRF

We use LCRF (Lafferty, McCallum, & Pereira 2001) for the recognition problem. LCRF is an instantiation of CRF that uses similar graphical structure with HMM. CRF models the conditional probability $P(A|O)$ by a set of feature functions $\{f_1(A,O,t), f_2(A,O,t), ..., f_J(A,O,t))\}$ and a weight vector $\{w_1, w_2, ..., w_J\}$. The conditional probability $P(A|O)$ is defined as

$$P(A|O) = \frac{exp(\sum_j \sum_t w_j f_j(A,O,t))}{Z(O)}$$

where $Z(O)$ is the normalization constant.

Here we use unigram feature functions between the activity label and a feature value (observation) in the same frame. For the discrete feature, a set of binary feature functions is defined for every combination of the activity and the feature value. For example, we define a binary feature function for one activity $a \in A$ and one feature value $o \in O$ as

$$f_{a,o}(A,O,t) = \begin{cases} 1, & \text{if } A_t = a \text{ and } O_t = o. \\ 0, & \text{otherwise.} \end{cases}$$

For the real value feature, a set of real value feature functions is defined for every activity. The value of the feature functions is defined as the feature value. For example, we define a feature function for the activity a and the mean value of volume (mv) as

$$f_{a,mv}(A,O,t) = \begin{cases} x, & \text{if } A_t = a \text{ and } O_t^{mv} = x. \\ 0, & \text{otherwise.} \end{cases}$$

where O_t^{mv} is the mean of volume at time t.

For modeling the temporal relationship, we use bigram feature functions between adjacent activities. A set of binary feature functions is defined for every combination of the activities. For example, we define a binary feature function for the consecutive activities of a as

$$f_{a,a}(A,O,t) = \begin{cases} 1, & \text{if } A_{t-1} = a \text{ and } A_t = a. \\ 0, & \text{otherwise.} \end{cases}$$

We choose CRF++ (Kudo 2007) for the implementation of LCRF. CRF++ is originally designed for discrete features. We extend CRF++ to handle the real value features. We predict the activity sequence with the maximum conditional probability given the observation sequence using Viterbi algorithm. Given the training data $D = (D_1, D_2, ..., D_N)$ where $D_i = (A_i, O_i)$, the learning criteria is to find a weight vector W that maximizes the log-likelihood of the training data. A zero mean Gaussian prior is assumed to avoid overfitting. A single variance σ^2 is used to control the degree of penalization for each weight w_i. Higher σ^2 makes the model tend to fit the training data. L-BFGS (Liu & Nocedal 1989) is used to train the model.

SVM^{hmm}

SVM^{hmm} (Joachims 2008) is a sequence labeling instantiation of structural SVM. Similar to LCRF, structural defines a linear discriminant function $D(A,O)$ by a set of feature functions $\{f_1(A,O,t), f_2(A,O,t), ..., f_J(A,O,t))\}$ and a weight vector $\{w_1, w_2, ..., w_J\}$. The linear discriminant function $D(A,O)$ is defined as

$$D(A,O) = \sum_j \sum_t w_j f_j(A,O,t).$$

Here we use the same feature functions in LCRF and SVM^{hmm}.

Unlike the maximum likelihood estimation in CRF and HMM, structral SVM does not model the probabilities but discriminate between different label sequences. The learning criteria is similar to conventional SVM that maximizes the margin. The loss function is the misclassified labels in a sequence. A cost factor c is used to control the trade off between the margin and loss. Higher c makes the model tend to fit the training data.

Other Approaches

To evaluate how we benefit from the modeling of the temporal relationship, we use three classifiers, NBC, maximum entropy classifier (MEC), and SVM for comparison. In this formulation, each time frame is viewed as an instance and the activity is independently classified with the observation in the frame. NBC, MEC and SVM can be viewed as a specialization of HMM, CRF and SVM^{hmm}. Here we use the same implementation in HMM, CRF and SVM^{hmm}.

Experiment

We use E-Home dataset as the evaluation dataset. E-Home dataset is collected in a home-like environment for the research of the activity recognition by Yen (Yen 2007). The dataset involves 13 subjects and each performs 12 activities. The order of activities is random and the parts of reading experimental instructions in the trace is manually eliminated. The dataset consists of totally 27818 seconds and the activity is annotated every second. In the dataset, three primary kinds of sensors including a microphone in the corner of the room, a wearable RFID reader and 40 load sensory blocks on the floor are used. For the audio stream from the microphone, 24 real value features like mean of volume, variance of volume, and mean of zero-crossing rate are extracted from it. A location system estimates the most active block of the 40 load sensory blocks by filtering out the deformation noise. The RFID reader returns null or one of the 24 tagged objects. The features are extracted every second. We do not know how many activities occur as well as the duration of each activity.

More details about this dataset, including RFID deployment and audio features, can be found in (Yen 2007).

Performance Measures

The output $(P_2, P_2, ..., P_T)$ is a string of T predictions of activities. The ground truth $(G_1, G_2, ..., G_T)$ is the annotated activity sequence. According to the E-Home dataset we used, the cardinality of the output set of these P_i's and G_i's are 12, and the time interval of each predicted and annotated activity is 1 second. To evaluate how our recognition algorithms perform, we use leave-one-subject-out cross-validation in our experiment. Since the data set involves 13 subjects, the model is trained using the data sequence containing 12 subjects and tested using the remaining one. The following two accuracy values are used for the comparisons.

Frame Accuracy The frame accuracy (FA) is the rate of matching frames between the prediction and the ground truth.

Average Class Accuracy The frame accuracy is easily affected by the long activity. For example, in E-Home dataset, a system that always predicts the activity as "watching TV" can be as accurate as 17% that is much higher than random guess due to the high coverage of watching TV. The average class accuracy (ACA) is the normalized frame accuracy by each activity.

Parameter Settings A weight w of the uniform distribution is used to control the smoothness of CPTs used by the viterbi algorithm in GMTK, the toolkit we used to run HMM experiments. We smooth each multinomial distribution using a weighted sum of the learned distribution and a uniform distribution. This parameter w is tested from 0.90 to 0.99 and 0.01 in our experiment. The variance σ^2 for LCRF and the cost factor c for SVM^{hmm} are tested from 2^{-3} to 2^4 and multiplied by 2 as a step. The parameters that achieve highest frame accuracy are used for each model.

Raw Features

We first use the raw features for the evaluation. Raw features include a 24-dimensional sequence of real value vector for the audio sensor, a discrete sequence for the RFID system and a discrete sequence for the location system. The features of these three sensors diverge in form. Audio features are real value vectors while RFID features and location features are discrete values. In additions, RFID features differ from location features in sparsity. In E-home dataset, RFID returns null in 90% of time.

Results The results are summarized in Table 1.

FA/ACA(%)	HMM	LCRF	SVM^{hmm}
Audio	23.5/24.8	37.6/27.3	45.0/36.4
RFID	44.9/44.1	51.4/44.4	59.5/50.9
Location	31.5/37.4	43.3/37.1	40.2/37.5
Audio+RFID	31.9/33.5	62.3/54.9	69.7/63.5
Audio+Location	39.5/41.9	56.3/48.6	61.0/56.0
Location+RFID	39.6/45.0	63.8/56.8	65.2/60.3
All	44.3/46.8	68.8/64.6	72.0/67.8

Table 1: Performance Comparison of HMM, LCRF and SVM^{hmm} using Raw Features.

In HMM, the result of fusing all sensors is even worse than using RFID only. The situation informs us that it can be dangerous to fuse sensors in a single HMM. Assuming independence for the 24-dimensional audio features in HMM is dangerous since these features are extracted from the same audio signal. With discriminative models such as LCRF and SVM^{hmm}, fusing more sensors generally performs better. The accuracy of LCRF and SVM^{hmm} using all sensors is much better than HMM. In our experiment, LCRF performs slightly worse than SVM^{hmm}. We can see that SVM^{hmm} show robustness in dealing with varieties of sensors while LCRF is relatively weak in dealing with RFID and audio features. It seems inappropriate to assume simple distribution between the activity and the audio features. Fitting parameters to a wrong distribution can result in severe bias. For RFID, since the event is sparse, there may not be enough counts for CRF to overcome the prior.

To show the usefulness of temporal relationship, we evaluate how sequence models improve over the frame-based classification by considering the temporal relationship. The results of different models are shown in Table 2.

FA/ACA(%)	NBC	MEC	SVM
Classification	35.5/34.2	42.7/39.7	43.7/40.9
FA/ACA(%)	HMM	LCRF	SVM^{hmm}
Sequence Models	40.9/43.1	68.8/64.6	72.0/67.8

Table 2: Performance Comparison of Frame-Based Classification and Sequence Models.

The results show that all sequence models outperform corresponding classifiers in both performance measures. The improvement of the frame accuracy by considering the temporal relationship can be up to 28.3% in SVM^{hmm}. The improvement of the average class accuracy can be up to 27% in LCRF. The significant difference shows us frame-based classification is not adequate in this problem.

Overlapping Features

Observation can be view in multiple ways. For example, in natural language processing, the word "White" can be viewed as the word itself or a capitalized word. In recognizing the name entity, the capitalization feature may be very informative. Discriminative models such as LCRF and SVM^{hmm} are shown to be able to utilize this kind of overlapping features. We describe three different strategies to extract features in our problem.

Generative Audio Probabilities The audio in a single second may not contain sufficient information. In additions, performing activities may generate different sounds in different stages. For example, in preparing meals, the sound of chewing can be very different from using microwave.

We use an HMM to model the relationship between a specific activity and a small segment of audio features.

Given the training data, at each time c, we segment a small length l of audio features $S_c^{LR} = (O_{c-l}^{Audio}, O_{c-l+1}^{Audio}, ..., O_c^{Audio})$ and associate the segment S_c^{LR} with the activity label A_c . For each activity i, we independently train an HMM parameter λ_i^{LR} using the segments with activity label i. By reversing the time index, we segment a set of audio features S_c^{RL} and train an HMM parameter λ_i^{RL} for each activity i. As a result, we have 24 HMM parameters.

For a testing sequence, we use a backward sliding window as well as a forward sliding window to segment two audio sequence of w frames with 50% overlapping. We then use the 24 HMM parameters to estimate the generative probabilities of the two segments. The probability estimations are logged and scaled to values ranging from 0 to 1. These 24

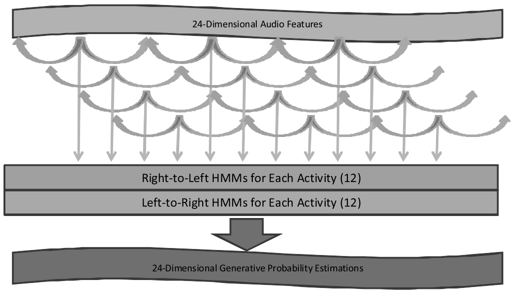

Figure 1: Generative Audio Probabilities.

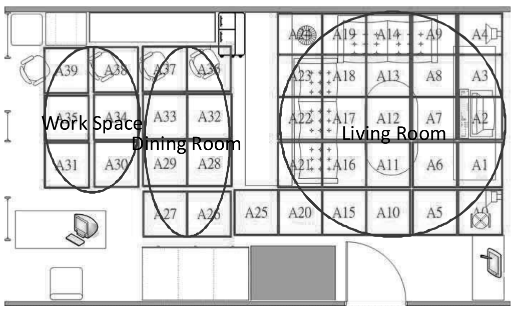

Figure 2: Regions and Sensory Blocks.

probability estimations are used as an additional feature vector. Figure 1 show the process of creating these features.

Region and Region Transitions We group the 40 load sensory blocks into 3 places including the living room, the dining room and the workspace. The location feature is the index of the 40 blocks while the region feature is the corresponding place. In additions, we consider all transitions between these 3 places as additional features. Figure 2 shows the relationship of the regions and location sensory blocks.

NextRFID and LastRFID When a non-null object is read by the RFID reader, it is usually very informative for disambiguating activities. Referring to the recent non-null object reading is helpful. For example, if we hold the TV remote control at last 3 seconds, it is very possible that we are watching TV currently. We define two features, NextRFID and LastRFID for expanding the raw RFID features.

For non-null object i at the time frame c, if the nearest RFID reading of object i is at the time frame c' where c' is larger than c, NextRFID distance is defined as $(c' - c)$. To prevent referring to the object that is irrelevant in time frames of other activities, the maximum distance is limited. In additions, the forward referencing process ended when it encounters the location change or another object. LastRFID distance is defined in the same way of the NextRFID except that the order of time frames is reversed. The resulting distances for objects are scaled to values ranging from 0 to 1. As a result, we have a new 48-dimensional features for RFID. These features are used as a replacement of the raw RFID readings.

Results To show the effect of incorporating these overlapping features, we evaluate the performance of LCRF and SVM^{hmm} with these overlapping features. The results are summarized in Table 3 and 4.

FA/ACA(%)	CRF	SVM^{hmm}
Audio	37.6/27.3	45.0/36.4
RFID	51.4/44.4	59.5/50.9
Location	43.3/37.1	40.2/37.5
All	68.8/64.6	72.0/67.8

Table 3: Performance Comparison of LCRF and SVM^{hmm} Using Raw Features.

FA/ACA(%)	CRF	SVM^{hmm}
Audio	40.1/29.8	45.9/37.4
RFID	63.6/56.8	63.0/55.4
Location	45.9/39.2	44.1/40.3
All	74.8/70.0	73.0/71.1

Table 4: Performance Comparison of LCRF and SVM^{hmm} Using Overlapping Features.

By combining these overlapping features, we improve the accuracy of LCRF and SVM^{hmm} in all sensor settings. Note that the frame accuracy of LCRF is improved from 51.4% to 63.6% with the NextRFID and LastRFID features because the two features solve the sparsity of the raw RFID readings. With these features, the performance of LCRF and SVM^{hmm} is close.

Conclusion and Future Work

This paper presents our formulation of continuous recognition of daily activities as a sequence labeling problem. We compare the capacity of several state-of-the-art models including HMM, LCRF, and SVM^{hmm} in handling heterogeneous sensors. In our experiment, discriminative models such as LCRF and SVM^{hmm} significantly outperform HMM. While LCRF is weak with respect to RFID and audio sensors, SVM^{hmm} is robust in dealing with all types of sensors in the E-home dataset. To better utilize the sensor data, we propose strategies for extracting overlapping features. For example, the NextRFID and LastRFID features significantly improve the accuracy of LCRF by eliminating the sparsity of the RFID sensor.

This research can be extended in several ways. The E-home dataset was collected by subjects following instructions in a lab environment. In reality, multiple activities may be carried out concurrently or with interruption. The highest frame accuracy of 74.8% in activity recognition, may not be good enough for practical applications. We plan to explore additional sensors such as accelerometers and biophysical sensors to improve recognition.

Acknowledgments

This work was partially supported by grants from the National Science Council, Taiwan (NSC 96-2628-E-002-173-MY3), Excellent Research Projects of National Taiwan University (97R0062-06), and Intel Corporation.

References

Altun, Y.; Tsochantaridis, I.; and Hofmann, T. 2003. Hidden markov support vector machines. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Beal, M. J.; Ghahramani, Z.; and Rasmussen, C. E. 2003. The infinite hidden markov model. *Advances in Neural Information Processing Systems 14* 577–584.

Bilmes, J. 2007. *The Graphical Models Toolkit (GMTK)*. `http://ssli.ee.washington.edu/~bilmes/gmtk/`.

Chieu, H. L.; Lee, W. S.; and Kaelbling, L. P. 2006. Activity recognition from physiological data using conditional random fields. Technical report, Singapore-MIT Alliance (SMA) Annual Symposium.

Ghahramani, Z., and Jordan, M. I. 1997. Factorial hidden markov models. *Machine Learning* 29(2-3):245–273.

Joachims, T. 2008. *SVM^{hmm} : Sequence Tagging with Structural Support Vector Machines*. `http://svmlight.joachims.org/svm_struct.html`.

Katz, S.; Ford, A.; Moskowitz, R.; Jackson, B.; and Jaffe, M. 1963. Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 185(12):914–9.

Keerthi, S. S., and Sundararajan, S. 2007. CRF versus SVM-struct for sequence labeling. Technical report, Yahoo Research.

Kudo, T. 2007. *CRF++: Yet Another CRF toolkit*. `http://crfpp.sourceforge.net/`.

Lafferty, J.; McCallum, A.; and Pereira, F. 2001. Conditional random fields: Probabilistic models. for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Lawton, M., and Brody, E. 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 9(3):179–86.

Liao, L.; Fox, D.; and Kautz, H. A. 2005. Location-based activity recognition using relational markov networks. In *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, 773–778. Professional Book Center.

Liao, L.; Fox, D.; and Kautz, H. A. 2007. Extracting places and activities from GPS traces using hierarchical conditional random fields. *International Journal of Robotics Research* 26:119–134.

Limketkai, B.; Liao, L.; and Fox, D. 2007. CRF-Filters: Discriminative particle filters for sequential state estimation. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*.

Liu, D. C., and Nocedal, J. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45:503–528.

Mccallum, A.; Freitag, D.; and Pereira, F. 2000. Maximum entropy markov models for information extraction and segmentation. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Nguyen, N., and Guo, Y. 2007. Comparisons of sequence labeling algorithms and extensions. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Oliver, N.; Garg, A.; and Horvitz, E. 2004. Layered representations for learning and inferring office activity from multiple sensory channels. *Comput. Vis. Image Underst.* 96(2):163–180.

Patterson, D. J.; Liao, L.; Fox, D.; and Kautz, H. A. 2003. Inferring high-level behavior from low-level sensors. In *Proceedings of the International Conference on Ubiquitous Computing (Ubicomp)*, volume 2864 of *Lecture Notes in Computer Science*, 73–89. Springer.

Rogers, W.; Meyer, B.; Walker, N.; and Fisk, A. 1998. Functional limitations to daily living tasks in the aged: a focus group analysis. *Human Factors* 40(1):111–125.

Shimosaka, M.; Mori, T.; and Sato, T. 2007. Robust action recognition and segmentation with multi-task conditional random fields. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*.

Sminchisescu, C.; Kanaujia, A.; and Metaxas, D. 2006. Conditional models for contextual human motion recognition. *Computer Vision and Image Understanding* 104(2):210–220.

Smyth, P.; Heckerman, D.; and Jordan, M. I. 1997. Probabilistic independence networks for hidden markov probability models. *Neural Computation* 9(2):227–269.

Tsochantaridis, I.; Joachims, T.; Hofmann, T.; and Altun, Y. 2005. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research (JMLR)* 6:1453–1484.

Wilson, D. H., and Atkeson, C. G. 2005. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In Gellersen, H.-W.; Want, R.; and Schmidt, A., eds., *Proceedings of the International Conference on Pervasive Computing (Pervasive)*, volume 3468 of *Lecture Notes in Computer Science*, 62–79. Springer.

Wu, T.-y.; Lian, C.-c.; and Hsu, J. Y.-j. 2007. Joint recognition of multiple concurrent activities using factorial conditional random fields. Technical Report WS-07-09, AAAI Workshop on Plan, Activity, Intent Recognition (PAIR).

Yen, C.-n. 2007. Utilizing heterogeneous sensors for everyday activity recognition. Master's thesis, National Taiwan University.