

# 行政院國家科學委員會專題研究計畫 成果報告

iCare: 社群化智慧型居家照護--子計畫三：多代理人式行為追蹤與主動控制系統研究(3/3)  
研究成果報告(完整版)

計畫類別：整合型  
計畫編號：NSC 95-2218-E-002-025-  
執行期間：95年10月01日至96年09月30日  
執行單位：國立臺灣大學資訊工程學系暨研究所

計畫主持人：許永真  
共同主持人：洪一平  
計畫參與人員：碩士班研究生-兼任助理：顏嘉楠、吳祖佑、李世偉、連家峻、陳麗徽、黃啟嘉

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中華民國 96 年 11 月 01 日

# 行政院國家科學委員會補助專題研究計畫 成果報告

## 多代理人式行為追蹤與主動控制系統研究

計畫類別：個別型計畫

計畫編號：NSC-95-2218-E-002-025

執行期間：95年10月1日至96年9月30日

計畫主持人：許永真

共同主持人：洪一平

計畫參與人員：顏嘉楠、李世偉、陳麗徽、吳祖佑、  
連家峻、黃啟嘉

成果報告類型(依經費核定清單規定繳交)：精簡報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：本計畫可公開查詢

執行單位：國立臺灣大學資訊工程學系暨研究所

中華民國 96 年 7 月 31 日

# Hierarchical Everyday Activity Recognition from Heterogeneous Sensors

Chia-Nan Yen Shih-Chieh Yen Jane Yung-jen Hsu

Department of Computer Science and Information Engineering  
National Taiwan University  
e-mail: [yjhsu@csie.ntu.edu.tw](mailto:yjhsu@csie.ntu.edu.tw)

**Abstract**—Activity recognition by fusing information from multiple sensors is a fundamental research question. Existing approaches to activity recognition tend to focus on short-term activities based on a single type of sensor. This research introduces a hierarchical activity recognition framework, in which the first layer detects meaningful sensor events from heterogeneous sensors, while the layer recognizes *activities of daily living* based on the detected sensor events. In particular, the system deciphers the state of a user's activity by integrating information from three types of sensors: voice, location, and object usage. We conducted experiments to collect data at the “NTU E-Home”, a home-like lab with built-in sensors. Our experiments show that using heterogeneous sensors can improve recognition accuracy from 59.52% to 66.03%. Moreover, by using the hierarchical activity recognition framework, the accuracy can be further improved to 77.56%. The results suggest that hierarchical activity recognition from heterogeneous sensors is promising both in terms of robust performance and reusability.

**Keywords** —activity recognition, heterogeneous sensors, multiple sensors, probabilistic reasoning

## I. INTRODUCTION

With the advances in modern medicine and technology, human life expectancy has increased significantly during the last century. Issues associated with the aging population have become critical in nearly every country around the world. One solution is to integrate techniques in information engineering, sensor networks, and artificial intelligent to create health monitoring systems that facilitate successful “aging in place”. [12]

In this paper, a hierarchical framework for recognizing activities in the home setting is presented. The system combines four components for recognizing everyday activities. First, an *audio event detection module* is used to capture audio events in the environment, such as music, speech, silence, or unknown. Second, a *location region detection module* tracks the whereabouts of the user, such as the living room, dining room, or study, inside the house. Third, an *object usage detection module*, which utilizes radio frequency identification (RFID) reader and a set of RFID tags, is used to monitor the object usage of the user. Finally, a *supervised pattern recognition algorithm* is proposed to fuse the information from the three sensor event detection modules, and to provide activity recognition prediction.

## II. RELATED WORK

Research on content analysis has attempted to detect a variety of activities from audio and video data. For example, video-based activity recognition has been applied to tennis [15], American Sign Language [11], and American football plays [7]. Similarly, audio has been used to recognize home activities [5] and bathroom activities [4]. Combination of video and audio has been used to recognize activities in an office environment [9]. Even though using video and audio for activity recognition has shown some promise, this approach also poses some limitations. First, many researchers using this approach have targeted the recognition of specific kinds of activities. However, in a natural environment, there might be highly varied activities taking place. Second, due to privacy concerns, cameras are less acceptable as a monitoring device in a home environment.

Besides video and audio data, a significant portion of work in this area has utilized wearable sensors. Among all, accelerometers have been widely used to detect posture and certain types of physical activities. For example, Bao and Intille used 4 wearable accelerometers to recognize activities such as walking, sitting, standing, and running with high accuracy [1]. In general, however, using only accelerometers is not enough for recognizing complex and long-term activities, which are expected to take longer time without repeated patterns.

Recently, researchers have explored the usage of ubiquitous sensors for activity recognition. They installed many simple binary sensors around the house, and use the collected sensor data to infer the user's activities. In the work done by Tapia et al [13], 77 simple state-change sensors were installed to recognize home activities, such as preparing lunch, toileting, etc. With the property of being cheap and easy to install, simple sensors provide solid potential for activity recognition in real homes. However, this approach can recognize activities only with explicit touching of objects. Therefore, the usage of other sensor modalities should be considered for a robust recognition system.

## III. SYSTEM FRAMEWORK

Rather than using a single type of sensor for short-term activity recognition, this research focuses on long-term activity recognition by utilizing heterogeneous sensors. The difficulty of this problem lies in the gaps between low-level sensor data and high-level human ac-

tivities. As a result, we found that directly combining heterogeneous sensors for activity recognition is simply not enough. Many semantic sensor events have completely different patterns or features at the low level. Different sensor data may denote similar high-level activities. To address this issue, a hierarchical framework is proposed for activity recognition from heterogeneous sensors, which consists of two stages: semantic sensor event detection and activity recognition, as illustrated in Fig. 1.

Consider an activity which has raw sensor data  $(x_1, x_2, \dots, x_t)$ , where  $x_i$  is used to denote the sensor data at time  $i$ . At the first layer, the semantic sensor event detection layer, raw sensor data  $(x_1, x_2, \dots, x_t)$  are used to identify semantic sensor events  $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_t)$ , as shown in Figure 1 (b). At the second layer, the activity recognition layer, the semantic sensor events  $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_t)$  are used to recognize the high-level human activity  $y$ , as shown in Figure 1 (a).

#### IV. IMPLEMENTATION

Our system HEAR, which stands for Hierarchical Everyday Activity Recognition, utilizes three kinds of sensors for everyday activity recognition: audio, location, and object usage. While object usage is already in the form of high-level information, the semantic sensor event detection in HEAR includes *audio event detection* and *location region detection*. The top layer performs activity recognition based on three types of semantic sensor events.

##### A. Audio Event Detection

The set of possible audio events modeled in this system include “music”, “speech”, “silence”, and “unknown”. Two approaches have been implemented in order to compare the performance of different methods for audio event detection. The first approach is a HMM-based approach, which model the temporal relationships between audio features through state transition schemes [10]. The second approach is a SVM-based approach, which constructs a linear decision boundary between classes based on the principle of structural risk minimization [2].

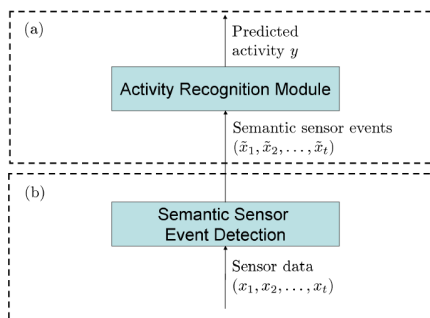


Fig. 1. The HEAR Hierarchical framework.

To better utilize the characteristics of the models, different features have been chosen for HMM and SVM. The features used by HMM are frame-level features. When frame-level features are used for HMM model, the state transition lies within HMM model can be used to capture the temporal variation of the audio signal. Unlike HMM, SVM do not have temporal modeling ability. One solution to this is to use features that already encode temporal information of the audio signals. These features represent properties of audio signal over a much longer period, usually from one second to several tens seconds. Such longer interval are called an audio “clip” or a “window”.

To extract audio features for HMM, a sliding window of 1 second moves through the input audio without overlapping. The signal in each sliding window is further divided into overlapped frames which are the basic units for feature extraction. Each frame is 25-ms long and the overlapping ratio is 0.5. We extracted a 42-dim feature vector from one frame, which includes: volume (VOL), band energy ratio (BER), zero-crossing rate (ZCR), frequency centroid (FC), bandwidth (BW), 13-order mel-frequency cepstral coefficients (MFCC), and the gradients between adjacent frames. These adopted audio features have been widely used in many audio applications, and are known to perform reasonably well [14].

To train the HMM models, the training audio streams are first segmented into overlapping frames. Then, the features described above are extracted from each frame and result into a 42-dimension vector. After that, we use one hidden Markov model (HMM) to represent the statistical patterns of one audio event. The parameters of each HMM are estimated by the well-known Baum-Welch algorithm [10]. The most appropriate state number of the HMM model is found through a 3-fold cross validation process, and the number of mixtures in each state is empirically set as 4. For both training and testing, we have used the Bayes net toolbox (BNT) [8].

To extract audio features for SVM, we first segment the input audio into non-overlapping 1-s clip. Then, a 26-dimension feature vector is extracted from each clip. These features are chosen based on their effectiveness in capturing both temporal and spectral properties of different audio classes. The extracted features include: mean of volume, variance of volume, low short time energy rate (LSTER), mean of zero-crossing rate, variance of zero-crossing rate, range of zero-crossing rate, high zero-crossing rate ratio (HZCRR), mean of the spectrum flux (SF), mean of band energy, variance of band energy, mean of band energy ratio, and variance of band energy ratio.

Before training and testing SVM models, the audio data is first scaled into  $[-1, 1]$ . Then, we use ‘LIBSVM’ [3] for training and testing SVM models. To build the SVM model, the RBF kernel has been adopted and the optimal parameters  $C$  and  $\gamma$  are selected using a grid search and a 10-fold cross-validation. Although SVM was originally used to solve the 2-class classification problem, LIBSVM can decompress the multi-class problem to binary (2-class) problems automatically via the

one-against-one method [6].

### B. Location Region Detection

The semantic events for location data are the location regions, which include “living room”, “dining room”, and “workspace”. To estimate the user’s location, a locating system based on load sensors is used in this system. The locating system includes 40 sensory blocks, each block with one load sensor. The layout of the whole locating system is shown in Fig. 2.

Served as the basis of location region detection, the raw location data is not quite reliable, though. In particular, we found that some sensory blocks might respond non-zero reading even when the user is not on that block. The reason for such situation is because the deformed wooden floor might influence the weight reading value of the load sensors. That is, when the deformed wooden floor get stuck, the load sensors will keep returning non-zero weight reading value, as if the user is on that sensory block. We call this kind of noise the “floor deformation noise”.

To eliminate the effect caused by floor deformation noise, some characteristics of this noise have been observed. We observed the weight reading value comes from deformation noise shows different patterns compared to normal situation. In normal situation (i.e. no deformation noise), there is a noticeable peak in the weight reading of one sensory block when the user step on or leave that sensory block. On the contrary, when deformation noise exists, the weight reading value tend to change smoothly, usually with similar weight readings over a period of time and without a noticeable peak during the whole period of time. The different weight reading patterns in normal situation and in deformation noise are demonstrated in Figure 3, where the left and middle figure is weight reading value in normal situation, and the right picture is in deformation noise.

With the properties of floor deformation noise described above, we first identify the sensory blocks which have significant change in weight reading value. Then, we use a voting scheme to estimate the most possible region user is located in. At each second, we estimate the most possible location region of the user as the region with most active sensory blocks within the decision window.



Fig. 2. The layout of the location regions in HEAR.

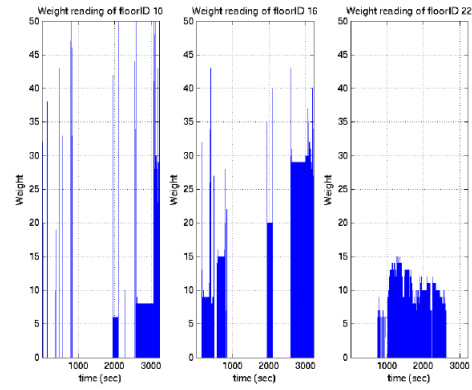


Fig. 3. Weight reading from the locating system shows different patterns in normal case and deformation noise.

### C. Activity Recognition

At the activity recognition layer, the second layer of the hierarchical framework, the semantic sensor events detected in the first stage are used to represent the characteristics of human activities, as shown in Figure 4.

To train the HMM model, the training data is first segmented into separate activities. Then, sensor events (audio event, location region, and object usage) are detected using corresponding sensor event detection modules. After that, we use one HMM to represent the statistical patterns of one activity. The parameters of each HMM are estimated by the Baum-Welch algorithm [10]. The most appropriate state number of the HMM model is found through a 3-fold cross validation process.

To recognize user’s activity, we first detect the semantic sensor events from the raw sensor data. With the sequence of recognized semantic sensor events, the posterior probability of each HMM model given this sequence is calculated. We say one activity is recognized if its HMM model has largest posterior probability.

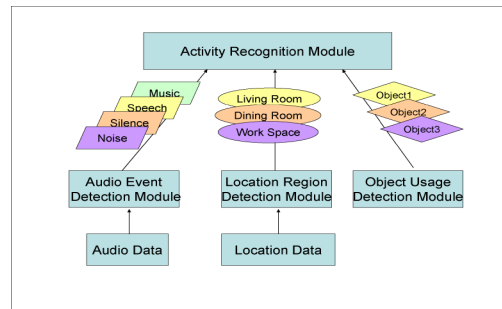


Fig. 4. Hierarchical activity recognition consists of semantic sensor event detection modules for different kinds of sensors.

## V. EVALUATION

To collect data about ADLs, a semi-naturalistic data collection experiment have been conducted. Semi-naturalistic data collection is a compromise between laboratory and naturalistic data collection. In the semi-naturalistic data collection, subjects ran an obstacle course consisting of a series of activities. The obstacle course was designed to minimize subject awareness of data collection. For instance, subjects were asked to “use

the web to find out what movie you would like to see this weekend” instead of being asked to “work on a computer”. During the obstacle course, subjects had freedom in how to perform each obstacle, and there was no researcher supervision.

In the semi-naturalistic data collection, the subjects wore one glove-based RFID reader and 4 accelerometers, which are used to detect the objects touched by the subjects and to detect the acceleration at different parts of the body respectively. During the experiment, the subjects are asked to follow the instructions of the obstacle course given in a slide, and to record the time they started and finished each obstacle. The obstacle course include 15 obstacles, each contain one ADL. Data collected between the start and finish time of one obstacle were labeled with the name of the activity behind that obstacle.

The semi-naturalistic data collection experiments are conducted in a home-like environment named “E-Home” (abbreviation for “electronic Home”). E-Home is a prototype for future living environment. Figure 5 shows the environment of the E-Home. This environment contains household appliance and furniture such as a TV, a sofa, a refrigerator, a micro-wave, a telephone, etc. Besides, several kinds of sensing technologies are installed in this environment as well. For example, several cameras and one microphone are installed in the corner of the E-Home environment. The floor of E-Home is equipped with weight sensor such that the user’s location can be estimated. Many objects in the E-Home are equipped with RFID tags, including TV remote control, cup, telephone, mouse, etc.



Fig. 5. The environment of the E-Home

#### A. Single Sensor and Heterogeneous Sensors for Activity Recognition

In this experiment, we use one-layer HMMs for three kinds of sensor separately. After that, we use one-layer HMMs using the three kinds of sensors altogether. The recognition accuracy is shown in Table 1. As shown in the table, object usage data can provide the most useful information for activity recognition among the three kinds of sensors. Nevertheless, when the information from three kinds of sensors is fused, an even better performance can be achieved. For the data collected from the E-Home experiment, the accuracy of using three kinds of sensors is 66.03%, while the best performance of using only one kind of sensor is 59.52% (using object usage data).

Table 1. Performance of activity recognition using single type of sensor versus using heterogeneous sensors.

Sensor	Overall Accuracy
Audio	46.15%
Location	45.00%
Object	59.52%
Audio+Location+Object	66.03%

#### B. Performance of Audio Event Detection

To evaluate the SVM-based approach and HMM-based approach for audio event detection, we label about 15 minutes data for each audio event class. Then, a 3-fold cross-validation is performed (i.e. 10 minutes training and 5 minutes testing at each iteration for each audio event) for HMM and SVM separately.

The overall accuracy for these two approaches is compared in Table 2. As shown in the table, the overall accuracy of HMM-based approach is about 77%, while the overall accuracy of SVM-based approach is about 78%. In both approach, the recall of “unknown” is particularly worse than others. There are two possible reasons. One reason is that some noise sounds which belong to “unknown” class are very short, and the training data are not labeled precisely enough. For example, the sound of “closing the cabinet” (which belongs to the “unknown” class in this work) is shorter than the length of one basic audio event unit 1s, and there are some silence period within the training data for “unknown” class. Another reason for the low recall rate for “unknown” is the diversity of sounds belonging to “unknown” class, which might result in varied patterns of audio features. The sounds defined as “unknown” include the sound of “step”, “opening or closing cabinet”, “opening or closing a container”, “operating appliance”, etc. If we neglect the data of “unknown” class and perform a 10-fold cross-validation on the data of other three audio event classes, the average precision of the SVM-based approach can achieve as high as 95.39%.

Overall, the SVM-based and the HMM-based approach have competitive performance in the audio event detection task. Besides, these two approaches both provide a promising basis as an audio event detection classifier for higher-level activity recognition. For SVM-based approach takes much less time for training and testing, this approach have been adopted for audio event detection.

Table 2. Performance of activity recognition using single type of sensor versus using heterogeneous sensors.

Audio Event	Accuracy (a)	Accuracy(b)
Music	89.87%	98.20%
Speech	73.11%	73.33%
Silence	64.20%	100.00%
Unknown	86.21%	56.80%
Overallt	76.96%	77.70%

### C. Performance of Activity Recognition

To evaluate the hierarchical activity models, we compare the performance of one-layer activity models and the performance of hierarchical activity models using the data collected from our E-Home experiment. In one-layer activity models, the raw features of audio, the active floor (after removing the deformation noise), and the object usage data are used directly for activity recognition. The classifiers were trained on E-Home experiment data for all subjects except one, and tested on the data for the only subject left out of the training data set. This leave-one-subject-out validation process was repeated for all 13 subjects.

The overall accuracy comparison of the one-layer activity models and the hierarchical activity models is given in Table 5. As shown in the table, the accuracy improves for most activities when the hierarchical activity models are used. Overall, the accuracy of one-layer activity models is 66.03%, while the accuracy of hierarchical model is 77.56%.

If we look more closely, we can find that the activities with significant improvement include “watching TV” (from 69.23% to 71.43%); “preparing meal” (from 61.11% to 81.82%). The results are reasonable because these activities are usually with diverse patterns in sensor data. For example, the audio data of “watching TV” might contain music, speech, and some special sound effects. In addition, the sound of speech in TV does not belong to a specific person. Instead, the speech sound in TV might be produced by people of different sex and age. In “preparing meal”, the audio data contains many sudden sounds and noise. The sudden sounds and noise contain many kinds of sound, such as “opening the plastic bag”, “opening the container”. The sound is so diverse that it is hard to learn a model which is general enough to inference high-level activity from such complex and noisy low-level sensor data.

To sum up, with the high complexity in sensor data, some activities might need more complex activity models (for example, more HMM states) or more training data, or even both. In contrast, with the hierarchical structure, we can decrease the complexity in sensor data by integrating sensor data into some meaningful semantic sensor events; therefore increase the accuracy without the need of more complex activity models.

The hierarchical activity model is not perfect, though. As we can see in Table 5, the accuracy of mopping drops from 100.00% to 58.82%. After analyzing the data of mopping, we found the reason for the decreased precision is because some information has been lost during the process of location region detection. In some cases, the trigger of some sensory blocks (for example, those near the location where the mop is usually put) can serve as a very strong evidence of user’s activity (for example, mopping), since these sensory blocks are less likely to be triggered in other activities. However, after location region detection, these floors are recognized as part of the living room, and lose the special information used to contained in the raw floor data. We think this problem can be avoided by defining the location region more properly.

Table 5. Overall performance of activity recognition by using (a) using one-layer activity model, (b) using hierarchical activity model

Activity	Recognition Accuracy(a)	Recognition Accuracy(b)
Listening to music	44.83%	84.62%
Watching TV	69.23%	71.43%
Reading book	66.67%	92.31%
Using telephone	100.00%	91.67%
Preparing meal	61.11%	81.82%
Preparing drink	62.50%	73.33%
Eating meal	50.00%	77.78%
Resting	83.33%	73.33%
Mopping	100.00%	58.82%
Taking out the trash	83.33%	90.00%
Taking medicine	60.00%	69.23%
Using computer	88.89%	78.57%
Overall	66.03%	77.56%

## VI. CONCLUSION

This paper proposed a multi-sensor and hierarchical approach for human activity recognition in a home environment. First, by fusion information from heterogeneous sensors, the system is able to utilize the different properties of human activities, such as sound, location and object usage. We have shown that by utilizing heterogeneous sensors, we can recognize activities that can not be recognized with single type of sensor. In particular, the accuracy increases from 59.52% (best accuracy using one sensor) to 65.00% (accuracy using all three sensors) when more types of sensors are used. Second, we have introduced a hierarchical activity model, which encodes more human knowledge and has the ability to capture different layers of abstraction from sensor data. This hierarchical structure appears to be an appropriate way for fusing information from different sensors. In our experiment, the accuracy can be further improved from 65.00% (using all three sensors) to 76.92% when the hierarchical structure is used.

## ACKNOWLEDGMENT(S)

This research is funded by the National Science Council in Taiwan under grants NSC 95-2218-E-002-025, NSC 95-2218-E-002-026 and NSC 94-2218-E-002-057, and the NTU Research Excellence Program. The authors would like to thank Prof. Li-Chen Fu and his students for supporting our experiments at the NTU e-Home.

## REFERENCES

- [1] L. Bao and S.S. Intille. “Activity recognition from user-annotated acceleration data”. In *Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004)*, pages 1–17, 2004.
- [2] C.J.C. Burges. “A tutorial on support vector machines for pattern recognition”. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [3] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [4] J. Chen, A.H. Kam, J. Zhang, N. Liu, and L. Shue. Bathroom activity monitoring based on sound. In *Proceedings of the 3rd*

*International Conference on Pervasive Computing (Pervasive 2005)*. Springer, 2005.

- [5] J. Fogarty, C. Au, and S.E. Hudson. Sensing from the basement: A feasibility study of unobtrusive and low-cost home activity recognition. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST 2006)*, pages 91–100, New York, NY, USA, 2006. ACM Press.
- [6] C.-W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.
- [7] S.S. Intille and A.F. Bobick. A framework for recognizing multi-agent action from visual evidence. In *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI 1999)*, pages 518–525, Menlo Park, CA, USA, 1999. AAAI Press.
- [8] K.P. Murphy. The Bayes net toolbox for MATLAB. *Computing Science and Statistics*, 33:331–351, 2001.
- [9] N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI 2002)*, pages 3–8, 2002.
- [10] L.R. Rabiner. “A tutorial on hidden Markov models and selected applications in speech recognition”. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [11] T. Starner and A. Pentland. Real-time American sign language recognition from video using hidden Markov models. In *Proceedings of the 1995 IEEE International Symposium on Computer Vision*, pages 265–270, 1995.
- [12] P. Tang and T. Venables. “Smart homes and telecare for independent living”. *Journal of Telemedicine and Telecare*, 6(1):8–14, 2000.
- [13] E.M. Tapia, S.S. Intille, and K. Larson. “Activity recognition in the home using simple and ubiquitous sensors”. In *Proceedings of the International Conference on Pervasive Computing (Pervasive 2004)*, pages 158–175, 2004.
- [14] Y. Wang, Z. Liu, and J.-C. Huang. Multimedia content analysis-using both audio and visual clues. *IEEE Signal Processing Magazine*, 17(6):12–36, 2000.
- [15] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden Markov model. In *Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 1992)*, pages 379–385, 1992.

# 行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2007年 5月 18日

報告人姓名	許永真	服務機構及職稱	國立台灣大學 資訊工程系所暨網路與多媒體研究所
時間 會議地點	2007年5月14~18日 Honolulu (美國 夏威夷)	本會核定 補助文號	
會議名稱	(中文) ACM 代理人基礎之普及計算研討會 2007年自主性代理人與多代理人系統國際會議 (英文) ACM International Workshop on Agents Based Ubiquitous Computing 2007 AAMAS 2007 (The 2007 International Joint Conference on Autonomous Agents and Multi Agent System)		
發表 論文題目	(中文) 智慧型空間之情境感知服務平台 (英文) A Context-Aware Service Platform in A Smart Space		
<p><b>一、參加會議經過</b></p> <p>AAMAS (The International Joint Conference on Autonomous Agents and Multi Agent System)是自主性代理人與多代理人系統的國際會議，每年舉辦一次盛會。自2002年起，此會議由 AGENTS (International Conference on Autonomous Agents)、ICMAS (International Conference on Multi-Agent Systems)、和 ATAL (International Workshop on Agent Theories, Architectures, and Languages) 三個國際研究組織所組成，目標是希望能提供一個高品質的學術研究交流會議，2006年的會議在日本函館舉行。本屆 AAMAS 2007 會議從5月14日至5月18日，歷時共五天，於風光明媚、氣候宜人的美國夏威夷舉行。會議地點則選擇在夏威夷國際會議中心舉行。</p> <p>在此會議底下並有許多相關議題的 workshop 研討會，而本人此行主要目的即參加 ACM International Workshop on Agents Based Ubiquitous Computing 2007 研討會，並發表論文。本屆 AAMAS 2007 會議從5月14日至5月18日，歷時共五天，而 ABUC workshop 則是5月14日為期一天研討會。均討論與 Ubiquitous Computing 相關的研究主題，本人發表論文題目是『A Context-Aware Service Platform in A Smart Space』有幸榮獲最佳論文獎。在開會期間與其他參加學者接觸，進行學術交流，直到整個會議結束，於5月18日返國。</p> <p><b>二、與會心得</b></p> <p>本次5/14日的 ABUC workshop 的會議，集合許多相關研究領域的人員來發表和討論自身目前所正在進行的研究構想和論文，而且提供一個開放自由的空間給對相關研究有興趣的學者進行學術上的交流。本報告人在此會議裡所發表的論文『A Context-Aware Service Platform in A Smart Space』也得到在場學者們的意見和建議，並進一步探討可能的發展性。會議期間本人積極參與諸多學者探討相關領域的問題，得到不少的寶貴意見以供日後研究的研究規劃與方向參考，實為珍貴的經驗。</p> <p>今年大會邀請了產業界及學術界學者來給 keynote speech。包括 Motorola 副總 John Strassner 演講關於代理人及 mobility services 之研究；IBM 研究中心的 Jeffrey O. Kephart 博士介紹與 agents for automatic computing 相關之研究。Bar Ilan University 的 Gal Kaminka 教授以“Robots Are Agents, Too!”為題，介紹代理人應用於機器人之概念；Sarit Kraus 教授，則探討代理人如何應用於 Automated Negotiation。</p> <p>除了 keynote speeches 外，在5/16日至5/18日的 AAMAS 國際會議本報告人也參加不</p>			

少的 session。本此會議主題包含與 agent 相關的各領域之研究，例如：Cognitive Models for Agents、Communications and Commitments、Partially Cooperative Multiagent Systems、Mechanism Design and Game Theory、Trust and Reputation、Cooperation, Coordination, and Teamwork、Multiagent Planning、e-Commerce, e-Government、Societal Aspects、Multiagent Learning，聽到許多與本身研究領域相關的論文報告。

其中 Expectation-driven trust 利用一個十分生動的動畫來介紹 Trust Expectation 的影響力，與會者均很欣賞這個有創意又有說服力的表達方式。此外，關於 Multi-agent learning 使代理人藉由簡單的任務去學會處理較大的任務，例如：由 3 對 2 的遊戲而學會處理 4 對 3 的遊戲。本人在各個 session 之間與來自各地的學者互動頻繁，藉以認識更多相關領域的學者並提升本校系所在國際上的知名度，故對於整個會議過程，本報告人深感獲益良多。

### 三、建議

AAMAS 國際會議會為自主性代理人與多代理人系統相關領域中的重要會議之一，發表的論文涵蓋面甚廣，且具有一定的深度，諸多學者人士雖未能夠發表亦參與此會議尋找相關資料，或與論文作者直接討論。而在此會議下舉辦許多相關的 workshop 研討會，規模雖然沒有主會議來的盛大，但也都有相當不錯的水準，此外更提供主會議無法提供的直接探討方式。本次會議來自台灣的學者並不多，但來自其他亞洲各地的學者卻不少，像是來自日本、中國大陸、新加坡等，這是本報告人認為比較可惜的地方，建議國內相關研究領域人員可考慮投稿，汲取參與高品質國際學術會議之經驗，也希望過內相關單位能給予適時的補助，提升過內學者與學生外出參與大型國際會議的經驗，以提升國內學術機構在國際上的知名度。

### 四、攜回資料名稱及內容

- 大會議程
- ABUC Workshop Proceedings
- Trust in MAS Workshop Proceedings
- AAMAS 2007 Proceedings (CD-ROM)
- CFP for AAMAS 2008
- CFP for PRIMA 2007

### 五、其它相關的國際研討會

- AAAI (Association for the Advancement of Artificial Intelligence)
- PRIMA (Pacific Rim International Workshop on Multi-Agents)
- IEEE Service-Oriented Computing and Applications