2006 IEEE International Conference on Systems, Man, and Cybernetics
October 8-11, 2006, Taipei, Taiwan

Visualization Design for Location-Aware Services

Chia-How Lin, Kai-Tai Song, Sheng-Po Kuo, Yu-Chee Tseng, and Yau-Jen Kuo

Abstract—**This paper presents a design and implementation of a location-aware service system, which combines a radio-frequency-based positioning engine and a multimedia human-machine interface (HMI). In this system, a number of Zigbee sensors are deployed for the localization purpose. Users' locations can be estimated according to their received signal strength samples from the Zigbee sensors. In order to improve the positioning efficiency, a hierarchical positioning architecture is proposed. The multimedia HMI is to help users to access the provided services. A speech recognition system is designed as the input of the system and a visualization system has been developed as the output interface of the location-aware system. Moreover, the visualization service is integrated with the Open Service Access (OSA) platform via mobile network systems. The whole system can be directed to future homecare applications.**

Keywords: Homecare, Location-aware service, Wireless networks, Location sensing.

I. Introduction

As a result of the progress of medical technology and the reduction in birthrate in recent years, the population ratio of the elderly has been increasing in many countries. Elderly-care services will soon become essential and cause social problems in the near future. Technology can play an important role by providing a more intelligent home environment to reduce the man power needed for such labor-intensive work and increase the quality of life of the elderly. One interesting area is the assistance provided by various context-aware systems. For example, an image surveillance system can transmit videos from home, so one can monitor the elderly through the Internet. However, one will still need to control the camera to find the elderly since the system is not able to find the location of the elderly automatically. An enhance visualization design using location information provides such services deserves urgent attention.

This paper emphasizes a human-machine interface for location-aware services. Fig.1 illustrates the architecture of the system. A family member (user B) can request or subscribe the images and location information of the elderly (user A) through e-mail or cell phone. The location is then estimated by the location server, and sent to the video server. The video server controls all the cameras in the house to acquire images. The selected camera will take pictures from the elderly according to the location information provided by the location server. Finally, the information will be sent to the user B through an Open Service Access (OSA) Gateway[1]. In addition to passive monitoring, the elderly can also ask for help via the speech recognition system. His/her location can be found and image can be sent to the user B in real time.

In the proposed system, the location is estimated by using radio signals transmitted or received by the active beacons such as WiFi or Zigbee modules. Location estimation is obtained by a pattern matching approach [2-4]. This approach is more proper to handle uncertainty and errors in signal strength measurement. However, such approach often incurs large computation overheads when the location database is large. In order to reduce the computation load for matching the signal strength patterns, a hierarchical positioning architecture is proposed. Under such architecture, the samples with similar characteristics will be classified into a cluster. In the positioning phase, the observed signal-strength sample will be compared with the pattern of each cluster and classified into the most likely one first. Then, the pattern matching process is launched again with the samples only belonging to this cluster. Because the samples in a cluster are much less than the whole ones in the location database, the positioning efficiency can be improved substantially.

On the other hand, considering that most elderly are not used to input devices such as keyboards, mouse or complicated touch panels, the system is also expected to provide an intelligent human-machine interface for easy and efficient access of the location results. A user can thus manipulate the system by speech and observe the location on a panel screen the image acquired by a camera. Flexibility and connectivity is also important for location-aware services. Many works have paid attention to creating an extendable framework with generalized application programming interface (API), in order to share the location information and integrate the location services with other useful applications. In this work, the location query service adopts the open and platform-independent Mobile Location Protocol (MLP) to communicate location information. Further, the video/audio information is integrated with an OSA platform, which provides an industrial standard of a common platform for value-add services via mobile network system such as GPRS.

This work was supported by the Institute for Information Industry, Taiwan R.O.C. under grant (94)0357.

Chia-How Lin and Kai-Tai Song are with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan R.O.C. (e-mail: jotarun.ece87@nctu.edu.tw, ktsong@mail.nctu.edu.tw).

Sheng-Po Kuo and Yu-Chee Tseng are with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan R.O.C. (e-mail: spkuo@cs.nctu.edu.tw, yctseng@cs.nctu.edu.tw).

Yau-Jen Kuo was with Institute for Information Industry, Taipei, Taiwan R.O.C. (e-mail: yzgao@iii.org.tw).

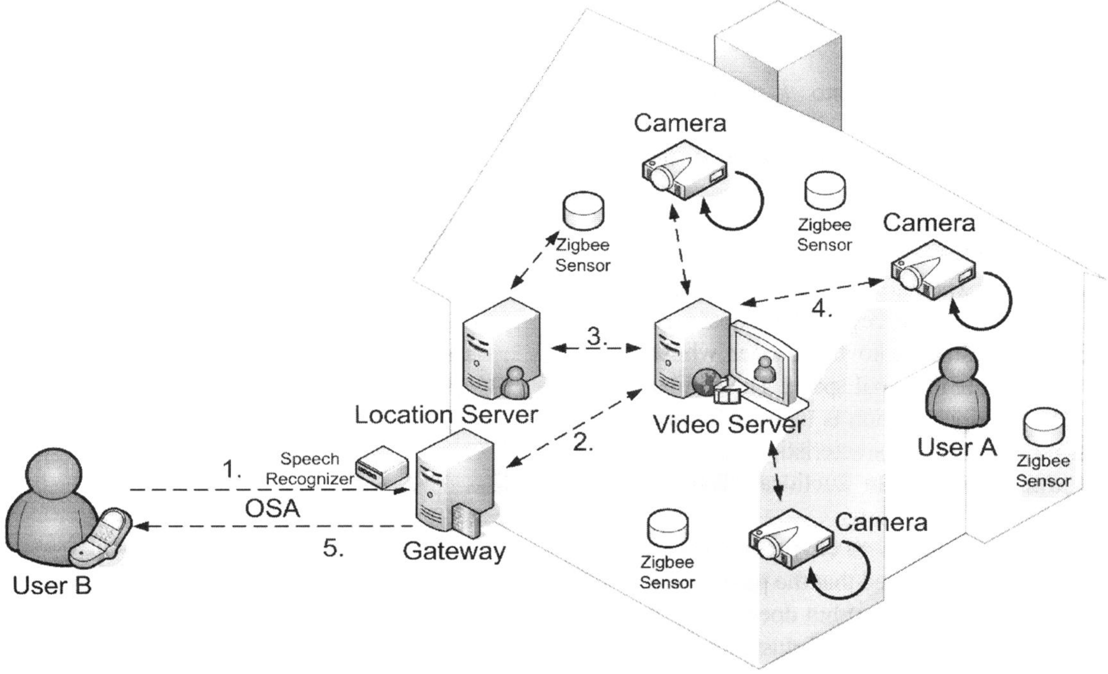

Fig.1. System architecture of the location-aware service system

The rest of the paper is organized as follows: the methods and functions of each subsystem are described in Section 2. In Sections 3 practical experiments of the integrated system are presented. The conclusions are summarized in Section 4.

II. Proposed System Framework

The developed location aware system consists of three main sub-systems: positioning engine, multimedia HMI, and OSA platform.

A. Positioning Engine

The positioning engine in this paper adopted the pattern matching algorithms presented in [6], [7]. In such design, a set of access points which can transmit radio signal for localization purpose are placed in the environment. A number of signal strength measurements are performed at some training locations, and then the measured signal samples and their corresponding training locations will be recorded in a location database. Afterward, given a signal sample with unknown location, the localization algorithm is able to estimate its location by classifying the signal sample to the training location which has the most likely signal characteristics in the location database. The positioning engine developed in this paper can be divided into three main parts: location estimation module, database enhancement method, and API.

1) Location estimation module

The location estimation module based on pattern matching method usually contains two phases. In the training phase, a set of Zigbee sensors $Z = \{z_1, z_2, ..., z_n\}$ are installed in a field. Within this field, a set of training locations $L = \{l_1, l_2, ..., l_m\}$ is defined and at each of which the signal strength is measured from each access point for a period of time to build up a location database.

In the positioning phase, given a set signal strength samples $S = [s_1, s_2, ..., s_n]$, the objective is to match the set to an training location in the database. The approach in this work is adopted from the nearest neighbor in signal space (NNSS) algorithm mentioned in [6] to match S to the most likely location. In NNSS, for each training location l_i, the location database will maintain a characteristic vector $C_i = [c^i_1, c^i_2, ..., c^i_n]$ such that c^i_j is the averaged signal strength received from z_j. Given a signal strength sample S, the NNSS algorithm computes the *Euclidean distance* in signal space between S and the characteristic vector of each training location l_i in the location database, for $1 \le i \le m$, i.e.,

$$dist(S, l_i) = \| S, l_i \| = \sqrt{\sum_{j=1}^{n} (s_j - c^i_j)^2} \quad (1)$$

According to the Euclidean distance of each training location, the one that has the minimum Euclidean distance will be determined as the estimated location.

2) Database enhancement method

Scalability is one of the critical issues for the localization algorithms based on pattern matching. When the number of training locations is large, the computation cost will be high. Hence, a novel clustering algorithm is proposed to improve the efficiency of the pattern matching process.

The basic idea of clustering is to format the signal strength measurements in the location database into a hierarchical structure. Similar samples are clustered beforehand during database buildup. Thus, the system is able to compare only the samples in a cluster rather than all the training data during the matching phase.

The clustering is implemented with *K-Means* algorithm. Given an integer k, the clustering algorithm can be summarized as follows:

1. Partition characteristic vectors into k nonempty subsets;
2. Compute mean as the centroids of the clusters of the current partition;
3. Relocate each characteristic vector to the nearest centroids of the cluster;
4. Go back to Step 2, stop when no more new relocation.

In the estimating-locations process, the observed signal-strength sample is classified into the cluster whose centroid is closest to the sample in signal space. The similar process mentioned in the previous section is then performed by comparing the sample with the characteristic vectors in this cluster. The one with the minimum Euclidean distance can thus be found.

Fig. 2 shows the simulation result in an environment with 600 training locations. It can be observed that the positioning accuracy of the proposed clustering algorithm does not have apparent negative effect when the number of clusters is less than 100. However, the clustering technique can substantially decrease the computation cost in the positioning phase.

3) Location Query API

The positioning engine provides a web container to accept queries from clients. The query format follows the Mobile Location Protocol 3.1 defined by Open Mobile Alliance (OMA) [9]. The protocol stack is shown in Fig. 3. A client can query the location of a target recorded in the location server, via the LBS API on a PC or the JSR-179 location API on a cell phone. These APIs will construct the corresponding XML message according to the query request and transmit it to the location server through the HTTP protocol.

B. Visualization System

The coordinates estimated and provided by the *Positioning Engine* are in format of numbers, which is difficult for human users to understand. Therefore in practice a usable system needs to represent the results in a more comprehensible way. The most common method nowadays is to use figure or images such as floor plan or map. However, in order to monitor the current state of the elders, the located elderly will be actually shown on the physical image obtained from a pan-tilt camera. To achieve this feature, one needs to apply camera calibration technique. Using camera calibration, the projection between world and image coordinates can be obtained. In order to increase the coverage area of the camera, the pant-tilt ability of the camera is also necessary. The system first calibrate the camera and build the camera model, which represents the mapping from the world coordinate to the image plane and the pan-tilt movement of the camera.

The calibration process starts from the geometric camera calibration procedure, which aims to estimate the intrinsic and extrinsic parameters of a static pin-hole camera model. The implemented algorithm is the linear estimation method given in [10]. Assumed that the target environment observed

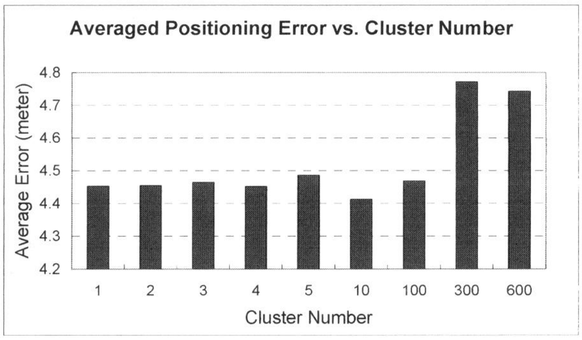

Fig. 2. The simulation result of the averaged positioning error when the number of clusters increases.

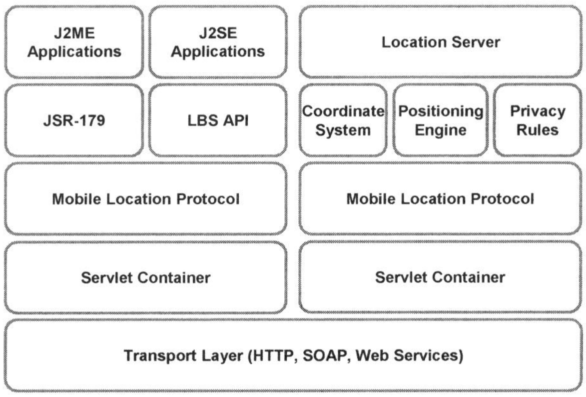

Fig. 3. Protocol stack of the location interface

by a camera and the image position $p_i=(u_i,v_i)^{\mathrm{T}}$, of n feature points with known world coordinate vectors $Pi=(x_i,y_i,z_i)^{\mathrm{T}}$ are matched in the image, $i=1\ldots n$ .The projection between world and image coordinates can then be modeled by

$$\mathbf{p}_i=\frac{1}{z_i}\mathbf{M}\mathbf{P}_i \text{ ,where } \mathbf{M}=\begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\\ \mathbf{m}_3\end{bmatrix}, \mathbf{p}_i=\begin{pmatrix}p_i\\1\end{pmatrix}, \mathbf{P}_i=\begin{pmatrix}P_i\\1\end{pmatrix} \quad (2)$$

M is decomposed to **M**=**K**(**R** **t**), where (**R** **t**) represents the extrinsic parameters, including rotation matrix **R** and translation vector **t**. **K** represents intrinsic parameters, including skew and aspect ratio. Collecting the constraints from (2) associated with n pairs points yields a system of $2n$ homogeneous linear equations in 12 coefficients of the matrix **M**-namely, $\underline{\mathbf{P}}\cdot\mathbf{m}=0$,where

$$\underline{\mathbf{P}}=\begin{bmatrix}\mathbf{P}_1^T & \mathbf{0}^T & -u_1\mathbf{P}_1^T\\ \mathbf{0}^T & \mathbf{P}_1^T & -v_1\mathbf{P}_1^T\\ \cdots & \cdots & \cdots\\ \mathbf{P}_n^T & \mathbf{0}^T & -u_n\mathbf{P}_n^T\\ \mathbf{0}^T & \mathbf{P}_n^T & -v_n\mathbf{P}_n^T\end{bmatrix}, \text{ and } \mathbf{m} = [\mathbf{m}_1^T\ \mathbf{m}_2^T\ \mathbf{m}_3^T] \quad (3)$$

As long as $n\geq 6$ (since $\underline{\mathbf{P}}$ is a $2n$ by 12 matrix), the homogeneous linear least-squares can be exploited to compute the unit vector **m** that minimize $|\underline{\mathbf{P}}\mathbf{m}|^2$ and thus estimate **M**. In this system, the n correlation points are marked manually from the captured image scene with a GUI interface. To increase the accuracy and avoid outliers, the Random Sample Consensus (RANSAC) algorithm [10] is also implemented. This algorithm picks n points from a large set of sample points randomly and the estimate the matrix iteratively until the fitting error is below the expected criterion. In the experiment, the average error is around 2 pixels.

The second step is to estimate the influence brought by the pan-tilt movement. In the proposed system a pan-tilt camera model, as shown in Fig.4, is used. A rotation around the x-axis, $\mathbf{R}_x$, and the y-axis, $\mathbf{R}_y$, correspond to tilt and pan respectively. With this model, the projection between world and image coordinates can then be modified as:

$$\mathbf{p}_i = \frac{1}{z_i}\mathbf{M}\mathbf{R}_y\mathbf{R}_x\mathbf{P}_i \quad (4)$$

The choice of this simplified camera model is a tradeoff for straightforwardness over accuracy. The model assumes that the center of rotation of the camera is fixed and coincides with the lens center of projection during operation, which many cameras may violate such an assumption. However, in this application, the deviation of the center is negligible.

Another problem is that many low-cost cameras cannot provide the actual pan-tilt angle. To overcome this problem, the kinematics model of the camera has to be calibrated beforehand. The kinematics calibration of this system is realized by randomly choosing several poses of the camera, and performing the camera calibration procedure. The kinematics model of pan and tilt can then be estimated from alternative results of calibrations iteratively. Furthermore, in order to guarantee that the target cab be shown completely in the image, the system will first query the height and width of the target, calculate its projected area on the image plane, and then move the camera to ensure that all the points are shown, if possible. A target can then be marked on the image according to its location.

C. Speech Recognition system

Considering that elderly members are generally few in a family, say around 3 persons, the mode of the speech recognition system is set as speaker-dependent. The command and user index are pre-defined integers to represent

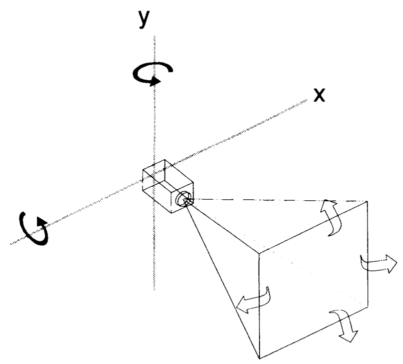

Fig. 4 . A Simplified model of the pan-tilt camera

the recognition results of the input speech signals. The recognition algorithm is based on the dynamic time warping (DTW) method[11]. The procedure of the speech recognition is divided into three parts.

The first part is the process of speech information sampling, which is the setting of gain control and sampling frequency when analog signals transforms to digital signals. The second part is the extraction of speech features. The endpoint detection process can determine the location of real speech signals by the short time energy detection and zero-crossing rate detection. The started 128 samples are used to determine the threshold value of the energy detection and zero-crossing rate detection. When the real speech signals are determined, they are divided into a 16ms length frame. Each frame has an overlap region. In order to do the encroachment of the high frequency spectrum, the pre-emphasis work is adopted. Considering the continuity of signals at two sides of a frame; Hamming window is used on every frame. It can hold the signals at mid parts and press down the signals at two sides. The Mel Frequency Cestral Coefficients (MFCCs) is extracted for every frame. A feature vector represents each frame. The DTW recognizer compares the input feature vectors with the reference speech and gets a minimum matching error as the recognition result. The third part is the process of feature matching by the dynamic time warping method. The speech recognition system is implemented in a TI TMS320VC5402 DSK board.

D. OSA platform

The OSA API is an open, standardized interface for applications to use the capabilities of a network without owning it or knowing its technology. It consists of a framework, which is in charge of access control and service discovery; and some Service Capability Features, which map to network capabilities. It is specified and standardized in the Joint API Group, with participation of 3GPP, ETSI and the Parlay Group: a single API for the whole developer community.

Parlay/OSA [1] enables operator and 3rd party applications to make use of network functionality through a set of open, standardized interfaces. This leads to several advantages such as shorter time to market (TTM) for applications, network independency, etc. Parlay/OSA Gateway consists of several Service Capability Servers (SCS): functional entities that provide Parlay/OSA interfaces towards applications. Each SCS is seen by applications as one or more Service Capability Features (SCFs): abstractions of the functionality offered by the network, accessible via the Parlay/OSA API. Sometimes they are also called services. The Parlay/OSA SCFs are specified in terms of interface classes and their methods.

In this system, the OSA gateway is connected with the location and visualization system in order to extend the service to mobile network. Fig. 5 shows components applied in this system. The OSA application server receives the request from users using mobile terminals, and request data from the OSGi gateway at home. The data can then be sent to the user through two different paths: by e-mail via the internet, or by MMS via cellular network.

III. Experimental Results

In order to evaluate the capability and practicability of the system, the system is tested in a sample room. Fig. 6 shows the layout and the physical scene of the room. The pan-tilt camera DCS-5300G from D-Link was installed in the middle of the room. Nine Zigbee modules were placed around the room, while three users carrying the Zigbee beacons.

A. Examining the Positioning Engine

The goal of this experiment is to test the accuracy of the estimated locations. 9 training locations is collected in the location database. Users wearing Zigbee beacons move around the room, while the system tries to keep positioning their locations. Fig. 7 shows the experimental result. The averaged location error of 1.86 meters was recorded in this experiment.

B. Examining the visualization system

This experiment aims to investigate the capacity of camera calibration when used to mark a person in the image. Eight positions were assigned on the floor, to emulate the result of the positioning engine. As shown in Fig. 8, given the exact coordinate, the camera successful turns and marks the person (in the bathroom) on the image. Thus the localized person can be visualized and the picture can be transferred through the OSA platform.

C. Examining the speech recognition system

As describe in the first section, it is desirable that an elderly can ask for help via the speech recognition system. Therefore, the system has to at least distinguish who the elderly is, and what he or she needs. The speech recognition system is tested with 3 different names and 6 phrases, for demonstration. Table I shows the result of the recognizing names of three persons. Table II shows the result of recognizing some daily-life Chinese phrases. The recognition rates are satisfactory for practical applications.

D. Integration test

The integrate system has been tested to exam the scenarios

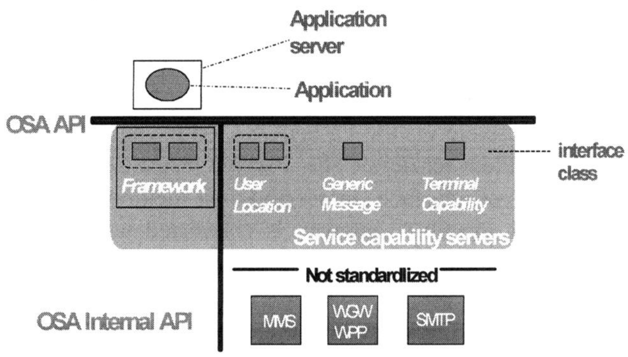

Fig. 5. OSA network services applied in the proposed system


Fig. 6. The scenario of experimental setup

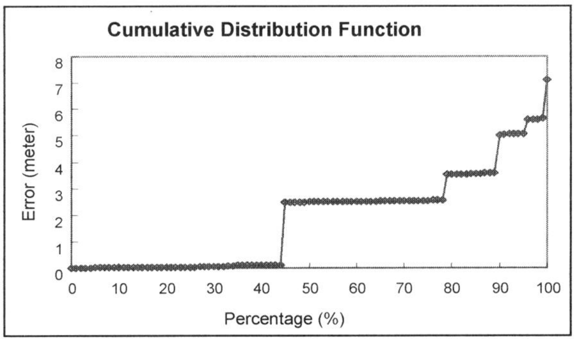

Fig. 7. The experimental result of the positioning engine

described in the first section: location request and "asking-for-help". However, the marker function is neglected due to the accuracy of the Positioning Engine. In the first scenario, several people with Zigbee beacons move around some room simultaneously. A user then requests to watch these people from the cell phone. The camera successfully moves to the target person and the user retrieves the target image for each request successfully. In the second scenario, several people wearing a headset microphone, as shown in Fig. 9. They took turns saying their name, to pretend they were in a situation to ask for help. The speech recognition system successfully recognized these people, translated the messages to the location server, and launched the camera to rotate and capture image of the person. The images successfully transmitted to other users' cell phones and PC via e-mails.

IV. Conclusions and Future Work

This paper proposes a location-aware system combining a RF-based positioning engine and a multimedia HMI. A hierarchical positioning architecture is applied in order to improve the positioning efficiency. The multimedia HMI facilitates a user to visualize the located target, and command the system by speech. A home-care application has been demonstrated to evaluate the performance and possible applications of the system. The developed RF-based positioning engine has been demonstrated its advantages in practical applications, but also reveals its problem on not accurate enough in position estimation. It would be beneficial to increase the accuracy of the visualization system by marking the target using image recognition and feature

TABLE I

EXPERIMENTAL RESULTS OF RECOGNIZING THE NAME

	ChiaHow	Sheng Po	Jin Huai	Recognition Rate
Chia How	25	0	0	100%
Shenh Po	0	23	2	92%
Jin Huai	0	1	24	96%

TABLE II

EXPERIMENTAL RESULTS OF RECOGNIZING COMMON PHRASES*

	Phrase						Recognition Rate
	a	b	c	d	e	f	
a	17	0	3	0	0	0	85%
b	0	16	0	0	4	0	80%
c	0	1	16	0	1	0	80%
d	0	0	0	12	8	0	60%
e	0	0	1	2	14	3	70%
f	0	1	1	0	2	16	80%

*Phrases a-f represent: help, blistered, ambulance, thief, fire, and hungry. These phrases are spoken in Chinese in the experiment

extracting techniques. Furthermore, the system can be applied to a "never-lost" system, while not only human, but also any object which is RF-tagged in the house. They can be found through location-aware system and visualized on a TV set, for instance.

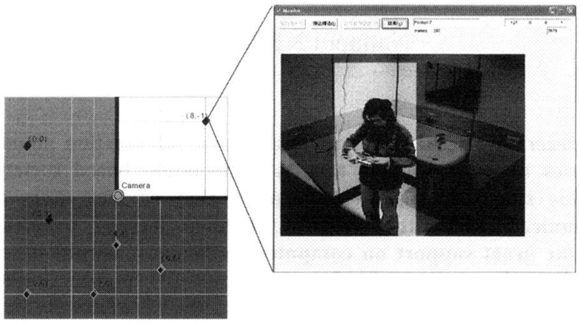

Fig. 8. An example of the "target marking" of the visualization system

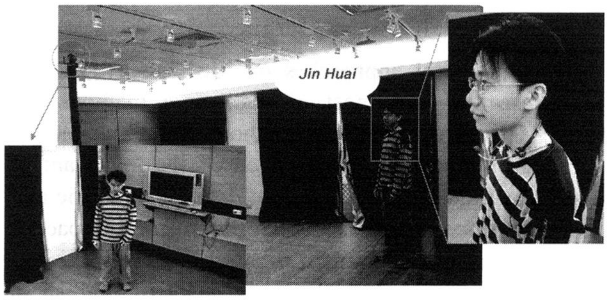

Fig. 9. An example of the "asking-for-help" scene

ACKNOWLEDGMENT

The authors would like to thank Mr. Kun-Ming Yan and Mr. Jin-Huai Hsu for their support in the speech recognition system and experiments.

REFERENCES

[1] " OSA/Parlay Overview," [Online document], 2002 Sep 7, [cited 2006 February 20], Available HTTP: http://www.openapisolutions.com/brochures/OSAParlayOverview.pdf

[2] P. Bahl and V. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," in *Proc. IEEE Infocom 2000*, IEEE CS Press, Los Alamitos,Calif., 2000, pp. 775-784.

[3] P. Castro, P. Chiu, T. Kremenek, and R. Muntz. "A Probabilistic Location Service for Wireless Network Environments," *International Journal of Wireless Information Networks*, Vol. 9, No. 3, pp. 155- 164, 2002.

[4] S. Banerjee, S. Agarwal, K. Kamel, A. Kochut, C. Kommareddy, T. Nadeem, P. Thakkar, B. Trinh, A. M. Youssef, M. Youssef, R. L. Larsen, A. U. Shankar, and A. K. Agrawala, "Rover: Scalable Location-Aware Computing," *IEEE Computer 35*, No. 10, pp. 46–53, 2002.

[5] T. Gu, H.K. Pung, and D.Q. Zhang, "Toward an OSGi-Based Infrastructure for Context-Aware Applications," *IEEE Pervasive Computing*, vol. 03, no. 4, pp. 66-74, 2004.

[6] Y. Chen , X. Y. Chen , F. Y. Rao , X. L. Yu , Y. Li, and D. Liu, "LORE: an infrastructure to support location-aware services," *IBM Journal of Research and Development*, v.48 n.5/6, p.601-615, 2004.

[7] X. Chen, Y. Chen, and F. Rao, "An Efficient Spatial Publish Subscribe System for Intelligent Location-Based Services," in *Proc. the 2nd International Workshop on Distributed Event-Based Systems (DEBS '03)*, June 2003, pp. 1- 6.

[8] A. K. Narayanan, "Realms and States: A Framework for Location-aware Mobile Computing," in *Proc. of the 1st International Workshop on Mobile Commerce*, July 2001, pp. 48–54.

[9] "Open Mobile Alliance (OMA):Mobile Location Protocol (MLP). Candiate Version 3.1- 16 März 2004," [Online document], 2004 March 16,[cited 2006 February 20], Available HTTP: http://www.openmobilealliance.org/release_program/docs/MLP/OMA-LIF-MLP-V3_1-20040316-C.pdf

[10] D.A. Forsyth and J. Ponce, *Computer Vision A Modern Approach*, New Jersey: Prentice Hall, 2003.

[11] C. S. Myers and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition," *The Bell System Technical Journal*, vol. 60, no. 7, pp. 1389-1409, 1981.