社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－UBI－8（74）
2005／6／10

CatchMe: Multi-Camera Person Tracking System for Indoor Public Space

Akinori Komaki[1] **Kohtake Naohiko**[1] **Kazunori Takashio**[12] **Tokuda Hideyuki**[12]

[1]Graduate School of Media and Governance, Keio University

[2]Faculty of Environmental Information, Keio University

{akinori,nao,kaz,hxt}@ht.sfc.keio.ac.jp

We propose CatchMe, a flexible sensing system for human' s location in indoor public space. This system analyzes a human's image by using multiple cameras and detects three-dimensional position and orientation. In this system, users do not have to bring any devices. Unspecified people exist in indoor public space. To provide location-aware services in indoor public space, sensing system that can apply to the movements of the unspecified people is necessary. This paper adapts a method that sets priority according to the accuracy of the image which captures the user's face. This model makes multiple cameras cooperate and detects the user's location. This tracking system can acquire the location information by flexibly enhancing the area.

1 Introduction

As computer hardware advanced, many kinds of sensing devices developed to perceive certain aspects of our environment[20]. As a result, various technologies to acquire environmental information of the real world was developed. Especially, location-sensing technologies are becoming pervasive throughout our surroundings[2]. Actually, as an outdoor location-sensing technology, GPS will be installed in all cellular phones in Japan by 2007[9].

While location-sensing technologies are developing, various location-aware services appear[16][15]. These services inform us about beneficial information based on our locations and movements. To increase the opportunity that we use these services, offered space of these services have expanded. Many researchers are proposing various location-sensing systems from a private space such as home and office to a public space such as department store and street.

Recent years, for the purpose of the crime prevention service and navigation service, there has been increasing interest in tracking system[21][13][11]. Various tracking systems are proposed for many location-aware services now. Unfortunately, most systems have focused on either research that fixes user's position to specific area and analyzes user's movement or research that gives person small sensor, and analyzes behavior. Accordingly, tracking system that can flexibly change area, doesn't give user any sensors and acquires user's orientation has not existed yet.

In this paper, we will attmpt to step back and consider the problem of person-tracking system in a indoor public space. We design and implement CatchMe, as a tracking system in a indoor public space. CatchMe is vision-based tracking system using multiple cameras. To make a tracking area extend flexibly, this system proposes Face Priority Set Model(FPSM). This system can flexibly enhance a tracking area, in comparison with the traditional vision-based systems. This paper presents a sample application using CatchMe and evaluates the comparison with an old research and this system. Final section concludes the paper and remarks future works.

2 Basic Problem of Tracking People in Indoor Public Space

Recently, various researchers are proposing the construction of location-aware services by the progress of the ubiquitous computing environment. These researchers' goal is to create supporting technologies that everyday tasks in our life. For instance, it is a surround application that considers person's position. This application automatically configures position of the sound source according to the listener's position and orientation, and enables sound effects that follows the listener's movement.

However, many existing researches targeted the limited space for a specific person such as members'office. We considered a basic problem of tracking people in indoor public space. Some requirements exist to provide these services in indoor public space. Four function requirements are following: generality, flexibility, accuracy, performance.

2.1 Generality

This space is an environment where unspecified people exist. In this space, the necessity of service intended for unspecified people is high. Tracking system in this space should give the generality that can deal with unspecified people.

To deal with unspecified people, the tracking system reduce a load to which the user carries many sensor devices. By reducing the user's load, each target of the service can be expanded.

2.2 Flexibility

Location-aware service provided in the space uses wide space compared with service as a private room. For instance, considering about an advertising application to follow to the movement of unspecified people. An existing system cannot easily enhance the tracking area because it installs many sensors on the environment and users have to bring any portable devices. According to the use of this space, a system that can flexibly enhance the area where position is acquired is necessary.

2.3 Accuracy

To spread various location-aware services, person's detailed location is requested. For example, there is a service that displays a sight message in user's direction of a glance. In this case, location sensing system should consider the user's orientation and three-dimensional position. Location-aware service can offer information with high quality, by receiving a detailed user' location. To offer a detailed user's locaton, many existing system has limited the tracking are. In this space, the system should not limit the position and make it to a possible tracking in wide space.

2.4 Performance

An important thing is a performance because it operates the system. When thinking about the person's tracking service, A high processing speed is demanded. In this space, many poeple are moving. Follow-me service cannot be provided if there is no high performance. For example, a human's walking speed is called four kilos per hour. To track the person's walking, the system that is treatable of this speed is necessary.

3 Existing Location Sensing Technology

A lot of location sensing systems already exist[1][19][17]. These location sensing systems have different strengthes and weaknesses according to the technology. This paper measures the progress level of those technology.

We compare existing technologies and selects the best technology for the construction of tracking system in the indoor public space. This paper shows the feature of an existing technology(Table 1). This

Table 1. Evaluation of Existing Technology

Technology	item1	item2	item3	item4
supersonic	×	5cm	○	○
RFID	×	15cm	×	×
Wireless LAN	×	300 - 400cm	○	×
Earth Magnetism	×	200 - 300cm	×	○
2D code	×	100cm	○	○
Image Analysis	○	10cm	×	×
Pressure	○	10cm - 100cm	×	×

item1: Form of Use: ○= not portable, ×= portable
item2: Granularity of location information
item3: 3D-position: ○= able, ×= disable
item4: Orientation: ○= able, ×= disable

table shows the achievement level of a tracking system. The item is four(item1∼4) because it chooses a suitable technology for the indoor public space.

As this table shows, the system filled with all requirements has not existed yet. Next, we show the realizability of each technology in the table(Table 2). The item that has already been achieved is "1", the item that can be achieved is "2", the item that cannot be achieved is "3" by now.

Table 2. Realizability of Existing Technology

Technology	item1	item2	item3	item4
supersonic	3	1	1	1
RFID	3	1	3	3
Wireless LAN	3	2	1	3
Earth Magnetism	3	2	3	1
2D code	3	2	1	1
Image Analysis	1	1	2	2
Pressure	1	1	3	3

item1: Form of Use
item2: Location Granularity
item3: 3D-position
item4: Orientation

We paid attention to the image analysis technology as shown in this table. The image analysis is a technology that has progressed rapidly in recent years. Especially, as the necessity for specifying the person by the image is caused in many fields, the research is actively done for the person detection and the face recognition. However, the element of a face still necessary for a specific person has not been clarified yet. It will be thought that the image analysis technology advances from various approaches in the future.

4 Approach of CatchMe

We designed CatchMe to sense the human's position and orientation in indoor public space. We constructed our tracking system to support location-aware service in intelligent environment.

This intelligent environment is designed like a living room, with a couch, a TV, a coffee table and many embedded sensors(mote, smart-its, webcamera etc). The room is shown in Figure 1. The area of this room where the person can move is about 20 square meters. We constructed vision-based system by setting up many cameras in this room.

This paper explains the system, and the assumption environment, potential application and apploach of this system.

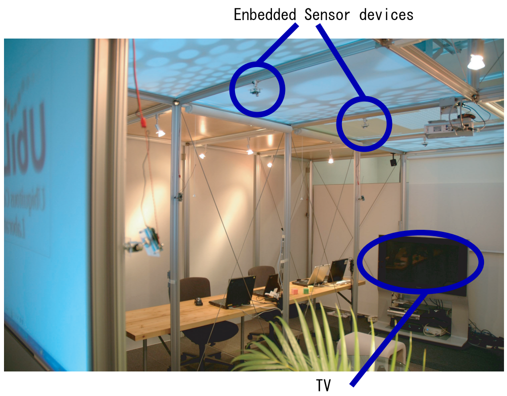

Fig. 1. Intelligent Environment

4.1 Assumption Environment

To provide location-aware services, location information is needed in a actuator and sensor. location sensor is installed in actuator and sensor, and construction of the location management system is assumed. In this system, user's location information is acquired by two or more embedded cameras. Therefore, this paper assumed that a AreaServer which manages location information of many actuators and sensors. The AreaServer always acquires the position and the orientation of actuator and the sensor.

4.2 Potential Application

There are many researches based on location-aware services now. This chapter introduces many services that this system targets. We classify the tracking service for which it is used by the indoor public space.

- **Follow-Me Service**
  Follow-me service offers information with the nearby device according to the person's movement. And triggering events based on location and person's motion, such as the sound effect example(Fig 5) and above. This service provides information matched to each people's behavior. To provide this service, a location granularity and a performance are requested.
- **Security Service**
  Security service observes user's movement and motion, and discovers the prowler or understanding a person's behavior in order to assist him or her. For example, there is a service that navigates the person who has lose his way in the airport. A demand for the prowler discovery service has risen along with the rapid increase of the crime rate in recent years. To provide these services, it is necessary not to give the user the sensor and to acquire person's position widely.
- **Notify Service**
  Notify service does the messaging according to person's position and orientation. The method of doing the messaging becomes abundant, The research of service is advanced in various fields. The demand is high in a indoor public space with a lot of actuators that do the messaging.
- **Device Control Service**:
  Device Control Service can offer various interactions by observing person's movement. According to the action of the indoor person, an intuitive interaction can be offered.

4.3 Approach

CatchMe acquires user's location information according to three steps. First, two or more cameras acquire user's face image. Next, CatchMe analyzes the face image, and sets priority to the image information. Finally, the system selects the image according to priority and calculates the location information.

In this research, we proposes Face Priority Set Model. Catch Me system specifies user's three-dimensional position and orientation according to this model. Figure 2 indicates it by a simple chart. The flow is shown in this chart. Arrows indicate some actions.

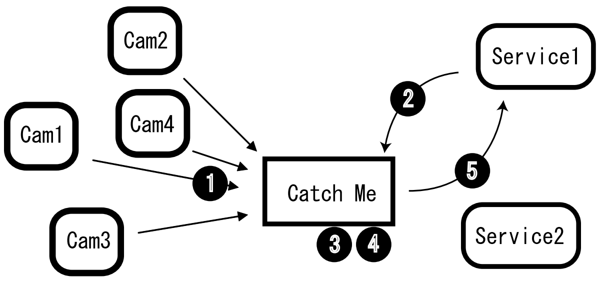

Fig. 2. Behavior of Face Priority Set Model

1. Each of the cameras notify user's location information
2. Each of the services notify the location granularity. For example, if a Follow-Me service is provided, the service requests a location information, about 30cm grain degree.

3. According to the service's request, face priority are listed. High priority is put on the camera to be able to acquire the user accurately.

4. A location information is calculated by using the camera with high priority.

5. Then, the location information is returned. This flow is repeated as follows at each service.

Key point of this model is how to set the priority. A basic way to put priority is accuracy of the face image. High priority is set to the image that it is the largest and the error is little. When a service demands a in-depth location granularity, a camera image that roughens more than the demand is deleted.

Many accident errors are caused in the face recognition by the image analysis in the place where a lot of backlights and obstacles exist. The accident error of the location information can be reduced by using this model.

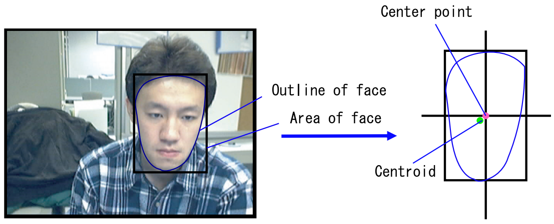

Fig. 3. face detect

5 Design & Implementation

In this section, Hardware composition and software architecture of CatchMe.

CatchMe uses multiple cameras, each connected to its own PC on Linux(2.6.4). CatchMe uses web cameras(300,000 pixels, viewing angles 44 degrees). Area server manages the location information of those cameras and many actuators. And, CatchMe Server that calculates a user'position and orientation exists. Each application program receives information with UDP. In this implementation, the cameras are connected on a local network, and we excluded the delay of the system as much as possible. Next, the outline of software is described.

This software's implementation used MALib[8] to detecte the face. The image processing technology improves, and various image processing libraries are developed in recent years. Now, a following library is represented as open source, OpenCV, ARToolKit, Malib. OpenCV[3] is image processing API that intel offers. This API is offering the library optimized to cpu made of intel. ARToolKit[10] is a library to make an application of AR(Augmented Reality). ARToolKit acquires a location of a marker by taking a picture of a marker. And, there is MALlib as image processing API for Linux. The feature of Malib is face detection API.

CatchMe detects the face area by using MALib, and calculates centroid and a center point from the face area(Figure 3). This centriod point is adjusted to the position of the face. And orientation of a face is detected from a difference between this centriod point and the center point.

An existing system that can acquire the user orientation was not able to part near the camera. In this system, a user orientation of the face about 5m away from the web camera can be detected by using this technique. However, this changes according to the performance of the camera and environment.

Priority is put on the image taken with each web camera(Figure 4). The system calculates a three-dimensional position by using two face images with high priority. The calculation of a three-dimensional position uses the triangulation. It is assumed that a series of camera parameter(viewing angle, number of pixels, and location information of camera) is already-known. Therefore, this system can calculate the angle per pixel.

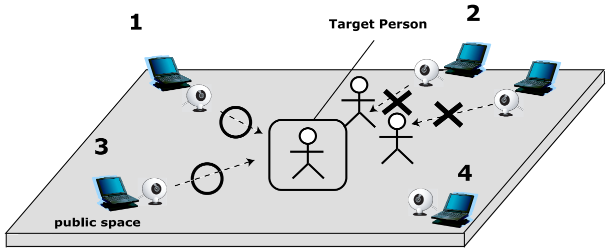

Fig. 4. tracking target

6 Evaluation

This paper implement CatchMe, a person-tracking system indoor public space as a prototype. This paper does a qualitative evaluation and a quantitative evaluation. A qualitative evaluation is compared with a existing system using image analysis. The quantitative evaluation is done about the processing time to calculate the location information and location granularity. The place of the evaluation used smart space in Keio University. Four cameras were set up in the space of 20 square meters. To avoid direct sunlight, the room put a blackout curtain. The evaluation result of the location granularity was as follows(Table 3).

As a result of showing here, it changes depending on the number of cameras. However, it is understood to be able to acquire person's location information accurately.

Table 3. location granularity

System Name	CatchMe
Location	25cm
Orientation	30 °
Number of Camera	4

We mesured the time spent on the image analysis. This paper evaluate the time that CatchMe calculated. The following evaluations were obtained in the position and the orientation calculation. The speed took the average of 100 operation. To measure the image processing speed, the measurement used the read-time stamp counter(RDTSC). The following are the results(Table 4).

Table 4. processing speed

system name	**location**	**orientation**
CatchMe	150.56ms	200.01ms

The processing time was made a minimum in not calculating the characteristic of the face in this research. The target of this research is to keep the performance to be able to follow to the person's movement. As a result, this system can provide those services.

We compared it with the existing system like to this system. A qualitative evaluation evaluated the following, Generality, Flexibility, Accuracy. The qualitative evaluation is as follows(Table 5).

Table 5. Qualitative evaluation

system	**Generality**	**Flexibility**	**Accuracy**
CatchMe	○	○	○
Aware Home	△	△	◎
Easy Living	△	△	○
Activity Zone	○	△	◎

evaluation： ◎=very good, ○=good, △=bad

As for CatchMe, it is understood that there are a lot of advantages in the indoor public space compared with other image analysis systems.

7 Discussion & Conclusion

As a demonstration of our system, we present TrackableSound(Figure 5). TrackableSound automatically configures position of the sound source according to the listener's position and orientation, and enables sound effects that follows the listener's movement. TrackableSound is one of the Follow-Me services. In this experiment, it was a performance enough for the user's follow.

CatchMe is a tracking system intended for indoor public space. There are some problems to offer CatchMe by the indoor public space. CatchMe was made by existing image processing API. MALib detected user's face from color information in the image. There is sometimes a mistake because only color spectrum is using it. It is necessary to improve the alogrithm of the face detection to decrease a judgment mistake of the image analysis.

CatchMe calculates rough, because the location is specified based on priority. To use the image analysis, a few errors are caused. It is necessary to improve the location granurality to make an interface of the eye. The location granurality is improved by increasing the number of cameras used.

There is privacy as a problem that should be thought in the future. Privacy is one of the hottest topics now. Many poeple meet with strong opposition setting camera in indoor public space, because the camera senses many information. However, there is a coping strategy by limiting limitation of installation area of camera and filtering image. There are research topics because it constructs a tracking system in indoor public space.

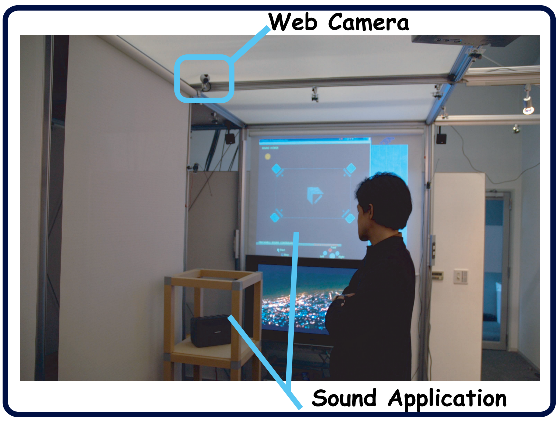

Fig. 5. TrackableSound

8 Related Works

The area of vision-vased tracking is very active with work in face tracking, gesture understanding, body-part tracking. We concentrate here on work in person's location tracking system. The work most closely related to ours is that of Haritaoglu and Daviswith a series of outdoor person-trakers named W^4[5], W^4S[7], and Hydra[6]. Their systems include grayscale appearance modelingthe ability to add and delete people over timeand the ability to track people whose silhouettes overlap

Another related work is that of Darrell et al- who use face detectionIn the part of the DARPA VSAM project[4], CMU has created an elaborate system for video-based surveillace. Using multiple pan/tilt/zoom cameras, their system classifieds and tracks multiple people and vehicles as they move about outdoor. Rosales and Sclaroff describe a multi-person tracking system that unifies object tracking, 3D trajectory estimation, and action recognition from a single video camera[18].

Recently, There is a Easy Living project[14] by John Krumm, Steve Harris. It uses two set of color stereo cameras for tracking multiple people, during live demonstrations in a living room. The stereo images are used for locating people, and the color images are used for maintaining their identities. As a research that makes the best use of the characteristic of the camera, there is a Activity Zones by Kimberle Koile, Konrad Tollmar[13]. It showed how to semi-automatically partition a space into activity zones based on patterns of observed user location and motion.

Rick Kjeldsen and Jacob Hartman researched the design issue ofthe vision-based computer interaction[12]. In this research, a method of designing vision-based system is dscribed from the aspect in interface.

Acknowledgments

This research has been conducted as part of Ubila Project supported by Japanese Ministry of Internal Affairs and Communications. And we thank all numbers of Tokuda Lab who wrote and provided helpful comments on previous versions of this document.

References

1. N. Adams, R. Gold, B. N. Schilit, M. M. Tso, and R. Want. An infrared network for mobile computers. In *Proc. of the USENIX Mobile and Location-Independent Computing Symposium*, pages 41–51, Cambridge, MA, 1993.
2. Jeffrey Hightower Borriello and Gaetano. *Location Systems Ubiquitous Computing.* August 2005.
3. Gary Bradski. The OpenCV library. 25(11):120, 122–125, November 2000.
4. Robert Collins, Alan Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, and Osamu Hasegawa. A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2000.
5. Ismail Haritaoglu, David Harwood, and Larry S. Davis. W4: Who? when? where? what? a real time system for detecting and tracking people. In *FG*, pages 222–227, 1998.
6. Ismail Haritaoglu, David Harwood, and Larry S. Davis. Hydra: Multiple people detection and tracking using silhouettes. In *ICIAP*, pages 280–285, 1999.
7. D. Harwood I. Haritauglu and L. S. Davis. W4s: A realtime system for detecting and tracking people in 2 1/2 d. In *presented at Fifth European demonstration, people enter and leave the living room, Conference on Computer Vision, Freiburg, Germany*, 1998.
8. Jun Iio. Malib: Media analysis library. *Linux Conference*, 2002.
9. Japanese and Ministry of Internal Affairs and Communications. Study committee concerning expansion of use of frequency for cellular phone. Technical report, 2005.
10. Hirokazu Kato and Mark Billinghurst. Developing ar applications with artoolkit. In *ISMAR*, page 305, 2004.
11. Cory D. Kidd, Robert Orr, Gregory D. Abowd, Christopher G. Atkeson, Irfan A. Essa, Blair MacIntyre, Elizabeth D. Mynatt, Thad Starner, and Wendy Newstetter. The aware home: A living laboratory for ubiquitous computing research. In *CoBuild*, pages 191–198, 1999.
12. Rick Kjeldsen and Jacob Hartman. Design issues for vision-based computer interaction systems. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–8, New York, NY, USA, 2001. ACM Press.
13. Kimberle Koile, Konrad Tollmar, David Demirdjian, Howard E. Shrobe, and Trevor Darrell. Activity zones for context-aware computing. In *Ubicomp*, pages 90–106, 2003.
14. John Krumm, Steve Harris, Brian Meyers, Barry Brumitt, Michael Hale, and Steve Shafer. Multi-camera multi-person tracking for easyliving. In *VS '00: Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000)*, page 3, Washington, DC, USA, 2000. IEEE Computer Society.
15. Chandra Narayanaswami and M. T. Raghunath. Application design for a smart watch with a high resolution display. In *ISWC '00: Proceedings of the 4th IEEE International Symposium on Wearable Computers*, page 7, Washington, DC, USA, 2000. IEEE Computer Society.
16. Ikuko Shimizu Okatani and Nishimura Takuichi. Location-based information support system using multiple cameras and led light sources with the compact battery-less information terminal (cobit). In *ECCV Workshop on HCI*, pages 142–152, 2004.
17. Jun Rekimoto and Katashi Nagao. The world through the computer: Computer augmented interaction with real world environments. In *ACM Symposium on User Interface Software and Technology*, pages 29–36, 1995.
18. Rómer Rosales and Stan Sclaroff. 3d trajectory recovery for tracking multiple objects and trajectory guided recognition of actions. In *CVPR*, pages 2117–2123, 1999.
19. Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. Technical Report 97.10, AT&T Laboratories Cambridge, 24a Trumpington Street, Cambridge CB2 1QA, England, 1997. Also in IEEE Personal Communications, Vol. 4, No. 5, October 1997, pp 42-47.
20. Mark Weiser. The computer for the 21st century. *Scientific American (International Edition)*, 265, No. 3:66–75, 1991.
21. Kevin Wilson, Vibhav Rangarajan, Neal Checka, and Trevor Darrell. Audiovisual arrays for untethered spoken interfaces. In *ICMI*, pages 389–394, 2002.