Development of a Ubiquitous Vision System for Location-awareness of Multiple Targets by a Matching Technique for the Identity of a Target: a New Approach

Chi-Ho Kim*,**, Bum-Jae You*, and Hagbae Kim**

* Intelligent Robotics Research Center, Korea Institute of Science and Technology, Seoul, Korea
C.Kim - (Tel : +82-2-958-5708; E-mail: chkim@yonse.ac.kr)
B.You - (Tel : +82-2-958-5760; E-mail: ybj@kist.re.kr)
**Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
H.Kim - (Tel : +82-2-2123-2778; E-mail: hbkim@yonsei.ac.kr)

Abstract: Various techniques have been proposed for detection and tracking of targets in order to develop a real-world computer vision system, e.g., visual surveillance systems, intelligent transport systems (ITSs), and so forth. Especially, the idea of distributed vision system is required to realize these techniques in a wide-spread area.

In this paper, we develop a ubiquitous vision system for location-awareness of multiple targets. Here, each vision sensor that the system is composed of can perform exact segmentation for a target by color and motion information, and visual tracking for multiple targets in real-time. We construct the ubiquitous vision system as the multiagent system by regarding each vision sensor as the agent (the vision agent). Therefore, we solve matching problem for the identity of a target as handover by protocol-based approach. We propose the identified contract net (ICN) protocol for the approach. The ICN protocol not only is independent of the number of vision agents but also doesn't need calibration between vision agents. Therefore, the ICN protocol raises speed, scalability, and modularity of the system. We adapt the ICN protocol in our ubiquitous vision system that we construct in order to make an experiment. Our ubiquitous vision system shows us reliable results and the ICN protocol is successfully operated through several experiments.

Keywords: Identified Contract Net (ICN) protocol, location-awareness for multiple targets, ubiquitous vision system, multiagent system, handover.

1. INTRODUCTION

Various techniques have been developed for detection and tracking of targets in order to develop real-world computer vision system, e.g., visual surveillance systems [1] and intelligent transport systems (ITSs). Especially, an idea of distributed vision system is required to realize these techniques in a wide-spread area. Distributed vision system can be used to help an agent for which it is difficult to notice one's location by oneself in complicated environments, and as a platform to access to physical world and virtual world. The navigation of a mobile robot and the virtual reality are representative examples, respectively.

Recently, researches for a location-awareness system that surveils a wide-spread area are being increased. Each vision sensor that constitutes the system is made up of a camera and an image processor, and connected by computer network each other. The system can monitor a wide-spread area in real-time because an image that is acquired from a camera is processed in each vision sensor. Redundant information can be occurred because many vision agents exist in the distributed vision system. This can contribute more reliability to the system [2]. However, if there is no coordination for redundant information, it cannot get proper results and can result in overloading in the network [3]. [4] and [5] use multiple camera approaches. Because their cameras are densely arranged with a short baseline, the scene where they are looking at is restrained to a small area. Therefore, their methods are not readily applied to our goal (location-awareness in a large area). Ng el al. [6] proposed a method that synthesizes a person simultaneously by using many omni-directional vision sensors. Nakazawa et al. [7] proposed a method that uses a state transition map and an action rule as tracking a person in a distributed vision system. At first, they divide tasks that each vision sensor can perform into three states, that is, tracking, acquisition, and idling. Next, they determine which task the vision sensor performs by using a state transition map, made by information for visible area, and an action rule. [6] and [7] successfully coordinate several vision sensors but consider tracking for only one person. Nakazawa et al. [8] improved their system to be able to track multiple persons. A matching technology is needed for each person to be able to track multiple persons in a distributed vision system. They called all targets in a distributed vision system the seeing agents and solved matching problem by gathering the seeing agents that had a same identity (they called the group of the seeing agents 'the agency'). Matsuyama and Ukita [9] propose a method that tracks multiple targets by the cooperation between vision sensors, regarding distributed vision system as a multiagent system. However, the methods that is used in [6], [7], [8], and [9] need calibration in coordinating between vision agents. This results in a barrier in raising the speed, modularity, and scalability of a distributed vision system.

In this paper, we develop a ubiquitous vision system for location-awareness of multiple targets in a large area. Here, each vision sensor that the system is composed of can perform exact segmentation for a target by color and motion information, and visual tracking for multiple targets in real-time. We construct the ubiquitous vision system as the multiagent system by regarding each vision sensor as the agent (the vision agent). Therefore, we solve matching problem for the identity of a target as handover by protocol-based approach. We propose the identified contract net (ICN) protocol for the approach. The ICN protocol is independent of the number of vision agents and the kind of a camera. And, if we use the ICN protocol, calibration is not needed and an overlapping area between neighborhood vision agents does not have to exist. This results in raising the speed, scalability, and modularity of the system.

Section 2 introduces elementary techniques that constitute each vision agent, that is, a target segmentation algorithm and an approach for real-time visual tracking of multiple targets. We represent architecture of the ubiquitous vision system constructed by us, and propose the ICN protocol that operates our ubiquitous vision system in section 3. We construct our ubiquitous vision system by setting up 12 vision agents to make experiments and estimate performance in section 4. This paper is concluded in section 5.

2. ELEMENTARY TECHNIQUES OF EACH VISION AGENT

Each vision agent must be able to track targets that enter its visible area. Because the number of target is not determined, it must be also able to track multiple targets. Therefore, our vision agents need to have algorithms for target extraction and multiple targets tracking.

2.1 Target extraction

A target is extracted by combining (AND gate) motion and color information. We use a deference image between a current image and a reference image with color information for motion segmentation. And, the approach proposed in [10] is adopted for color segmentation. The algorithm is based on the fact that a color has different statistical characteristics at different intensity level. The statistical characteristics of a color are represented by the means and standard deviations of hue and saturation with respect to intensity. After a number of image patches in different illumination conditions are stored, a color model is generated by 2nd order B-spline curve fitting for statistical characteristics of hue and saturation with respect to intensity.

2.2 Real-time visual tracking of multiple targets [11]

Our algorithm for real-time visual tracking of multiple targets is made up of following two modes.:

1. *Detection*: to surveil and detect new targets that enter a visible area
2. *Tracking*: to track targets that was tracked in previous frame

We use the discriminative focus of attention for the detection mode and line-based trackers for the tracking mode.

2.2.1 Discriminative focus of attention

We modify a concept of the focus of attention introduced in [12] to reduce searching time of new targets. It is very similar with human eyes in the sense that imaging resolution of outer area of human eyes is lower than that of fovea area. That is, targets that are being tracked are processed with high resolution and low resolution is used to surveil new targets. We call it the discriminative focus of attention. We divide the focus of attention into the following two levels:

- Level 1: This process is to track targets that were tracked in a previous frame. This level is executed with a high resolution.
- Level 2: This process is to find new tracks in a current frame. This level is process with a low resolution.

We use 10×10 super pixel in the level 2. If a cue for a new target is detected in the level 2, the target is tracked with higher resolution from the initial location where it is detected. The discriminative focus of attention contributes to high-speed visual tracking by decreasing the resolution of backgrounds.

2.2.2 Line-based tracker

We perform real-time visual tracking of a target by adjusting four line-based trackers (top line tracker, bottom line tracker, left line tracker, and right line tracker). Figure 1 shows how four line-based trackers find the location of a target (face). At first, a line-based tracker is moved with fixed distance to the outer direction of a target. If the number of interest pixels on the line-based tracker is equal to zero, it is moved with fixed distance to the opposite (inner) direction. Until the number of interest pixels on the line-based tracker is not equal to zero, it is moved to the inner direction repeatedly. If the number of interest pixels on the line-based tracker is not equal to zero at the first step, it is moved with fixed distance to the same (outer) direction. And, the line-based tracker is moved repeatedly until the number of interest pixels on the line-based tracker becomes zero. The locations of other three trackers are determined by the same method. These four line-based trackers provide a region for each target. Especially, if a top line tracker is lower than the corresponding bottom tracker or a left line tracker is more right than the corresponding right line tracker, this system regards that the target is disappeared is its visible area.

Fig. 1 Line-based trackers for face tracking

The line-based tracker can not only realize faster visual tracking than region-based approaches but also solve the motion correspondence problem.

3. UBIQUITOUS VISION SYSTEM

We develop a ubiquitous vision system that is made up of independent vision agents. Here, the techniques explained in section 2 are loaded in each vision agent. The aim of this system is aware of location of targets in a large area from location information of those that each vision agent computes.

3.1 Architecture of our ubiquitous vision system

Our ubiquitous vision system is made up of vision agents (VAs) that consist of a camera and an image processor. All vision agents and a directory vision agent (DVA) for interface with users are connected by wireless computer network each other. Figure 2 shows the architecture of our ubiquitous vision system.

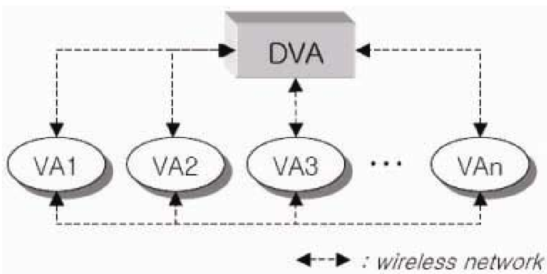

Fig. 2 Architecture of our ubiquitous vision system

3.2 Identified contract net (ICN) protocol

Many researches have tried to develop a ubiquitous system by using vision sensors. The ubiquitous vision system must solve matching problem for the identity of a target as handover. Most researches solved this problem by relation between neighborhood vision sensors. We call it calibration-based approach. When we use a calibration-based approach to solve matching problem for the identity of a target, we notice the following problems:

1. It will decrease the speed of a ubiquitous vision system. When a target exists in an overlapped region that several vision sensors cover together, the system is more loaded relatively by including all vision sensors in the computation process. Therefore, we do not use the merit of distributed process.
2. It will hinder the scalability of a ubiquitous vision system. Whenever we set up an additional vision sensor to extend applicable area, we must perform calibration with neighborhood vision sensors and fix the location and angle (pan, tilt) of the vision sensor that is once set up. Therefore, it stiffens extension of a system.
3. It will hinder the modularity of each vision sensor. Whenever we set up each vision sensor, we must consider relation with its neighborhood vision sensors. Therefore, it prevents a vision sensor from being independently developed.

We solve matching problem for the identity of a target as handover without a calibration-based approach to prevent the above problems. We construct the ubiquitous vision system as the multiagent system by regarding each vision sensor as the agent. That is, we solve the matching problem by interaction between agents. We propose and use a communication protocol between vision agents for the interaction. Therefore, we solve matching problem for the identity of a target as handover without calibration-based approach, but with a protocol-based approach. We propose the ICN protocol that can be used for multiple targets as well as a single target. The basic steps in the ICN protocol are the followings:

1. ***Request Propagation*** (RP, Figure 3) - If an agent that is seeing a target needs cooperation with other agents, that is handover, the agent informs the other agents of a message that includes the identity of the target and a task.

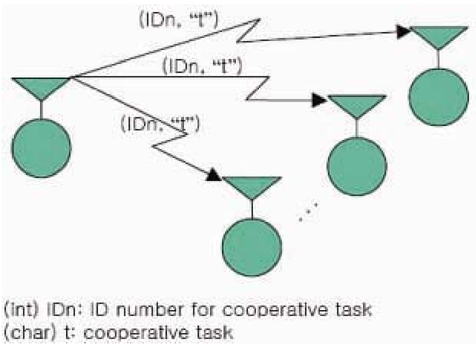

Fig. 3 Request propagation

2. ***Bid Feedback*** (BF**,** Figure 4) - The agents that get the message evaluate it. The agents for which it is possible to cooperate submit bids to the agent that sends the message.

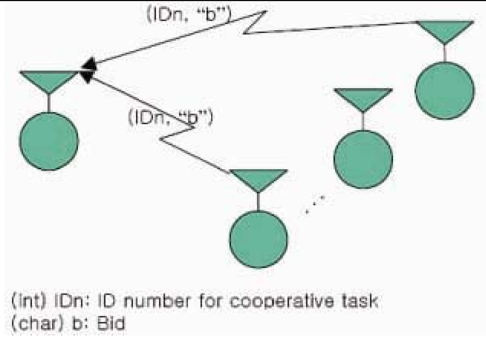

Fig. 4 Bid feedback

3. ***Contract Propagation*** (CP**,** Figure 5) - The agent that get the bids evaluates them. The agent awards a contract to an agent that sends the most appropriate bid.

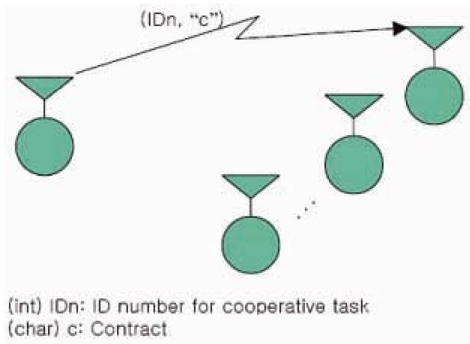

Fig. 5 Contract propagation

4. ***Acknowledgement Feedback*** (AF**,** Figure 6) – The awarded agent sends an acknowledgement to a directory vision agent. At this time, the directory vision agent perceives handover for the target.

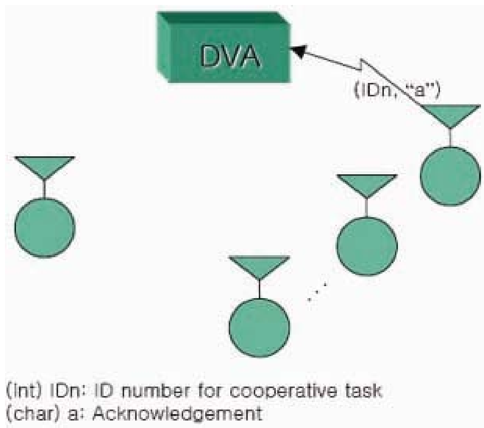

Fig. 6 Acknowledgement feedback

If a target tracked in a vision agent escapes its visible area, the target needs handover and interaction between vision agents is happened from that time. The target tries to escape its visible area in this process, interaction between vision agents starts by the ICN protocol (ICN initiation). At that time, the vision agent that is tracking the target is assigned to master agent and its neighborhood agents become slave agent. In the case of our ubiquitous vision system, the ICN protocol is initiated when the target touches the edge of the image that the vision agent acquires. Because a target disappears in a field of vision through its edge such as human eyes. If the target does not disappear in the field of vision and separates from its edge, the ICN protocol is released. At first, after roles of all agents involved in a handover process are assigned to master or slave, the master agent sends 'request propagation' signal to the slave agents. The slave agents check which an unknown target exists or not in their visible area. If an unknown target exists, the slave agent submits its world coordinate as bid, 'bid feedback' signal, to the master agent. The master agent analyzes bids. That is, it computes difference between a world coordinate that it computes and each bid. The master agent sends 'contract propagation' signal to a slave agent that has the smallest result, difference. If the master agent gets no 'bid feedback' signal, that is, no vision agent detects the target, it repeats the ICN initiation process until it receives at least one 'bid feedback' signal. The slave agent receiving the 'contract propagation' signal sends an 'acknowledgement feedback' signal to a directory vision agent. The 'acknowledgement feedback' signal means that handover for the identity of the target to be accomplished to the agent sending the signal. The slave agent with which the master agent makes a contract tracks the target from this time and the result is transmitted to the directory vision agent. This contract means that a system hands over a target, that an agent tracks, to other agent. We can use the ICN protocol to track multiple targets by indicating the identity of a target in a message header. The overview of the ICN protocol is shown in Figure 7.

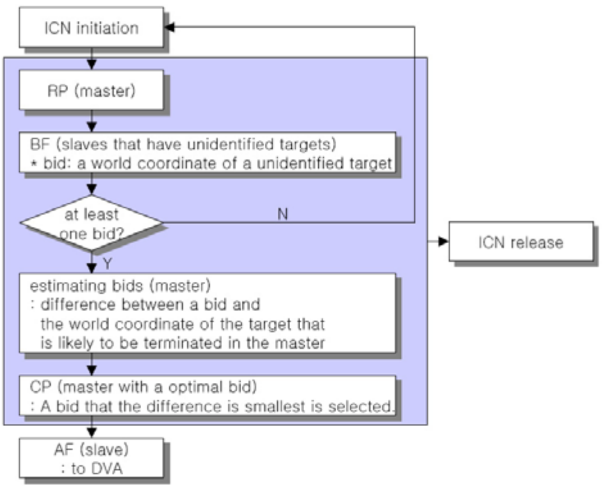

Fig. 7 The overview of the ICN protocol

4. EXPERIMENTAL RESULTS

At first, we implement the elementary techniques of vision agents that constitute our ubiquitous vision system. The result of the target extraction algorithm is shown in figure 8. Figure 8 shows a result of extraction for skin color. A target is finally extracted by combining (AND gate) motion and color information. From figure 8, it is verified that our algorithm successfully extracts a target (skin color).

We use a yellow green ID card in this experiment. That is, we make a color model for yellow green and perform tracking and location-awareness based on the color model. The yellow green color model made by us is shown in figure 9. Here, average of hue, average of saturation, standard deviation of saturation, and standard deviation of hue with respect to intensity are represented to the clockwise rotation from the left-top graph.

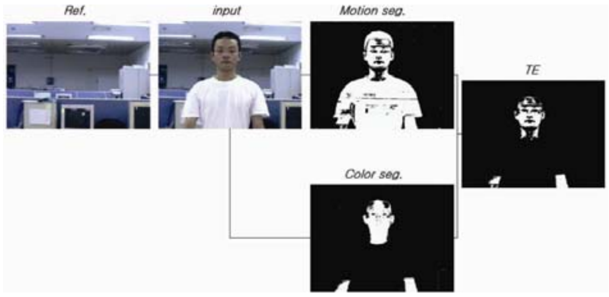

Fig. 8 Skin area extraction

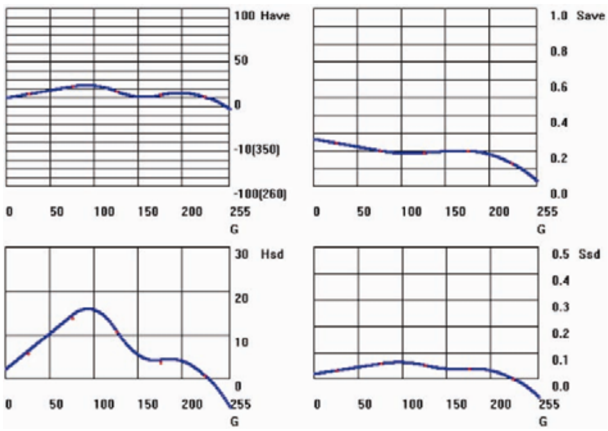

Fig. 9 Color model for yellow green

We develop a visual tracking algorithm for multiple targets based on this color model. Figure 10 shows a result tracking skin color by using our visual tracking algorithm for multiple targets (face and hands).

Table 1 shows frame rates with respect to number of targets, using our tracking algorithm. From table 1, it is verified that our visual tracking algorithm for multiple targets has real-time property. Therefore, our ubiquitous vision system can also be aware of location in real-time.

We make an experiment for a ubiquitous vision system by using the elementary techniques. We constitute our ubiquitous vision system in 6m × 4.5m space. A vision agent is made up of a color camera and an image processor. We constitute three vision agents with three IMI-TECH IMC-80F IEEE 1394 color camera and an Intel Pentium M Processor 1.6GHz laptop computer. That is, one laptop computer is connected with three cameras. We use an Intel Pentium 4 3.4GHz CPU PC as the directory vision agent. We arrange 12 vision agents and one directory vision agent as shown in figure 11.

We make a ubiquitous vision system tracking yellow green IDs card in the environment of figure 11. Therefore, the goal of the experiment is aware of location of persons that wear the ID card. We use a simple method of the perspective projection in order to estimate location (world coordinate) of a target in each vision sensor. The results of location-awareness for our ubiquitous vision system are shown in figure 12 and 13. They show a result of location-awareness for one target and for two (multiple) targets, respectively. Here, figure 13-(a) is a result of visual tracking for multiple targets when two targets exist at once in a visible area of a vision agent. And, figure 13-(b) is a result when two targets exist in different vision agent. Results for speed and accuracy of location-awareness are shown in table 2. The speed means the time the system spends for one location-awareness. We get average 0.190913 second. It is noted that our ubiquitous vision system can be aware of location in real-time. Accuracy of location-awareness means deviation from the traffic line of a person that is already determined. We get average 3.671209cm. Considering uncertainty of a person's movement, this result is very accurate. Therefore, It is noted that location-awareness for our ubiquitous vision system has very high accuracy.

(a) The number of targets is 1.

(b) The number of targets is 2.

(c) The number of targets is 3.

(d) A new target enters the visible area.

(e) A tracked target disappears in the visible area.

Fig. 10 Multiple targets tracking

5. CONCLUSION

Computer vision systems have been used in many application fields. An idea of distributed vision system is required in order to realize these systems in a wide-spread area. In this paper, we have implemented a distributed vision system that can be aware for location for multiple targets in a wide-spread area environment by using a vision-based sensor network. The developed system is built upon the concept of ubiquitous and especially uses vision sensor. We construct our ubiquitous vision system as the multiagent system by regarding each vision sensor as the agent. Here, each vision agent is made up of a camera and an image processor. We have developed the system based on two key ideas: (1) exact segmentation and real-time visual tracking of multiple targets and (2) handover protocol (ICN protocol) for interaction between vision agents. Our ubiquitous vision system needs a rule for handover when a target goes and comes through an overlapped region that several vision sensors cover together. Most distributed vision system solved matching problem for the identity of a target as handover by calibration-based approach. However, we solved the problem by protocol-based approach. That is, we proposed the ICN protocol to solve the problem for multiple targets. This protocol is not only independent of the number of vision agents but also don't need calibration between vision agents. Therefore, this approach raises speed, scalability, and modularity of the system. We used a target segmentation algorithm and an approach for real-time visual tracking of multiple targets by line-based tracker as elementary techniques that constitute each vision agent. Each vision agent can segment an accurate location of a target and track multiple targets in real-time by using these elementary techniques. Therefore, our ubiquitous vision system can accurately be aware of location in real-time. The proposed protocol is successfully performed in a vision-based sensor network that we construct for experiments. From several experiments, we verify that our ubiquitous vision system can be aware of location with high accuracy in real-time.

Table 1 Frame rate for multiple targets tracking

Number of target (s)	Frame rate (Hz)
0	14.99
1	14.85
2	10.26
3	10.05

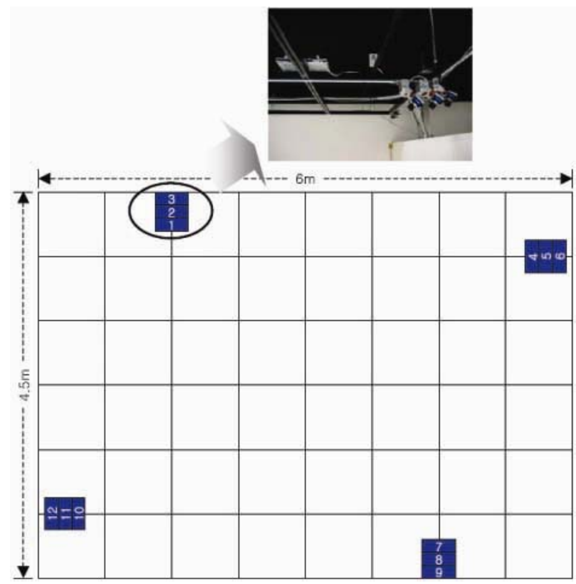

Fig. 11 Experimental set-up for our ubiquitous vision system

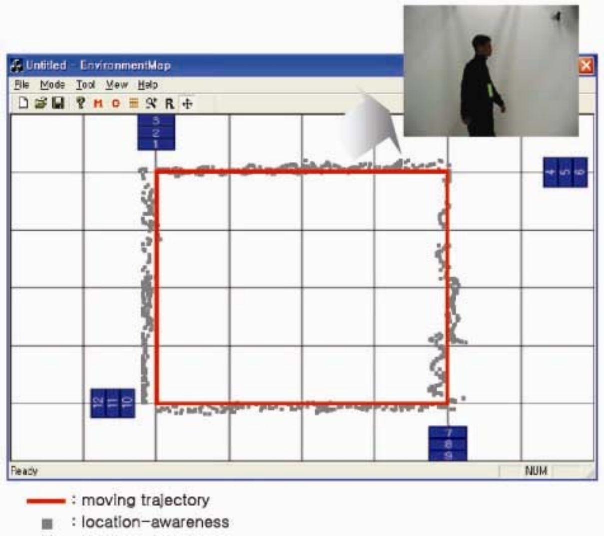

Fig. 12 Location-awareness for one person

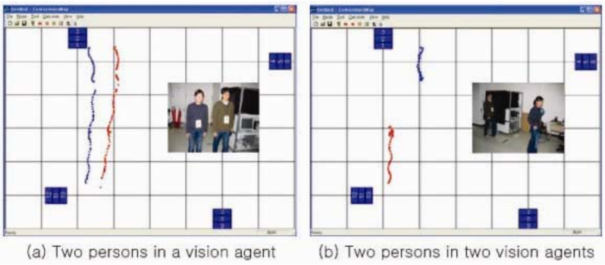

Fig. 13 Location-awareness for two persons

Ubiquitous age is coming soon. This paper verifies that a ubiquitous system can be implemented by using vision-based sensors. Many issues for the ubiquitous vision system remain and they depend on an application. Location-awareness can be viewed as one example. However, two key issues proposed in this paper can be adapted for any application. This is a primary contribution of this paper.

REFERENCES

[1] http://www-2.cs.cmu.edu/~vsam/.

[2] B. S. Rao and H. Durrant-Whyte, "A Decentralized Bayesian Algorithm for Identification of Tracked Targets," *IEEE Trans. on System, Man, and Cybernetics*, vol. 23, no. 6, Nov. 1993.

[3] P. K. Varshney and I. L. Coman, "Distributed Multi-sensor Surveillance: Issues and Recent Advances," *in Video-Based Surveillance Systems: Computer Vision and Distributed Processing*, P. Remagnino, G. A. Jones, N. Paragios, and C. S. Regazzoni: Kluwer Acardemic Publichers, 2002, pp. 239-250.

[4] P. J. Narayanan, P. W. Rander and T. Kanade, “Constructing Virtual World using Dense Stereo,” *Proc. of IEEE International Conference on Computer Vision*, pp. 3-10, 1998.

[5] S. M. Seitz and K. N. Kutulakos, “Plenoptic image editing,” *Proc. of IEEE International Conference on Computer Vision*, pp. 17-24, 1998.

[6] K. C. Ng, H. Ishiguro, M. Trivedi, and T. Sogo, "Monitoring Dynamically Changing Environments by Ubiquitous Vision System," *2nd IEEE Workshop on Visual Surveillance*, pp.67-73, 26 June 1999.

[7] A. Nakazawa, H. Kato, and S. Inokuchi, "Human Tracking using Distributed Vision Systems," *Proc. of 14th International Conference on Pattern Recognition*, vol. 1, pp. 593-596, 16-20 Aug. 1998.

[8] A. Nakazawa, H. Kato, S. Hiura, and S. Inokuchi, "Tracking Multiple People using Distributed Vision Systems," *Proc. Of International Conference on Robotics and Automation*, vol. 3, pp.2974-2981, 11-15 May 2002.

[9] T. Matsuyama and N. Ukita, "Real-time Multitarget Tracking by a Cooperative Distributed Vision System," *Proc. of the IEEE*, vol. 90, pp.1136-1150, July 2002.

[10] Y. B. Lee, B. J. You, and S. W. Lee, "A Real-time Color-based Tracking robust to Irregular Illumination Variations," *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1659-1664, 2001.

[11] C. H. Kim, B. J. You, and H. Kim, “Detection and Real-time Tracking of Moving Targets using a Color Segmentation Algorithm robust to Irregular Illumination Variation and a Line-based Tracker,” *The Institute of Electronics, Information and Communication Engineers Trans. Commun.*, to be published.

[12] J. K. Tsotsos, "Towards a Computational Model of Visual Attention," *in Early Vision and Beyond*, T. Papathomas, C. Chubb, A. Gorea, and E. Kowler: MIT Press, 1995.

[13] G. Weiss, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, MIT Press, 1999.

Table 2 Results for speed and accuracy of our ubiquitous vision system

Trial	Speed (sec./pt.)	Deviation (cm)
1	0.178806	2.750000
2	0.190061	4.918367
3	0.188227	6.068182
4	0.188596	3.280702
5	0.187057	6.028572
6	0.186588	8.294118
7	0.178326	5.239130
8	0.185378	8.135135
9	0.192904	8.346154
10	0.184000	2.375000
11	0.213120	1.800000
12	0.211574	1.803279
13	0.202289	1.763158
14	0.180977	2.930233
15	0.181775	1.714286
16	0.173105	1.131579
17	0.179867	2.133333
18	0.201656	2.062500
19	0.201800	1.228571
20	0.212156	1.421875
Average	0.190913	3.671209