A Study on the Application of Spatial-Knowledge-Tags using Human Motion in Intelligent Space

Tae-Seok Jin*, Kazuyuki Morioka**, Mihoko Niitsuma*, Takeshi Sasaki*, and Hideki Hashimoto*

* Institute of Industrial Science, the University of Tokyo, Tokyo, Japan
(Tel : +81-3-5452-6258; E-mail: {jints, niitsuma, sasaki, hashimoto}@hlab.iis.u-tokyo.ac.jp)
**Department of Electrical Engineering, Tokyo University of Science, Chiba, Japan
(E-mail: morioka@itohws01.ee.noda.tus.ac.jp)

Abstract: Intelligent Space (iSpace) is the space where many intelligent devices, such as computers and sensors, are distributed. According to the cooperation of many intelligent devices, the environment comes to have intelligence. In iSpace, the locations of multiple humans and other objects are obtained and tracked by using multiple camera and color-based method. In addition, we describe a context-aware information system which is based on Spatial-Knowledge-Tags (SKT). SKT system enables humans to access information and data by using spatial location of human and stored information in storage. The proposed tracking method is applied to the intelligent environment and its performance is verified by the experiments.

Keywords: Multi-Vision, Intelligent Space, Mobile Robots, Spatial-Knowledge-Tags.

1. INTRODUCTION

In recent years, the research field on the intelligent environment has been expanding[1][2]. An intelligent Environment is the space where many intelligent devices, such as computers and sensors, are distributed. According to the cooperation of many intelligent devices, the environment comes to have intelligence. The environment supports human, who exists in the intelligent environment, physically and informationally, so that he can use advanced computers and complicated mechanical system without feeling the stress.
It is necessary for the intelligent environment to acquire various information about humans and robots in the environment. When the environment does not know where humans and robots are respectively, the environment can not give the enough se r vice to the appropriate user as for the physical service especially. Therefore, it is considered that how to get the location information is the most necessary of all. The system with multiple color CCD cameras is utilized as one of the means to acquire the location information in an intelligent environment. It can achieve the human centered system because the environment acquires the location of human noncontactly and the equipment of the special devices isn't required for human. Moreover, camera has the advantage in wide monitoring area. It also leads to acquisition of details about objects and the behaviour recognition according to image processing.

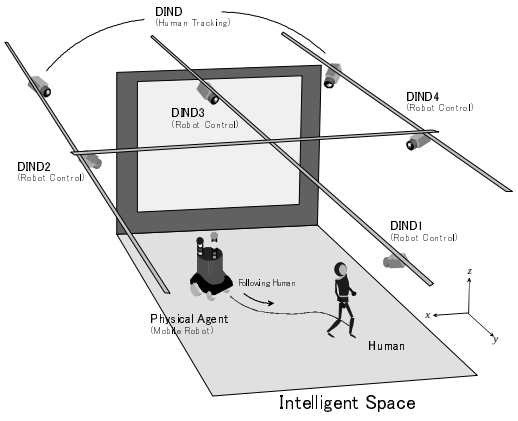


Fig. 1 Intelligent environment by distributed Cameras.

Our intelligent environment is achieved by distributing small intelligent devices which don't affect the present living environment greatly. Color CCD cameras, which include processing and networking part, are adopted as small intelligent devices of our intelligent environment. We call this environment "Intelligent Space (ISpace)"[3]. Intelligent Space is constructed as shown in Fig.1. This paper introduces the basic functions of the vision sensory system, what is the core of the iSpace, and the artificial spatial memory, the context-aware information exchange system for human-iSpace interaction is also introduced.

2. VISION SYSTEMS IN INTELLIGENT SPACE

2.1 Basic Scheme

Fig.2 shows the system configuration of distributed cameras in Intelligent Space. Since many autonomous cameras are distributed, this system is autonomous distributed system and has robustness and flexibility. Tracking and position estimation of objects is characterized as the basic function of each camera. Each camera must perform the basic function independently at least because over cooperation in basic level between cameras loses the robustness of autonomous distributed system. On the other hand, cooperation between many cameras is needed for accurate position estimation, control of the human following robot[4], guiding robots beyond the monitoring area of one camera[5], and so on. These are advanced functions of this system. This distributed camera system of Intelligent Space is separated into two parts as shown in Fig.2. This paper will focus on the tracking of multiple objects in the basic function.

2.2 Tracking of Moving Objects

Various tracking methods of moving objects using a vision system have been investigated. These can be separated in two major compartments. One is the method of matching and clustering of feature points extracted from an input image. The other is the method that the knowledge on objects is given to the system as an object advance and the model and an input image are compared. For example, the 3D ellipse model is used for human tracking in [7].

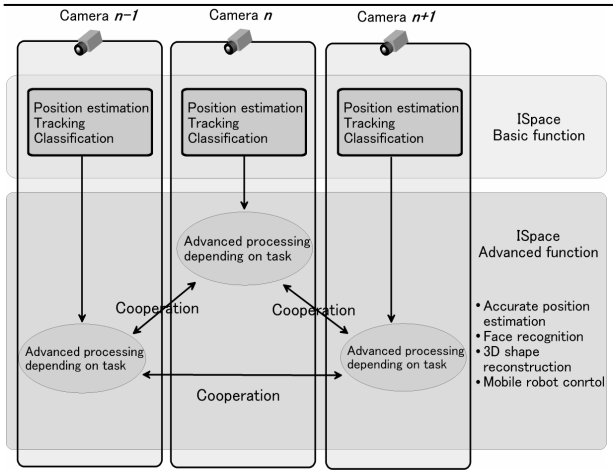


Fig. 2 Configuration of distributed camera system.

The former has the merit that various feature points can be extracted according to image processing, because the whole of the captured image can be always observed. However, matching of feature points between successive frames become difficult and computational cost increases, according to number of the feature points in the complicated scene and by the effect of noise. The other hand, in the latter method, only comparison between the model and input image is required. Tracking of moving objects is achieved by comparing the real image with the model. Computational cost is lower than the former. However, tracking systems have to prepare the models of the objects in advance. For example, human tracking for surveillance system needs human models[8] and vehicle tracking for ITS needs vehicle models[9]. Tracking cannot be achieved without object models. It is impossible to build a model of every object which exists in our daily life.

2.3 Requirements of Vision Systems in ISpace

Each camera of the vision system in Intelligent Space has to satisfy the following conditions in order to achieve the basic tracking function.

- Real time processing
- Response to multiple objects
- Extension to multiple cameras
- Overcoming partial occlusion

Especially, many kinds of objects, which are humans and the mechanical system like the mobile robots as physical agents, exist simultaneously in Intelligent Space. Matching and clustering of the feature points extracted from an input image increases the computational cost, so processing performance cannot be kept adequately. It is difficult to adjust the model based tracking to the tracking system of Intelligent Space where many kinds of objects exist. Therefore, the tracking method which has advantages of both methods is required for the vision system in Intelligent Space.

3. PROCESSING FLOW

3.1 Extraction of Objects

Our system uses many low-cost cameras to improve recognition performance. Position and viewing field of all cameras are fixed. Each camera is connected to a normal computer with a video capture board. It is necessary to extract only the moving objects robustly in order to simplify the matching process. Background subtraction is simple and efficient to recognize the moving objects in fixed camera image. Following process based on background subtraction is performed to extract the object region.

1) Background subtraction: Feature points are separated from captured image by comparison with the background image. In this part, the system doesn't discriminate if each feature point belongs to the moving objects or is just noise.

2) Clustering of feature points: Feature points which gather in the certain range are merged into a cluster which corresponds to one object. If number of pixels of which a cluster consists doesn't get to minimum number of the cluster treated as the object, the cluster is removed as the noise.

Fig.3 shows the example of results of this object extraction process mentioned above. Fig. 3(a) is the raw image captured by the CCD camera. Extracted objects, which are human and robot, are shown in Fig. 3(b), 3(c). It is clear that this can extract the multiple objects simultaneously. When the image of the size of 320X240 pixels is captured and Pentium III 866 MHz PC is used, this process is performed at the speed of 28 to 30 frames per second. In this process, a lot of processing time is not required. Matching process of the objects between successive frames is based on the information acquired from these extracted objects.

(a) (b) (c)

Fig. 3 Extracted objects.

3.2 Tracking Methods

For objects tracking, color histogram is configured from the objects extracted and separated by the method mentioned above. Since color histogram is stable to deformation and occlusion of the objects relatively[10], it is qualified as unique feature value to represent each object. Compared with the contour and so on, color histogram of the object stays largely unchanged against the various images that are captured by the distributed cameras. Therefore, it is considered that representation using color histogram is suitable for tracking multiple objects seamlessly in wide area that distributed cameras are monitoring. Color histogram of each object is acquired per each video frame. Feature vector of the object is configured by using color histogram. The cluster corresponding to each object is created dynamically based on feature vectors acquired in definite period of time. The regions which express the possibility of object existence are defined by clusters in the feature space.

At first of tracking process, it is considered which region feature vector of the object belongs to. When feature vector doesn't match the cluster accumulated in the system, it is considered that new object appeared in the monitoring area. Then, the process creating new cluster is performed. This process is iterated through successive video frames and tracking of multiple objects is achieved as the result of verification with the cluster. Here, the system doesn't recognize what the object is and cluster is created without depending on the kind of the object, such as human, robot and the other things. All objects extracted from the comparison

between background image and input image are tracked as the basic function in Intelligent Space. This method has advantages of both tracking method. Since cluster is created on the basis of the background subtraction and updated dynamically in each frame, this method is always able to observe the overall of the input image. Moreover, feature vectors acquired from different cameras for the same object are supposed to have similarity and gather in a cluster. This characteristic is useful for distributed camera system. Therefore, this is suitable for the use with tracking of multiple objects in real time. Tracking processing flow is shown as Fig.4.

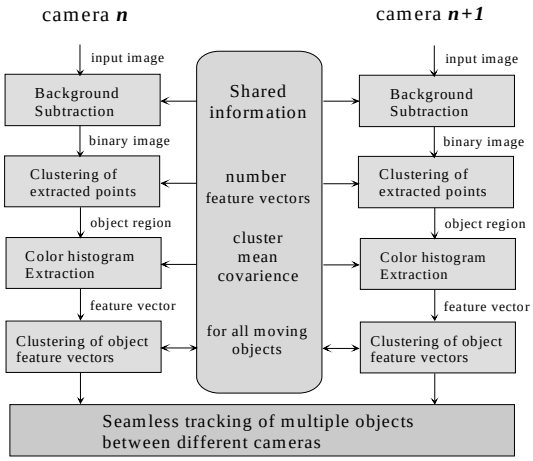


Fig. 4 Flow of the matching process.

4. OBJECT REPRESENTATION USING COLOR HISTOGRAM

4.1 Color Space Selection

Object representation using color histogram has some advantages as mentioned above. It is required that adequate color space is chosen to configure the color histogram for the object tracking. Many color spaces have been investigated before. RGB color space has a problem that each value of RGB changes significantly according to the variance of the illumination.

The other color spaces that linear transformation is applied to also have a similar tendency to the illumination. On the other hand, HSV color space that is nonlinear transformation of RGB color space is used frequently. HSV color space is expressed based on similarity of the three basic color attribute. Hue, saturation, and value are used in HSV color space. While value is corresponding to intensity of a pixel, hue and saturation have a little relation to the variance of the illumination. Therefore, hue and saturation can represent the objects in the wide area more robustly than RGB color space.

4.2 Color Histogram Configuration

In this research, the feature vectors of the objects are represented by the histogram based on HSV color space. Value is affected by the variance of the illumination, so it is desirable that the feature vector of the object consists of only hue and saturation. The histogram of each object is normalized by the all number of pixels in each object region, in order to cancel the effect from the size of the object region. Here, the range of hue and saturation is separated to 20 parts respectively to generate the histogram. This separation decides the resolution of the object recognition, however the number of part is decided experientially at present. Since hue has less reliability in the low saturation and low value domain, unreliable hue is eliminated by thresholding of saturation and value. Feature vector of the object is configured as shown Fig.5. Feature vector consists of 40 components derived from hue and saturation.

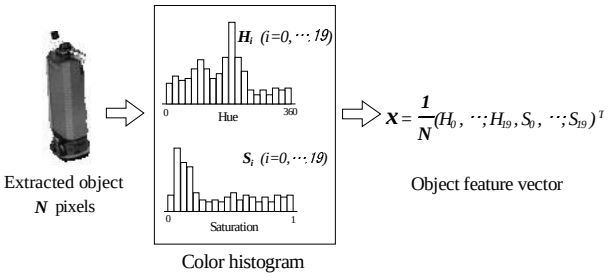


Fig. 5 Feature vector generation method.

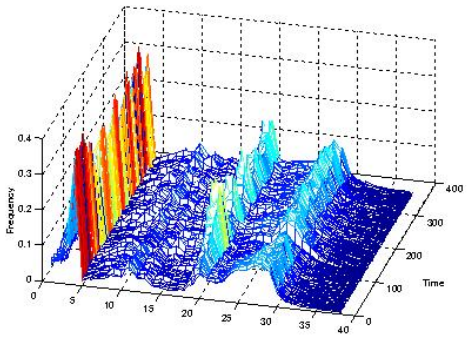


Fig. 6 Change of normalized color histogram against time.

Fig.6 shows the components of the feature vectors of human. These feature vectors were extracted continuously by one camera of Intelligent Space, when human was walking around in the monitoring area of the camera. The horizontal axis shows the components of the feature vectors, the vertical axis shows the color frequency in the object, and the depth expresses the time axis. The size of the object region increased and decreased, and the direction and the shape of the object were changed during human walking. However, sudden fluctuation of the feature vector doesn't appear in this figure. Therefore, the feature vector using normalized color histogram is supposed to represent the moving object reasonably.

Moreover, this representation of the objects has an advantage for information sharing between the distributed cameras, because feature vector of the object isn't affected by the direction and shape of the object. When different cameras observe the same object, the extraction results of the object are similar. This representation is useful for the distributed camera system.

5. OBJECT TRACKING

5.1 Clustering of Feature Vector

Feature vectors of the objects are extracted per a frame from different cameras. Feature vectors of the same object are

supposed to come together in a cluster in the feature space shared by adjacent cameras. Matching of the objects between frames and different cameras are achieved by online clustering of these feature vectors sequentially. Multiple objects tracking is realized based on iteration of this matching process. The diagrammatic illustration of this method is shown in Fig.7.

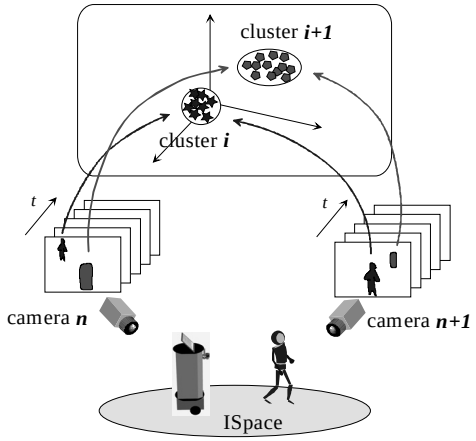


Fig. 7 Clustering at each sampling time and camera.

1) Initialization: Objects feature vectors are obtained after the objects extraction process that consists of background subtraction and clustering of feature points in binary image. When the number of objects is c in initialization, each obtained feature vector becomes the initial vector and the mean vector of each cluster. Here, c clusters are generated and i th cluster is represented as D_i. In following sections, m_i and n_i represent the mean vector and the number of feature vectors of D_i respectively. Feature vector is represented as x.

2) Clustering: In this process, it is decided whether obtained feature vector x belongs to any clusters in existence or a new cluster is generated. At first, square distance between feature vector x and each mean vector m_i of clusters is calculated to decide nearest neighbour cluster. Here, it is assumed that cluster $\hat{i}(D_{\hat{i}})$ gives least square distance. Next, it is evaluated whether feature vector x belongs to $D_{\hat{i}}$ or generate new cluster with Eq.(1). Left side of Eq.(1) represents square distance with mean vector $m_{\hat{i}}$ of cluster $D_{\hat{i}}$. Right side represents degree of scattering of cluster. This distance is compared with scattering of cluster in this equation. If this distance is longer than degree of scattering of cluster, it is justifiable that new cluster is generated. If not so, it is decided that x belongs to cluster $D_{\hat{i}}$. Here, α is a parameter that is decided experientially.

$$\| x - m_{\hat{i}} \|^2 > \alpha \cdot \mathrm{tr}[S_{\hat{i}}] \tag{1}$$

where, $S_{\hat{i}}$ is covariance matrix in D_i.

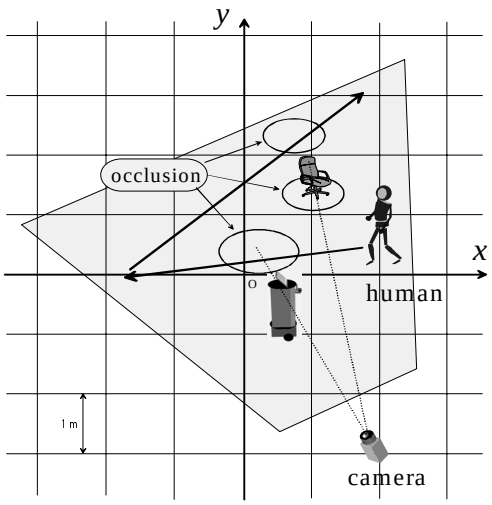
$$Si = \frac{1}{n_i} \sum_{x \in D_i} (x - m_i)(x - m_i)^T \tag{2}$$

If it is decided that feature vector x belongs to $D_{\hat{i}}$, mean vector $m_{\hat{i}}$ and covariance matrix $S_{\hat{i}}$ is updated. If not so, feature vector x becomes mean vector of the new cluster just like initialization process. Multiple objects tracking is achieved by iteration of this clustering process.

This criterion cannot classify clusters which are overlapped. Therefore, it is required that this method has to be extended to classify complicated clusters as the future work. Although this paper doesn't describe about sharing information of clusters between adjacent cameras, this will be shown in camera ready paper.

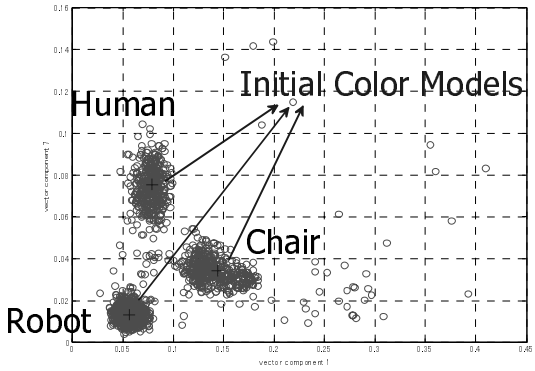
5.2 Tracking experiments

Some experiments are performed to verify this tracking met hod. Fig.8 shows the experimental environment and objects th at should be tracked by this method. Three objects, which are human, a mobile robot and a chair, exist in this environment. I n this experiment, the system does not have object models for these objects in advance. A mobile robot and a chair are static at the beginning and human is walking between them afterwar d. Since only one camera is used for this experiment, occlusio n between human and the other objects is supposed to happen as shown Fig. 8(a). Fig. 8(b) shows the clustering result of the feature vectors obtained in a given time, when three objects ex ist in the space as shown in Fig.8 (c). Object classification of t he new object is achieved based on the color pattern. In this cl ustering, the reference point of each cluster is treated as the ini tial local color model. Stable color model can be obtained by t his process.



(a)

(b)



(c)

Fig. 8 Experiment: moving area.

Fig. 9 shows the captured image by a camera in this experiment. Experimental result of multiple objects tracking is shown in Fig. 10. X axis and Y axis represent X and Y pixel coordinate of captured image respectively. Central pixels of each object are plotted.

Fig. 9 Multi-Objects detection and tracking in ISpace.

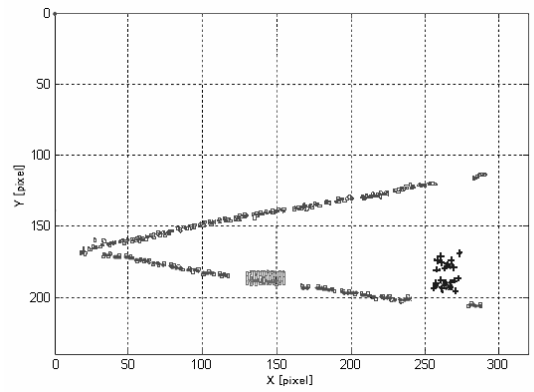


Fig. 10 Experiment results: Multi Objects tracking.

Although occlusion between human and other objects was observed during tracking of walking human, matching and tracking of each object achieved without fail. In this case, this system doesn't have the complex object models, however tracking of multiple objects was performed in low processing time.

6. ARTIFICIAL SPATIAL MAMORY FOR HUMAN-ISPACE INTERACTION

6.1 Context-Aware Information Exchange in iSpace

Serving context-aware information according to the human's current physical context is useful for the Human-Machine Interaction in order to enhance the speed of the information exchange. By building intuitive and instantaneous interaction among human and iSpace that facilitates manipulations of a large amount of data and a utilizing of DIND capability, we can combine human's ability with machines far greater computational capacity. Therefore, we focus on a new interface based on activity history of human such as trajectories or gestures in iSpace. We have proposed the spatial memory [15] which regards three dimensional space as external storage unit, i.e. three dimensional point is treaded as memory address to access stored memory such as various data and commands for DINDs. Consequently, we can access stored memory by using body action such like indication at the point. We call the body action human indicator. The spatial memory system is implemented by Spatial-Knowledge-Tags (SKT) shown in Fig.11. The spatial memory has two advanced points shown as follows;

(1) Humans can reduce stress to access stored memory because a spatial location as memory address plays a role of remembrance trigger. Additionally, human can intuitively and instantaneously access stored memory by moving their arms without using particular tools and particular training.

(2) Even though the spatial memory requires gesture recognition, it can be performed easily and definitely, because the requirement for DINDs is only to detect three dimensional points of human body and hand.

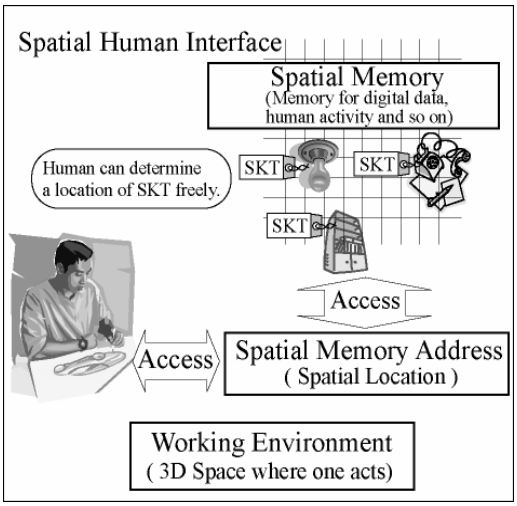


Fig.11 Whole system components of the spatial memory and data flows

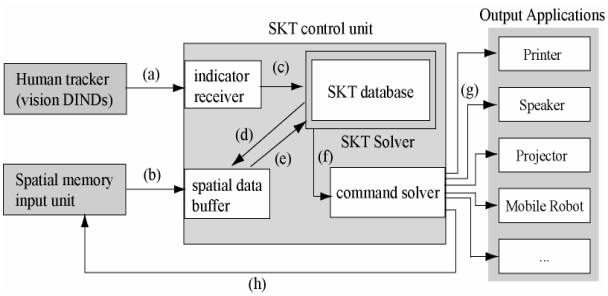


Fig.12 Whole system components of the spatial memory and data flows

6.2 Spatial memory

The SKT is a virtual tag, which has the spatial location and the path of stored memory. When human determines a spatial location and data, the SKT is created automatically and it becomes a spatial memory. Spatial memory system consists of four units shown in Fig.12. Human tracker is implemented by vision DINDs. Spatial memory input unit comprises a GUI application, cameras and microphones. SKT control unit is implemented by relational databases and control programs. Output applications include printer, projector, mobile robot, and so on.

Table.1 shows usages of spatial memory and required information to actualize them. Three usages are currently supposed. Spatial memory as a reminder is one of the most simple computational human memory augmentation system based on the location of human. Human leaves information or data in iSpace and the data is stored as spatial memory attached SKT which has the location information in the room. If human reaches the location where the data is stored, the data is provided for human. In the case of prepared information by iSpace, spatial memory plays a role of interface to interact between human and iSpace.

Table. 1 Usages of spatial memory and required information

Usage of spatial memory	Basic required information	Advanced required information
Workspace to handle specific visualized data	Positions of hands and body, posture of body	Accurate gesture recognition
Storage for Digital data	Positions of hands and body, posture of body	Hierarchy construction of memory, intention estimation
Reminder	Position of body, posture of body	Intention estimation

7. CONCLUSION

In this paper, the basic function of the vision system in iSpace and Spatial-Knowledge-Tag in order to support human memory were described. The vision system of iSpace needs real time processing, tracking of multiple objects, extension to cooperative multiple cameras network and overcoming partial occlusion. To realize them, it is required that model based method and feature based method are combined efficiently. Then, new tracking strategy was proposed based on extracting the objects by background subtraction and creating color appearance model dynamically with color histogram. This strategy achieved real-time and robust tracking of multiple objects. Especially, correct matching had been kept after the occlusion among objects happened in the experimental results.

As a future work, representation method of objects that are close to achromatic color will have to be investigated. Next, recognition of the wide area using the distributed cameras should be performed. It will need that different cameras share information about clusters and the feature space. Then, sharing method of the information that each camera acquires will be investigated. Additionally, we will implement SKT system in iSpace by utilizing the proposed tracking system and its performance will be evaluated.

REFERENCES

[1] B. Brumitt, B.Meyers, J.Krumm, A.Kern, S.Shafer, “EasyLiving: Technologies for Intelligent Environments”, Proceedings of the International Conference on Handheld and Ubiquitous Computing, September 2000.

[2] Rodney A.Brooks, “The Intelligent Room Project”, Proceedings of the Second International Cognitive Technology Conference(CT’97), Aizu, Japan, August 1997.

[3] Joo-Ho Lee, Hideki Hashimoto, “Intelligent Space -concept and contents”, Advanced Robotics, Vol.16, No.3, pp.265-280, 2002.

[4] Kazuyuki Morioka, Joo-Ho Lee, Hideki Hashimoto, “Human Centered Robotics in Intelligent Space”, IEEE International Conference on Robotics and Automation(ICRA’02), Washington D.C., USA, May 2002.

[5] Joo-Ho Lee, Kazuyuki Morioka, Hideki Hashimoto, “Cooperation of Intelligent Sensors in Intelligent Space”, IEEE Transactions on industrial electronics (submitted)

[6] Peter Norlund and Jan-Olof Eklundh, “Towards a Seeing Agent”, Proceedings of First International Workshop on Cooperative Distributed Vision, pp.93-120, 1997.

[7] Nakazawa Atsushi, Kato Hirokazu, Hiura Shinsaku, Inokuchi Seiji, “Tracking Multiple People using Distributed Vision Systems”, Proceedings of the 2002 IEEE International Conference on Robotics & Automation, pp.2974-2981, Washington D.C, May 2002.

[8] Christopher Wren, Ali Azarbayejani, Trevor Darrell, Alex Pentland, “Pfinder: Real-Time Tracking of the Human Body”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, pp.780-785, 1997.

[9] W.F.Gardner and D.T.Lawton, “Interactive modelbased vehicle tracking”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.18, pp.1115-1121, 1996.

[10] M.J. Swain, and D.H. Ballard, ”Color indexing”, International Journal of Computer Vision, Vol.7, No.1, pp.11-32, 1991.

[11] J.-H. Lee, T. Yamaguchi, and H. Hashimoto, “Human comprehension in intelligent space,” in Proc. IFAC Conf. Mechatronic Systems, pp. 1091–1096, 2000.

[12] J.-H. Lee, G. Appenzeller, and H. Hashimoto, “Physical agent for sensored, networked and thinking space,” in Proc. IEEE Int. Conf. Robotics and Automation, pp. 838–843, 1998.

[13] J.-H. Lee, N. Ando, and H. Hashimoto, “Design policy of localization for mobile robots in general environment,” in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 1733–1738, 1999.

[14] J.-H. Lee and H. Hashimoto, “Mobile robot control by distributed sensors,” in Proc. IFAC Workshop Mobile Robot Technology, pp. 85–90, 2001.

[15] M. Niitsuma, H. Hashimoto, H. Hashimoto and A. Watanabe, “An Architecture of Spatial- Knowledge-Tags to Access Memory of Human Activity,” Conf. of Japan Industrial Applications Society, 2004

[16] M. Niitsuma, H. Hashimoto, H. Hashimoto and A. Watanabe, “Spatial Human Interface in Working Environment -Spatial-knowledge-tags to Access the Memory of Activity-,” the 30th Annual Conf. of the IEEE Industrial Electronics Society, TC5-5, 2004.