International Journal of Fuzzy Logic and Intelligent Systems, vol. 4, no. 2, September 2004 pp. 165-171

Appearance Based Object Identification for Mobile Robot Localization in Intelligent Space with Distributed Vision Sensors

TaeSeok Jin*, Kazuyuki Morioka*, and Hideki Hashimoto**

***, ****Institute of Industrial Science, University of Tokyo,**
4-6-1 Komaba Meguro-ku Tokyo 153-8505, Japan

Abstract

Robots will be able to coexist with humans and support humans effectively in near future. One of the most important aspects in the development of human-friendly robots is to cooperation between humans and robots. In this paper, we proposed a method for multi-object identification in order to achieve such human-centered system and robot localization in intelligent space. The intelligent space is the space where many intelligent devices, such as computers and sensors, are distributed. The Intelligent Space achieves the human centered services by accelerating the physical and psychological interaction between humans and intelligent devices. As an intelligent device of the Intelligent Space, a color CCD camera module, which includes processing and networking part, has been chosen. The Intelligent Space requires functions of identifying and tracking the multiple objects to realize appropriate services to users under the multi-camera environments. In order to achieve seamless tracking and location estimation many camera modules are distributed. They causes some errors about object identification among different camera modules. This paper describes appearance based object representation for the distributed vision system in Intelligent Space to achieve consistent labeling of all objects. Then, we discuss how to learn the object color appearance model and how to achieve the multi-object tracking under occlusions.

Key words : Mobile robot, intelligent space, multi-vision sensor, position estimation, tracking

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. We proposed "Intelligent Space (iSpace)" [3][4] in order to achieve a human-centered services by accelerating the physical and psychological interaction between humans and environments. Color CCD cameras, which include processing and networking part, are distributed as the intelligent devices of the iSpace. We call these intelligent devices "DINDs (Distributed Intelligent Network Devices)". The Intelligent Space is constructed as shown in Fig.1. DINDs observe the positions and behavior of both humans and robots coexisting in the iSpace, and communicate with other DINDs. Based on the accumulated information in the DINDs, the position estimation of human and mobile robots based on the tracking of color marker[5], human behavior recognition[6] and mobile robot control under iSpace[7] have been studied.

It is important to track the objects without fail and to get the location of objects by DINDs, in order to make the Intelligent Space. There are two major problems in the multiple-camera multiple-object tracking system. One is the traditional correspondence problem from frame to frame over time. The other is the correspondence problem among different camera modules in order to achieve seamless tracking and location estimation. The later problem is called the consistent-labeling problem[8]. There are several approaches to solve this problem in the recent literatures. These approaches include feature matching[9], 3D information[10], [11], and alignment approach[12]. If all cameras are calibrated in advance, consistent labeling can be established by projecting the location of each object in the world coordinate system. Alignment approaches rely on recovering the geometric transformation between the cameras.

However, it is difficult for these approaches to establish consistent label without overlapping of the monitoring areas among different cameras. Feature matching approaches based on the color or others are the simplest scheme to establish consistent labeling. However, color feature matching is not reliable when the disparity is large in location and orientation. For example, if a person is wearing a shirt that has different colors on front and back, simple color matching among different cameras doesn't work. On the other hand, color information is useful for recognition and identification of objects in the interpersonal communication. If color representation, that absorbs the differences among different cameras and includes the color appearance model of all round the object, is achieved, color information is also useful for object identification in the communication among different camera modules. In this paper, color appearance based object representation for the distributed vision system in the Intelligent Space is described. At first, vision system in Intelligent Space will be explained. Then, this paper will show how to learn the object color appearance model, track the

Manuscript received Aug. 20, 2004; revised Sep. 2, 2004.

multi-object under occlusions, and achieve the correspondence among different cameras.

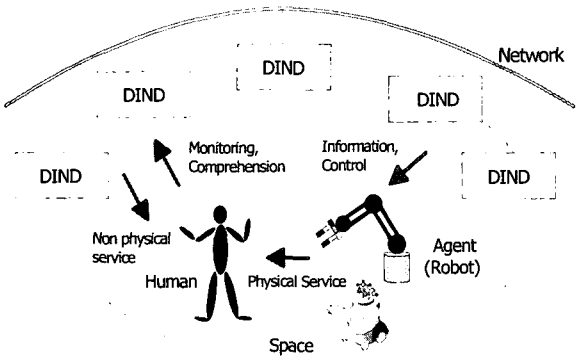


Fig. 1. Intelligent Space with DINDs.

2. Vision system in Intelligent Space

2.1 System Configuration

Figure 2 shows the system configuration of distributed cameras in Intelligent Space. Intelligent Space uses many low-cost cameras redundantly to improve recognition performances. Positions and monitoring areas of all cameras are fixed. These camera modules are regarded as DINDs. This distributed camera system of Intelligent Space is separated into two parts as shown in Fig.2. Tracking, consistent classification and position estimation of all objects are the basic functions of each camera. Each camera has to perform the basic function independently of condition of other cameras, because of keeping the robustness and the flexibility of the system. On the other hand, cooperation between many cameras is needed for accurate position estimation, control of mobile robots to supporting human[14], guiding robots beyond the monitoring area of one camera[13], and so on. These are advanced functions of this system.

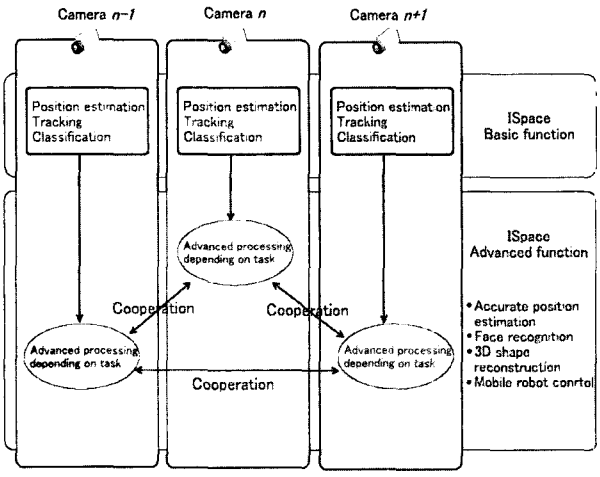


Fig. 2. Configuration of distributed camera system.

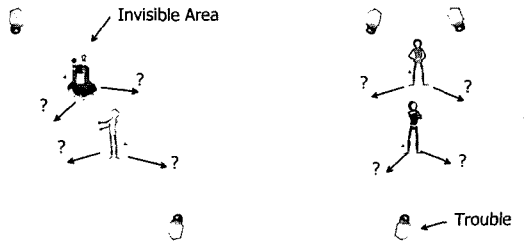
2.2 Required information for vision system

There are many camera modules and many kinds of objects, which are humans and mechanical systems like the mobile robots, in Intelligent Space. Each camera module should have the basic information about the environment, the objects and the other cameras. It is not realistic that human operators manually teach them to all camera modules in advance. The camera modules need the functions learning them automatically. Followings are the required information.

- Self-identification of camera parameter
- Object model learning for seamless tracking and labeling (focused in this paper)
- Environmental map generation Self identification algorithm for camera modules of Intelligent Space are also investigated[15].

3. Appearance based Object Identification

Consistent labeling capability, which identifies all objects based on the appearance models, is necessary in the situations shown in Fig.3. One is the case that the monitoring areas don't overlap and one camera does not work. For example, when two objects are in invisible area at once, the system cannot identify the objects without the appearance based models.



(a) Unoverlapping the monitoring areas (b) Trouble with the camera module

Fig. 3. Conditions requiring appearance based consistent labeling.

The color histograms of the extracted objects are used for the appearance based identification. The object representation based on the color histogram is stable against deformation and occlusion relatively [16]. 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. The object representation using color histogram is suitable for appearance based correspondence of multiple objects seamlessly in wide area. It is difficult to realize the correspondence among different cameras using the simple color histogram measured from one direction. On the other hand, color histogram representing the current appearance is needed for color region tracking in one camera image. Two kinds of object models are defined as follows to satisfy these requirements.

· Local Color Model

Local Color Model is used for object tracking and object segmentation under the occlusion in one camera image. It is updated according to the appearance of the object every frame.

· Global Color Model

Global Color Model means the object appearance model for matching among camera modules. It includes the information of the local color models in several postures of the object. Whole system of seamless labeling and tracking is shown in Fig.4. The system is separated by three parts: Object Finding process, Object Tracking process, and Global model learning process.

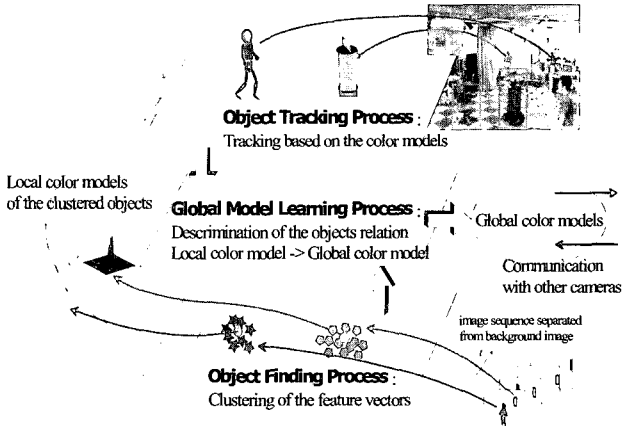


Fig. 4. Whole system configuration of the camera module.

3.1 Object Finding Process

Object finding process is the process to find the new moving objects in the monitoring area of the camera module. The local color models of the new objects is also acquired. Background subtraction is simple and efficient to find the new moving objects in fixed camera image. This background subtraction uses the background model updated from frame to frame adaptively. In Intelligent Space, this background subtraction works well since the lights and the floor are configured as reducing the effects of the shadow and lighting condition. The candidate regions of moving objects are extracted after the dilation, erosion, and clustering to the binary image separated from captured image by comparison with the background image. The small object region is removed as the noise.

The initial local color model is defined as follows. $\{x_i\}=1,\cdots,n$ is the pixel locations in the region extracted as the object. The function b associates to the pixel at location x_i the index $b(x_i)$ of its bin in the quantized feature space. Feature space is represented by two-dimensional normalized color space, e.g. $r=R/(R+G+B)$, $g=G/(R+G+B)$. The component p_u, $u=1,\cdots,m$ of the feature vector p in the object is then computed as

$$p_u=\frac{1}{n}\sum_{i=1}^{n}\delta[b(x_i)-u] \tag{1}$$

where δ is the Kronecker delta function.

Since the region extracted by background subtraction is unstable, several sets of p are required for each object in order to stabilize the initial local color model. There is also a probability that multiple objects are found simultaneously. The set of p should be clustered to some categories by the online clustering algorithm. It is decided whether obtained feature vector p belongs to any existing clusters or a new cluster is generated. The number of existing cluster is N at that time. At first, the similarity between feature vector p and each reference vector r_k of cluster is calculated to decide nearest neighbor cluster by Eq.(2). $p_{j,t}$ denotes jth object at the current time t.

$$S(p_{j,t}, r_{k,t})=\sum_k \min(p_{j,t}, r_{k,t}) \tag{2}$$

It is assumed that c represents the adequate cluster, and it is computed as

$$c=\begin{cases}\arg\max_k S(p_{j,t}, r_{r,t}) & \text{for } S(p_{j,t}, r_{k,t})\\ N+1 & otherwise\end{cases} \tag{3}$$

where, T is the threshold to evaluate the similarity between feature vectors.

The reference vector of each cluster is updated by Eq.(4). Updated vector is used as the reference vector at the next time $t+1$. α is the learning coefficient.

$$r_{k,t+1}=r_{k,t}+\alpha\,\delta_{ck}\{p_{j,t}-r_{k,t}\} \tag{4}$$

When the vectors beyond the threshold are gathered in one cluster, object candidate which correspond with this cluster is treated as the target object. The tracking process for each target object runs at that time. The reference vector $r_{k,t}$ of each cluster is treated as the local color model $l_{k,t}$.

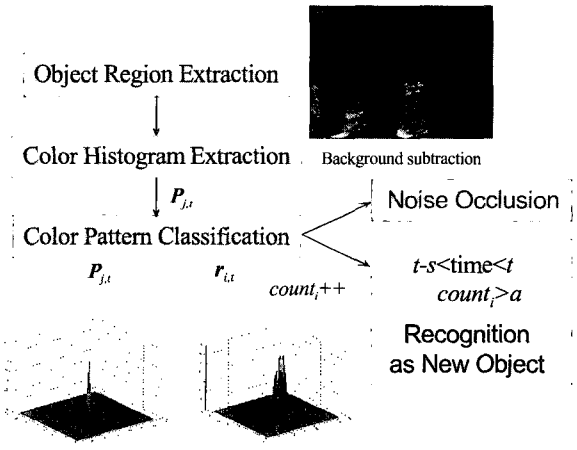


Fig. 5. Object Finding Process.

3.2 Tracking Process

Tracking process works for tracking of object region recognized in the object finding process. Tracking process receives the local color model, initial location and size of bounding box from the object finding process. Tracking process consists from two part as shown in Fig.6. One is the segmentation based on the local color model. Another is the position estimation by Kalman filter.

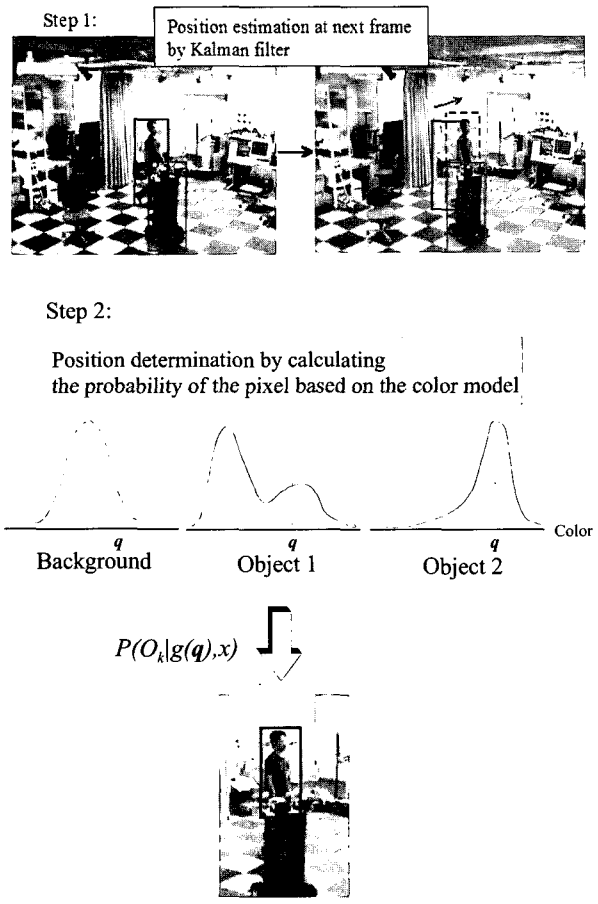


Fig. 6. Tracking Process.

Recently, the tracking system based on mean shift algorithm is reported that it is suitable for the color region tracking[17]. In this system, weighted mean shift is used for multiple color region tracking. Weight is computed as follows based on the color information of the object regions and background image. Object overlapping is detected by the overlap among the bounding boxes estimated by Kalman filter at the next frame. When n objects are overlapping and each object is denoted by $O_k(k=1,\cdots,n)$, the probability that color q at pixel x belongs to the bounding box of object O_k is computed as Eq.(6). The function g associates to the color q the index $g(q)$ of its bin in the quantized feature space. The local color model l_k is represented as,

$$l_i=[l_1^{(k)}, l_2^{(k)}, \cdots, l_m^{(k)}]^T \tag{5}$$

When the color q belongs to the bin u of the color histogram, $P(g(q)|O_k)$ corresponds to the component $l_u^{(k)}$ of the local color model of object O_k.

$$P(O_k|g(q),x)=\frac{P(g(q)|O_k)P(O_k|x)}{\sum_{i=0}^{n}P(g(q)|O_i)P(O_i|x)+P(g(q)|B,x)P(B|x)} \tag{6}$$

where, each pixel of the background image is modeled by Gaussian distribution. $P(g(q)|B, x)$ is computed from the background model as the probability of background color appearance at the pixel x. $P(O_k|x)$, $P(B|x)$ is then computed as

$$P(O_k|x)=P(B|x)=\frac{1}{n+1} \tag{7}$$

$P(O_k|g(q), x)$ is used as the weight per each pixel to compute the candidate of the centroid $\hat{x}_{g,k}$ of the object O_k.

$$\hat{x}_{g,k}=\frac{\sum_i x_i P(O_k|g(q),x_i)}{\sum_i P(O_k|g(q),x_i)} \tag{8}$$

This process is iterated several times changing the size of the bounding box until the candidate of the centroid converge. The pixel at the convergence is treated as the centroid $x_{g,k}$ of object O_k at that frame.

The local color model $l_{k,t}$ of the object k is updated when the objects don't overlap. The local color model $l_{k,t+1}$ at the next frame is computed by Eq.(9) based on the color feature vector $p_{k,t}$ made from pixels belonging to the object k. It is decided by $P(O_k|g(q),x)$ whether the pixel belongs to the object k. The color feature vector $p_{k,t}$ is configured as well as $p_{j,t}$ in the object finding process.

$$l_{k,t+1}=(1-\beta)\, l_{k,t}+\beta\, p_{k,t} \tag{9}$$

3.3 Global Model Learning Process

The global model means the object model for matching of objects measured by the different camera modules as mentioned above. Global model learning process runs when the occlusion among the objects doesn't happen in the tracking process. The global model is produced from local color models which have been measured since the object finding process started. This model should cancel the effects of the object posture, scaling or the direction of measurement by cameras for matching between the different camera modules. At first, it is decided whether the global model is learned, based on objects conditions such as overlapping and approaching. This condition is decided by evaluating the distance between the bounding boxes.

The local color models can be configured every frame in the tracking process. For example, the back of tracked human is captured in Fig.8(2). In Fig.8(8), the front of human is tracked. The set of the local models includes the color appearances that changes according to the posture of the object. If the color appearances of all round the object is

included in the object model, it is useful for consistent labeling from the view of different cameras. It is not reasonable to store all local color models to each objects in terms of the memory size. Global color models requires the effective representation from all local color models. The theory of Eigenspace Method, which is excellent in compression of large amount of image data and calculation of the correlation among images, is reported in [18]. Eigenspace method is applied to configuration of the global color model from a lot of local color models. Fig.7 shows the details of this process.

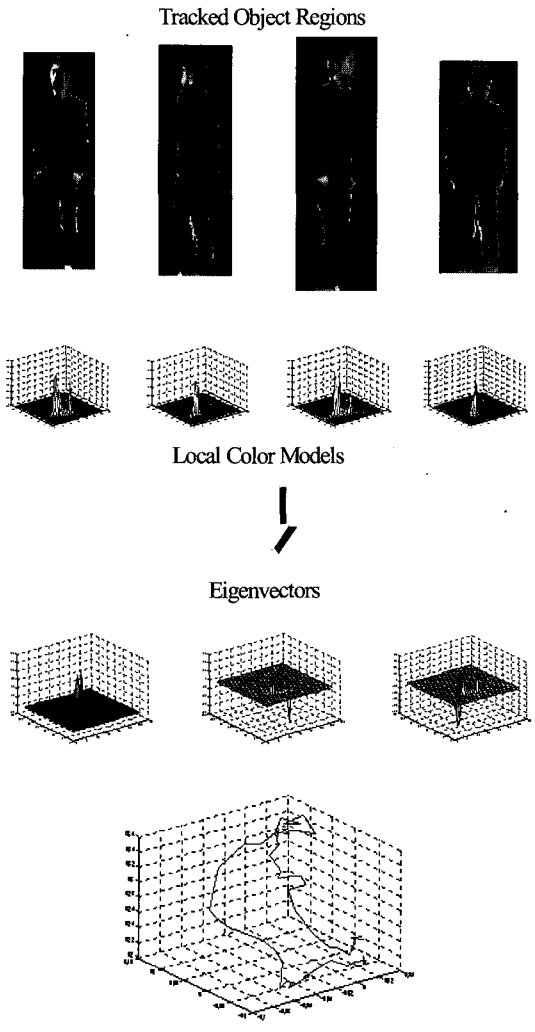


Fig. 7. Global Model Learning Process.

Global color model g_k of the object O_k at time t is acquired as follows. The covariance matrix of local color models is computed as,

$$Q = E\{(l_k - \bar{l})(l_k - \bar{l})^T\} \tag{10}$$

where l_k is the set of the local color models obtained until t and $\bar{l}$ is the mean vector for l_k. d eigenvectors $e_1, e_2, \cdots, e_d$ $(\lambda_1 > \cdots > \lambda_d > \cdots > \lambda_m)$ are determined by solving eigenvalues problem:

$$\lambda_k e_k = Q\ e_k \tag{11}$$

The d-dimensional subspace spanned by these d eigenvectors corresponding to d large eigenvalues is called the eigenspace. By ignoring the small eigenvalues, dimension of the local color model data is reduced. The cumulative proportion of eigenvalues in Eq.(12) is evaluated in order to determine the effective dimension.

$$W_d = \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{m} \lambda_i} > T_s \tag{12}$$

Then, one local color model is projected onto the eigenspace by

$$z_{k,t} = [e_1, \cdots, e_d]^T\ l_{k,t} \tag{13}$$

$z_{k,t}$ is a point that the local model $l_{k,t}$ at t is mapped to the eigenspace. The local models to each object can be represented as a manifold in the eigenspace. This manifold includes the local color models changing according to the posture of the object. Global color model g_k is represented as this manifold.

4. Experiment

Tracking experiments were performed to verify the object finding process and the tracking process. Experimental results of multiple object regions tracking are shown in Fig.8. Since monitoring area of each camera is limited and fixed for this system, occlusion between human and the other objects is supposed to happen as shown in Fig.8(1)~(8). In this experiment, the system does not have object models for these objects in advance. The object finding process found the objects #1,#2 at first. Human of object #3 was found and tracked under the occlusion afterward.

Fig.9 shows the global color model of the human tracked by object finding process and tracking process. A lot of local color models are projected onto the eigenspace spanned by three eigenvectors.

The global model is represented as the data sets of the local models compressed in this case. The cumulative proportion of eigenvalues in Eq.(12) and the threshold generally set 0.8 or 0.9 determine the effective dimension. This model can represent the change of the color appearance of the tracked human.

In Fig.10, the global color model obtained in the different camera is compared with the eigenspace as shown in Fig.9. Although the local color models change according to the difference of the camera, correspondence among cameras can be evaluated by the comparison of the manifold shape. The global model by camera2 is in process of the complete global model.

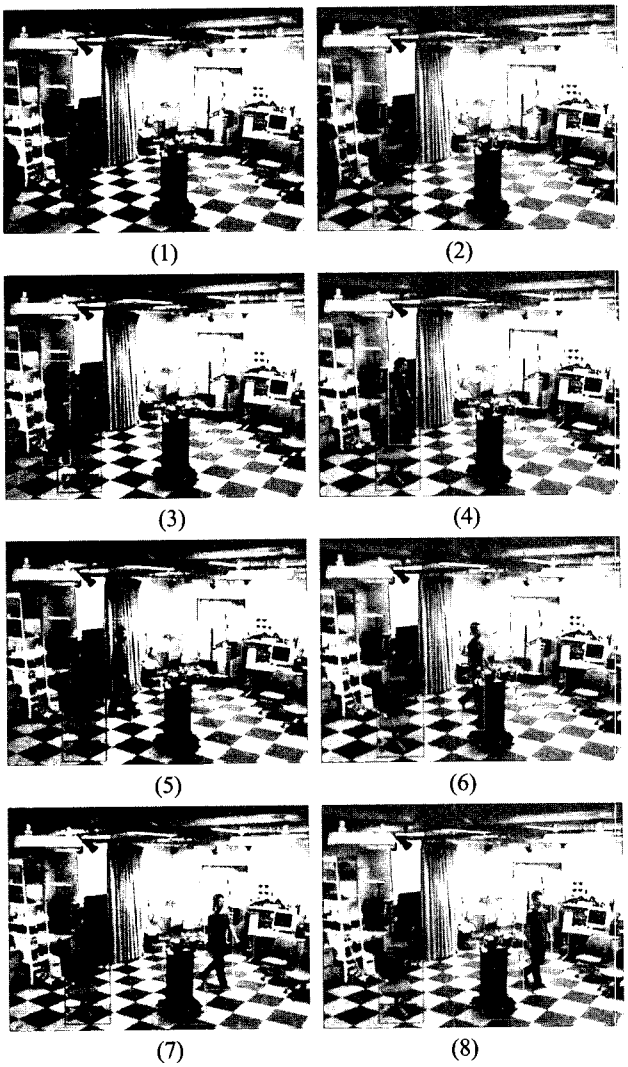


Fig. 8. Multi-objects Tracking results.

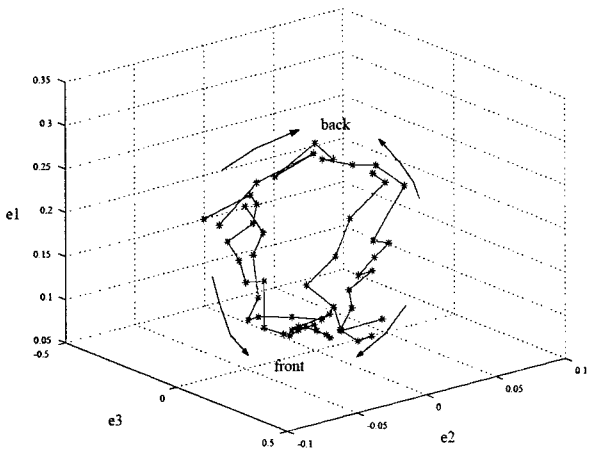


Fig. 9. Color Information in the Eigenspace.

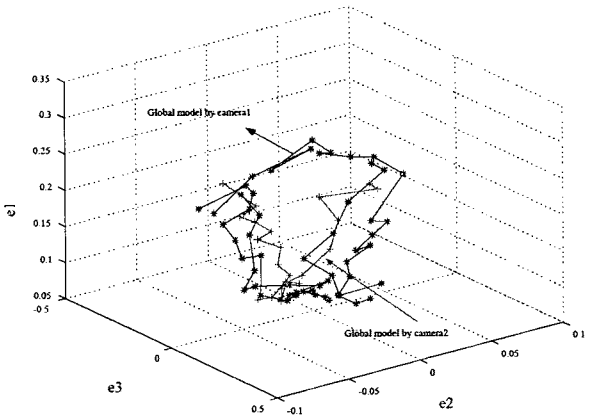


Fig. 10. Comparison of the Global Models.

5. Conclusion

In this paper, the basic function of the vision system in Intelligent Space was described. The vision system of Intelligent Space needs tracking of multiple objects, correspondence among different cameras network and overcoming partial occlusion. To satisfy these, it is required the appearance model based method. Then, the local color model and the global color model was proposed based on extracting the objects by background subtraction and creating color histogram. This strategy achieved robust tracking of multiple objects and consistent labeling among different cameras. Some experimental results are shown in Section IV. As a future work, representation method of objects that are close to achromatic color will have to be investigated. Next, recognition system in the wide area using the distributed cameras should be constructed. It will require the cooperation algorithm for different cameras to share information about the objects. In the camera ready paper, more detailed experimental results and analysis will be added.

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, pp.12-29.

[2] Rodney A.Brooks, "The Intelligent Room Project", Proceedings of the Second International Cognitive Technology Conference(CT'97), Aizu, Japan, August 1997, pp.69-74.

[3] J.-H. Lee, H.Hashimoto, "Intelligent Space - concept and contents", Advanced Robotics, Vol.16, No.3, 2002, pp. 265-280.

[4] H.Hashimoto, "Intelligent Space -How to Make Spaces Intelligent by using DIND", Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'02), 2002, pp.14-19.

[5] G. Appenzeller, J.-H. Lee and H.Hashimoto, "Building

Topological Maps by Looking at People: An Example of Cooperation between Intelligent Space and Robots," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'97), 1997, pp.1326-1333.

[6] J.-H. Lee, T. Yamaguchi and H. Hashimoto, "Human Comprehension in Intelligent Space," IFAC Conference on Mechatronic Systems, 2000, pp.1091-1096.

[7] J.-H. Lee, H.Hashimoto, "Controlling Mobile Robots in Distributed Intelligent Sensor Network", IEEE Transactions on Industrial Electronics, Vol. 50, No. 5, 2003, pp.890-902.

[8] S.Khan and M.Shah, "Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View", IEEE Transactions on Pattern Analysis and machine Intelligence, Vol.25, No.10, 2003, pp.1355-1360.

[9] A.Utsumi and J.Ohya, "Multiple-Camera-Based Human Tracking Using Non-Synchronous Observations", Proc. Asian Conf. Computer Vision, 2000, pp.1034-103.

[10] T.Matsuyama and N.Ukita, "Real-Time Multi-Target Tracking by a Cooperative Distributed Vision System", Proc. IEEE, Vol.90, No.7, 2002, pp.1136-1150.

[11] N.Atsushi, K.Hirokazu, H.Shinsaku, I.Seiji, "Tracking Multiple People using Distributed Vision Systems", Proceedings of the 2002 IEEE International Conference on Robotics & Automation, Washington D.C, May 2002, pp.2974-2981.

[12] Y.Caspi and M.Irani, "A Step Towards Sequence-to-Sequence Alignment", IEEE Conf. Computer Vision and Pattern Recognition, June 2000, pp.682-689.

[13] J.-H. Lee, K.Morioka, H.Hashimoto, "Cooperation of Intelligent Sensors in Intelligent Space", IEEE Transactions on industrial electronics (submitted).

[14] K.Morioka, J.-H. Lee, H.Hashimoto, "Human Centered Robotics in Intelligent Space", IEEE International Conference on Robotics and Automation(ICRA'02), Washington D.C., USA, May 2002, pp.2010-2015.

[15] H.Hashimoto, J.-H. Lee, N.Ando, "Self-Identification of Distributed Intelligent Networked Device in Intelligent Space", Proc. IEEE Int. Conf. on Robotics & Automation, 2003, pp.4172-4177.

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

[17] D.Comaniciu, V.Ramesh, P.Meer, "Kernel-Based Object Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.5, 2003, pp.564-577.

[18] H.Murase and S.K.Nayer, "Visual Learning and Recognition of 3-D Objects from Appearance", International Journal of Computer Vision, Vol.14, 1995, pp.5-24.

TaeSeok Jin

He received the B.Sc. degree from Jinju National University, M.Sc. and Ph.D. degrees from Pusan National University, Busan, Korea, in 2000 and 2003, respectively, all in electronics engineering. He is currently a Postdoctoral Researcher at the Institute of Industrial Science, The University of Tokyo, Japan. His research interests include sensor fusion, mobile robots, computer vision, redundant manipulator, and intelligent control. Dr. Jin is a Member of the KSME, JSME, IEEK, ICASE, and KFIS.

Phone : +81-3-5452-6258
Fax : +81-3-5452-6259
E-mail : jints@hlab.iis.u-tokyo.ac.jp

Kazuyuki Morioka

He received the B.E. and the M.S. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 2000 and 2002, respectively. He is currently working toward the Ph.D. degree at University of Tokyo, Tokyo, Japan. His research interests are intelligent environment, mobile robots, and computer vision. He is a student member of the Robotics Society of Japan.

Phone : +81-3-5452-6258
Fax : +81-3-5452-6259
E-mail : morioka@hlab.iis.u-tokyo.ac.jp

Hideki Hashimoto (S'83-M'84)

He received the B.E., M.E., and Dr.Eng. degrees in electrical engineering from The University of Tokyo, Tokyo, Japan, in 1981, 1984, and 1987, respectively.
He is currently an Associate Professor at the Institute of Industrial Science, The University of Tokyo. From 1989 to 1990, he was a Visiting Researcher at Massachusetts Institute of Technology, Cambridge. His research interests are control and robotics, in particular, advanced motion control and intelligent control. Dr. Hashimoto is a Member of the Society of Instrument and Control Engineers of Japan, Institute of Electrical Engineers of Japan, and Robotics Society of Japan

Phone : +81-3-5452-6258
Fax : +81-3-5452-6259
E-mail : hashimoto@iis.u-tokyo.ac.jp