분산다중센서로 구현된 지능화공간의 색상정보를 이용한 실시간 물체추적

Real-Time Objects Tracking using Color Configuration in Intelligent Space with Distributed Multi-Vision

진 태 석*, 이 장 명, 하시모토 히데키
(Tae-Seok Jin, Jang-Myung Lee, and Hideki Hashimoto)

Abstract : Intelligent Space defines an environment where many intelligent devices, such as computers and sensors, are distributed. As a result of the cooperation between smart devices, intelligence emerges from the environment. In such scheme, a crucial task is to obtain the global location of every device in order to offer the useful services. Some tracking systems often prepare the models of the objects in advance. It is difficult to adopt this model-based solution as the tracking system when many kinds of objects exist. In this paper the location is achieved with no prior model, using color properties as information source. Feature vectors of multiple objects using color histogram and tracking method are described. The proposed method is applied to the intelligent environment and its performance is verified by the experiments.

Keywords : multi-vision, tracking, intelligent space, distributed sensors, mobile robots

I. Introduction

In recent years, the research field on intelligent environment has been expanding[1,2]. An intelligent Environment represents the space where many smart and interconnected devices, such as computers and sensors, are distributed. As a result of the cooperation between them, the intelligence emerges from the environment. Such scheme supports human, who exists in the intelligent environment, physically and as another source of information.

It is a necessary task for the intelligent environment to obtain basic information about humans and robots inside the perception area of the space. It is of obvious importance that the environment obtains robust location information about such human users and robots inside it. Otherwise the basic services for which the environment is designed to be accomplished. Therefore, it is considered that the localization process should be the most basic and the most necessary among the rest.

The present paper proposes a system doted with multiple color CCD cameras as the perception layer, which is able to solve the location information in an intelligent environment. Among the numerous advantages about using cameras in an intelligent space approach, the most important is that it is able to achieve localization without the use of special beacons or devices in each object or human from which the location information is required. 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 image processing.

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. In this paper is described a method to represent feature vectors of multiple objects using color histogram. Then, the technique to achieve the tracking by using color information is described.

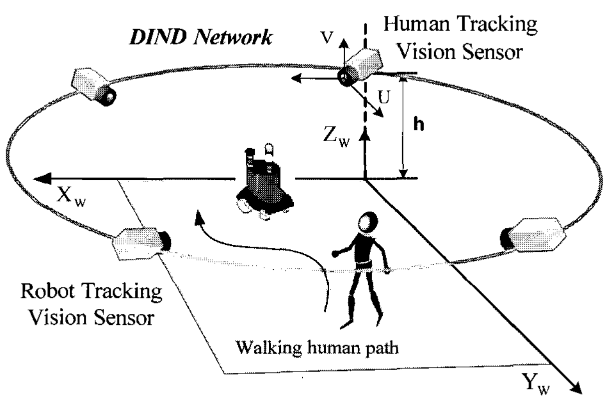


그림 1. 분산카메라로 구현한 지능형 공간의 개념.
Fig. 1. Concept of intelligent environment by distributed cameras.

II. Vision systems in intelligent space

1. Basic scheme

Fig. 2 shows the system configuration of distributed cameras in Intelligent Space. Since many autonomous cameras are placed, this system can be considered as an autonomous distributed system. Such scheme will present robustness and flexibility. Tracking and position estimation of objects is characterized as the basic function of each camera. Each camera is designed to perform the basic function independently. The

* 책임저자(Corresponding Author)
논문접수 : 2006. 4. 24., 채택확정 : 2006. 6. 10.
진태석 : 동서대학교 메카트로닉스공학과(jints@dongseo.ac.kr)
이장명 : 부산대학교 전자공학과(jmlee@pusan.ac.kr)
하시모토 히데키 : 동경대학 생산기술연구소
(hashimoto@iis.u-tokyo.ac.jp)

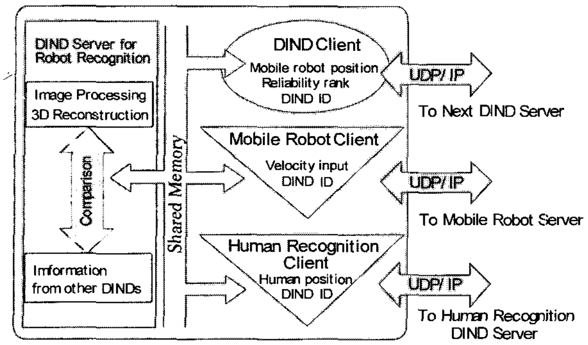


그림 2. 분산카메라시스템의 구성.

Fig. 2. Configuration of distributed camera system.

main reason to force a separate basic level localization process in each camera is justified as an improvement in the robustness of the whole distributed system. On the other hand, once the basic localization is obtained a better 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 can be considered as 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 as the basic function.

2. Tracking of moving objects

Various tracking methods of moving objects using a vision system have been extensively investigated. These can be separated in various approaches. First group use methods of matching and clustering of feature points extracted from an input image. For example, optical flows are extracted in a image, and tracking is achieved by clustering [6]. Another approach makes use of prior knowledge about objects which is given to the system as an object model in advance. The model and an input image can then be compared. For example, a 3D ellipse model is proposed for human tracking in [7]. The former has the merit that various feature points can be extracted according to image processing, because the whole captured image can be always observed. However, matching process of feature points between successive frames become difficult and computational cost increases. Success of such process is related to the number of the feature points in the complicated scene and by the effect of noise. The other approach, 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 model [8] and vehicle tracking for ITS needs vehicle models[9]. Tracking cannot be achieved without object models. However is not possible to build a model of every object which exists in our daily life. In this paper the location is achieved with no prior model, using color information as information source. Feature vectors of multiple objects using color histogram and tracking method are described in the following sections.

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 a tracking process that has advantages of both methods is required for the vision system in Intelligent Space.

III. Processing flow

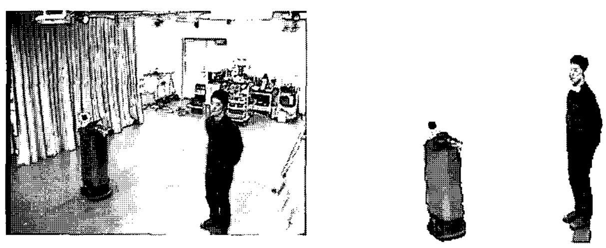
1. Extraction of objects

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 that gather in a certain range are merged into a cluster corresponding to one object. If number of pixels belonging to a cluster cannot reach a minimum number, the cluster is removed as a 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), (c). It is clear that this method is able to extract multiple objects simultaneously. When the image of the size of 320X240 pixels is captured and Pentium



(a) (b) (c)

그림 3. 추출된 영상 및 물체.

Fig. 3. Extracted objects.

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.

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 relatively stable to deformation and occlusion of the objects [10], then 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. Object's feature vector 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 feature vector region 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. A new cluster is created without depending on the kind of the object, such as human, robot or obstacles. 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.

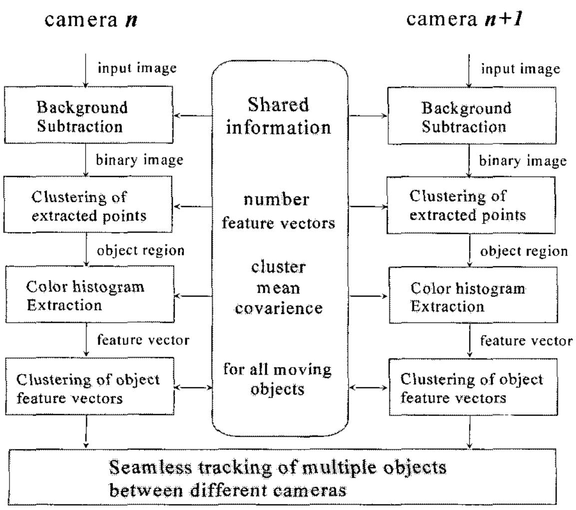


그림 4. 매칭과정의 흐름도.

Fig. 4. Flow of the matching process.

IV. Object representation using color histogram

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.

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 number of pixels in each object region. The goal is to cancel the effect from the variation in size of the object region. Here, the range of hue and saturation is separated in 20 parts respectively to generate the histogram.

This separation decides the resolution of the object recognition, however the number of separations is decided experientially in this paper. 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.

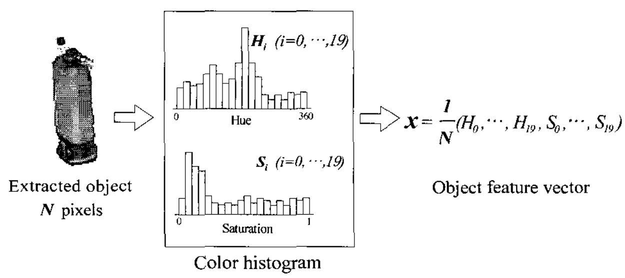


그림 5. 특성벡터 추출방법.

Fig. 5. Feature vector generation method.

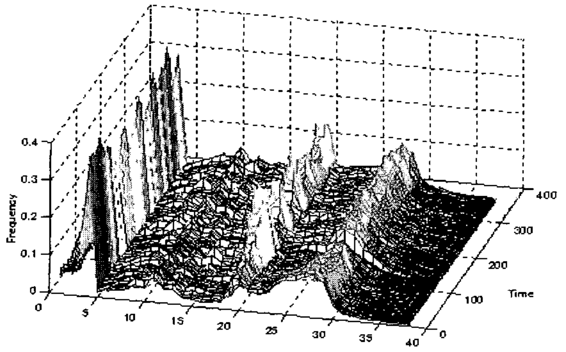


그림 6. 표준화된 칼라히스토그램의 변화.
Fig. 6. Change of normalized color histogram against time.

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.

V. Object tracking

1. Clustering of feature vector

Feature vectors of the objects are extracted each frame from different cameras. Feature vectors of the same object are supposed to fuse together in the same cluster of 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.

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 $\hat{i}(D_i)$ gives least square distance. Next, it is evaluated

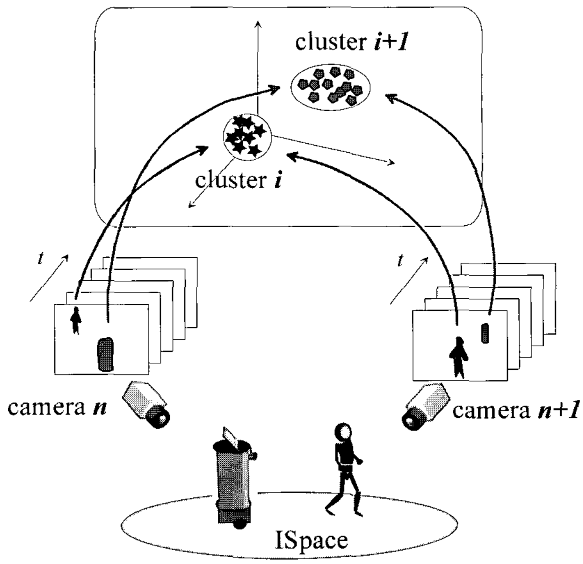
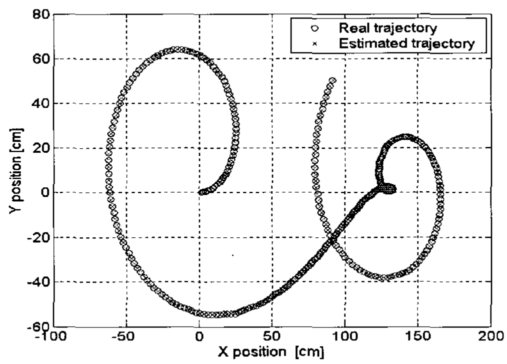
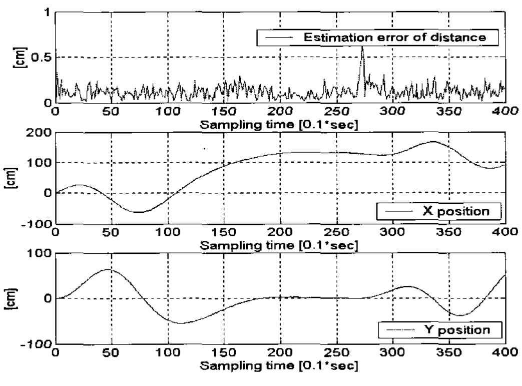


그림 7. 샘플링 타임 및 카메라에서의 클러스터링.
Fig. 7. Clustering at each sampling time and camera.



(a) Trajectory of moving object.

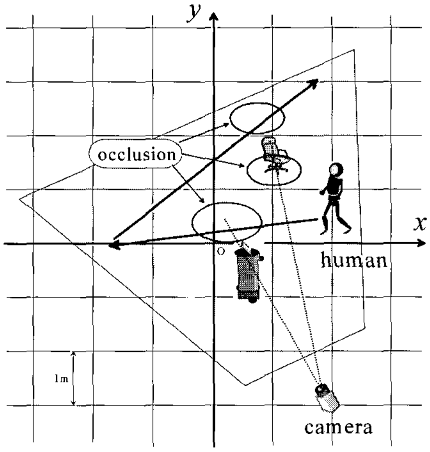


(b) Estimation error along the trajectory.

그림 8. 칼라모델에 의한 상태추정.
Fig. 8. State estimations using a color models.

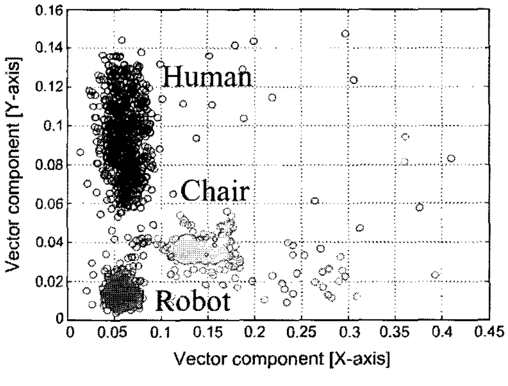
whether feature vector x belongs to D_i or generate new cluster with Eq.(1). Left side of Eq.(1) represents square distance with mean vector m_i of cluster D_i. Right side represents degree of scattering of cluster. This distance is compared with scattering of cluster in this equation. If this distance 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_i. Here, α is a parameter that is decided experientially.

$$\| x - m_i \|^2 > \alpha \cdot \mathrm{tr}[S_i] \tag{1}$$



(a)

(b)



(c)

그림 9. 실험: 다중물체에 따른 이동공간과 벡터성분.

Fig. 9. Experiment: moving area and vector component for objects.

where, S_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_i, mean vector m_i and covariance matrix S_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. Fig. 8(a) represents a real and an estimated trajectory of a moving object, while Fig. 8(b) represents the estimation error when the trajectory was estimated by the color histogram.

2. Tracking experiments

Some experiments are performed to verify this tracking method. Fig. 9 shows the experimental environment and objects that should be tracked by this method. Three objects, which are human, a mobile robot and a chair, exist in this environment. In 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 afterward. Since only one camera is used for this experiment, occlusion between human and the other objects is supposed to happen as shown Fig. 9(a). Fig. 9(c) shows the clustering result of the feature vectors obtained in a given time, when three objects exist in the space as shown in Fig. 9(b). Object classification of the new object is achieved based on the color pattern. In this clustering, the reference point of each cluster is treated as the initial local color model. Stable color model can be obtained by this process.

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. Experiment was performed to show the detecting and tracking a walking human. Fig. 10 shows

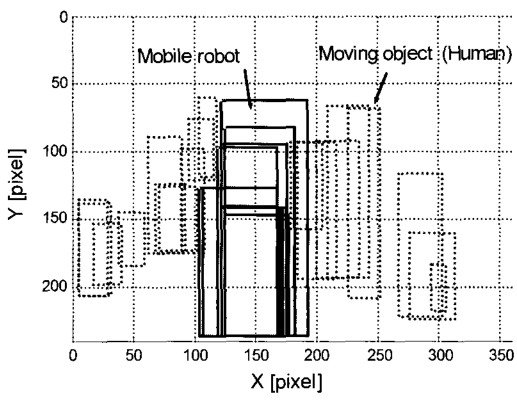


그림 10. 실험결과: 다중물체 추적.

Fig. 10. Experiment result: Multi-objects tracking.

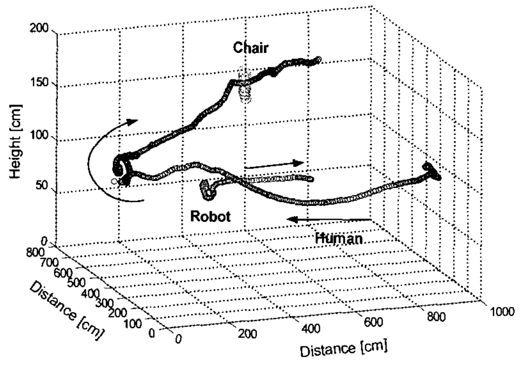
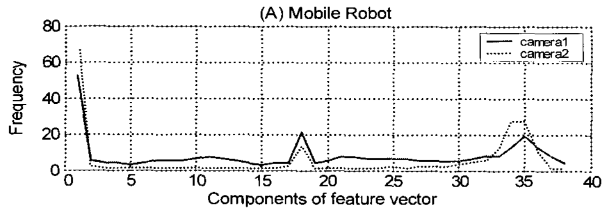
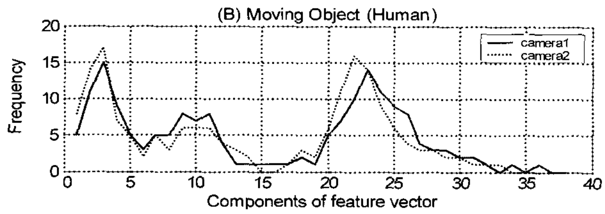


그림 11. 인간의 이동경로.

Fig. 11. Human's walking routes.



(a)



(b)

그림 12. 두 카메라에 의한 물체색상비교.

Fig. 12. Comparison of object color between two cameras.

the experimental results for detecting and tracking a walking human and obstacle. Mobile robot is attached with 70x20[cm] red-colored panels. Human walks random velocities in the range of 45-50[cm/sec]. First, mobile robot detects the moving human using cameras in ISpace. When a walking human is detected within view, mobile robot tracks it following the proposed method. And central pixels of each object are plotted. Although occlusion between human and other objects was observed during tracking of walking human, matching and tracking of each object achieved without fail.

Fig. 11 shows the route of the human tracked by object finding process and tracking process. Several local color models are obtained as shown in Fig. 9 during the tracking. Objects, human, robot, and chair, were tracked by the tracking process which was performed independently with the other process.

Fig. 12 shows the comparison of feature vectors extracted in each camera. Fig. 12(a) shows the feature vector of a mobile robot, and fig. 12(b) shows human. Since feature vectors have simirality between Cameral and Camera 2, this system can identify the object extracted from different cameras by communication between cameras. In this case, this system doesn't have the complex object models, however tracking of multiple objects was performed in low processing time. The image correspondence approach provided a fast and efficient mechanism to match objects across multiple viewpoints. The human image sequence had several instances of the object being partially occluded and in full visibility within the field of view of each camera, but this did not affect the performance of the matching or height estimation. The value of detecting occlusion is that it facilitates the learning of occlusion plane regions for each camera viewpoint that provides information on the scene. An image sequence where objects move behind and in front of occlusion planes could be used to constrain the occlusion plane region position in 3D based on the location and uncertainty of each 3D measurement.

VI. Conclusion

In this paper, the basic function of the vision system in Intelligent Space was described. The vision sys tem of Intelligent Space 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.

References

[1] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, "EasyLiving: Technologies for intelligent environments," *Proceedings of the International Conference on Handheld and Ubiquitous Computing*, September 2000.

[2] R. A. Brooks, "The intelligent room project," *Proceedings of the Second International Cognitive Technology Conference (CT'97)*, Aizu, Japan, August 1997.

[3] J.-H. Lee and H. Hashimoto, "Intelligent space-concept and contents," Advanced Robotics, vol. 16, no. 3, pp. 265-280, 2002.

[4] K. Morioka, J.-H. Lee, and H. Hashimoto, "Human centered robotics in intelligent space," *IEEE International Conference on Robotics and Automation(ICRA'02)*, Washington D.C., USA, May 2002.

[5] J.-H. Lee, K. Morioka, and H. Hashimoto, "Cooperation of intelligent sensors in intelligent space," *IEEE Transactions on industrial electronics (submitted)*.

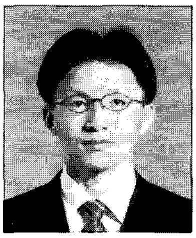
[6] P. Norlund and J.-O. Eklundh, "Towards a seeing agent," *Proceedings of First International Workshop on Cooperative Distributed Vision*, pp. 93-120, 1997.

[7] N. Atsushi, K. Hirokazu, H. Shinsaku, and I. 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] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780-785, July 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, Nov 1996.

[10] M. J. Swain, and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

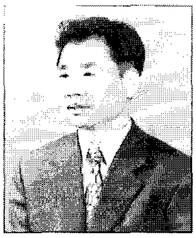


Tae-Seok Jin

He received the B.Sc. degree in Jinju National University, M.Sc. and Ph.D. degrees in Pusan National University, Korea, in 2000 and 2003, respectively, all in electronics engineering. From 2004 to 2005, he was a Researcher at the Institute of Industrial Science, The University of Tokyo, Japan. He is currently a full-time lecturer at the Dept. of Mechatronics engineering, DongSeo University. His research interests include intelligent space with multi-sensor fusion, mobile robot control. Dr. Jin is a Member of the ICASE, IEEK, and KFIS.

Hideki Hashimoto

He received the B.E., M.E., and Dr. Engineering 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. Prof. 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.



Jang-Myung Lee

Jang-Myung Lee has been a professor in the Department of Electronics Engineering, Pusan National University. He received his B.Sc. and the M.Sc. degrees in Electronics Engineering from Seoul National University in 1980 and 1982, respectively and his Ph.D. degree in Computer Electronics from the University of Southern California in 1990. His current research interests include intelligent robotic systems, integrated manufacturing systems, cooperative control and sensor fusion. Prof. Lee is an IEEE Senior member, and a member of ICASE and IEEK.