Human-following Robot Using the Particle Filter in ISpace with Distributed Vision Sensors

Tae-Seok Jin, Kazuyuki Morioka, and Hideki Hashimoto

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

Abstract

We present a method for representing, tracking and human following by fusing distributed multiple vision systems in intelligent space, with application to pedestrian tracking in a crowd. In this context, particle filters provide a robust tracking framework under ambiguity conditions. The particle filter technique is used in this work, but in order to reduce its computational complexity and increase its robustness, we propose to track the moving objects by generating hypotheses not in the image plan but on the top-view reconstruction of the scene. Comparative results on real video sequences show the advantage of our method for multi-object tracking. Simulations are carried out to evaluate the proposed performance. Also, the method is applied to the intelligent environment and its performance is verified by the experiments.

Keywords: Multi-vision sensors, Tracking, Intelligence Space, Mobile robot, Particle filter

1. Introduction

Video object tracking in dense visual clutter, although being notably challenging, has many practical applications in scene analysis for automated surveillance, such as the detection of suspicious moving objects (pedestrians or vehicles), or the monitoring of an industrial production [1][2][3][4]. The quality of an object tracking system is very much dependent on its ability to handle ambiguous conditions, such as occlusion of an object by another one. To cope with such ambiguities, multi-hypotheses techniques have been developed [5]. In the standard techniques using multi-hypotheses for the state estimation and tracking, the Kalman filter is used under the premise that the noise distributions are Gaussian and the system dynamics are linear [6]. However, when tracking human movements, non-linear and non-stationary assumptions make it suboptimal to use. In this context particle filter algorithms are attractive because they are both simple and very general. The particle filter algorithms track objects by generating multiple hypotheses and by ranking them according to their likelihood. They suppose that the correct hypothesis is retained [7][8]. Many tracking filters have been proposed using this approach, defining the states as being each static posture or position of the objects and modeling a motion sequence by the composition of these states with some transitional probabilities [9][10]. Those state-of-the-art techniques perform efficiently to trace the movement of one or two moving objects but the operational efficiency decreases dramatically when tracking the movement of many moving objects because systems implementing multiple hypotheses and multiple targets suffer from a combinatorial explosion, rendering those approaches computationally very expensive for real-time object tracking.

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, how to represent feature vectors of multiple objects using particle filter is described. Then, the technique to achieve the tracking by using color information is described.

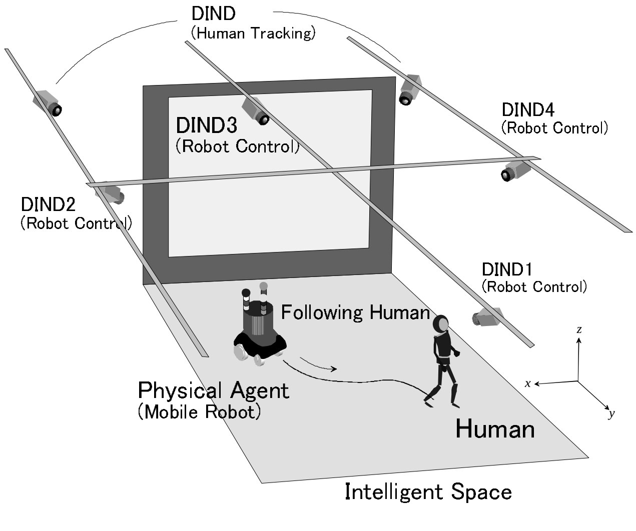


Fig. 1. Intelligent environment by distributed Cameras.

2. Vision Systems in Intelligence 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.

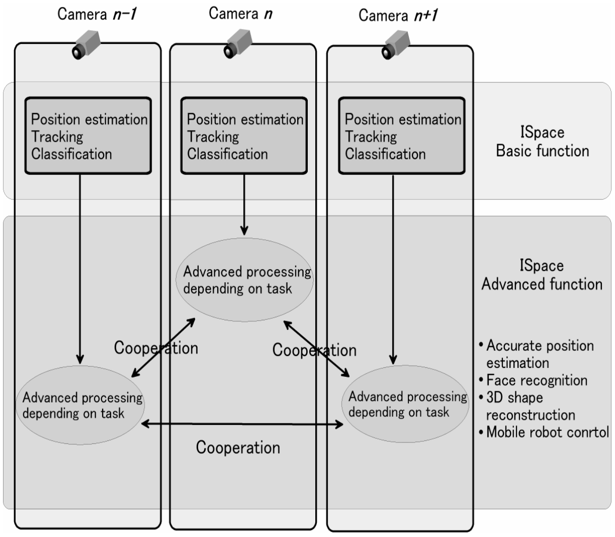


Fig. 2. Configuration of distributed camera system.

2.2 Previous Research for Moving Object Tracking

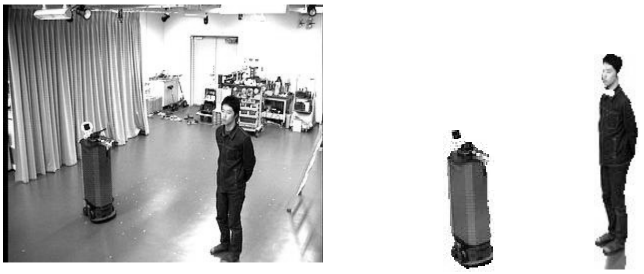
Various tracking methods of moving objects using a vision system have been investigated. These can be separated in two major compartments. One is the method of matching and clustering of feature points extracted from an input image. For example, optical flows are extracted in a image, and tracking is achieved by clustering of them[11]. The other is the method that the knowledge on objects is given to the system as an object model in advance and the model and an input image are compared. For example, the 3D ellipse model is used for human tracking in [12]. 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.

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.

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. Captured image and extracted objects.

3.2 Target regions encoded in a state vector Using Particle filter

Particle filtering provides a robust tracking framework, as it models uncertainty. Particle filters are very flexible in that they not require any assumptions about the probability distributions of data. In order to track moving objects (e.g. pedestrians) in video sequences, a classical particle filter continuously looks throughout the 2*D*-image space to determine which image regions belong to which moving objects (target regions). For that a moving region can be encoded in a state vector.

In the tracking problem the object identity must be maintained throughout the video sequences. The image features used therefore can involve low-level or high-level approaches (such as the colored-based image features, a subspace image decomposition or appearance models) to build a state vector.

A target region over the 2*D*-image space can be represented for instance as follows:

$$r = \{l,\ s,\ m,\ \gamma\} \tag{1}$$

where l is the location of the region, s is the region size, m is its motion and γ is its direction. In the standard formulation of the particle filter algorithm, the location l, of the hypothesis, is fixed in the prediction stage using only the previous approximation of the state density. Moreover, the importance of using an adaptive-target model to tackle the problems such as the occlusions and large-scale changes has been largely recognized. For example, the update of the target model can be implemented by the equation

$$\overline{\mathrm{r}}_t = (1-\lambda)\overline{\mathrm{r}}_{t-1} + \lambda \mathrm{E}[\mathrm{r}_t] \qquad (2)$$

where λ weights the contribution of the mean state to the target region. So, we update the target model model during slowly changing image observations.

4. Tracking moving objects

4.1 State-space over the top-view plan

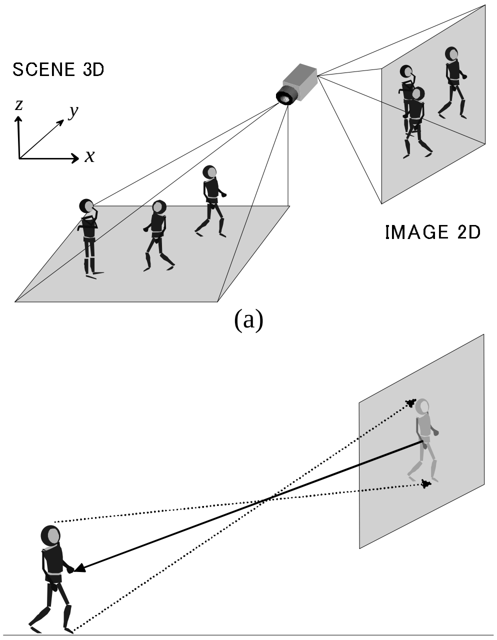
In a practical particle filter implementation, the prediction density is obtained by applying a dynamic model to the output of the previous time-step. This is appropriate when the hypothesis set approximation of the state density is accurate. But the random nature of the motion model induces some non-zero probability everywhere in state-space that the object is present at that point. The tracking error can be reduced by increasing the number of hypotheses (particles) with considerable influence on the computational complexity of the algorithm. However in the case of tracking pedestrians we propose to use the top-view information to refine the predictions and reduce the state-space, which permits an efficient discrete representation. In this top-view plan the displacements become Euclidean distances. The prediction can be defined according to the physical limitations of the pedestrians and their kinematics. In this paper we use a simpler dynamic model, where the actions of the pedestrians are modeled by incorporating internal (or personal) factors only. The displacements $\mathrm{M}^t_{topview}$ follows the expression

$$\mathrm{M}^t_{topview} = \mathrm{A}(\gamma_{topview})\mathrm{M}^{t-1}_{topview} + \mathrm{N} \qquad (3)$$

where $\mathrm{A}(.)$ is the rotation matrix, $\gamma_{topview}$ is the rotation angle defined over top-view plan and follows a Gaussian function $g(\gamma_{topview};\sigma_\gamma)$, and N is a stochastic component. This model proposes an anisotropic propagation of M : the highest probability is obtained by preserving the same direction. The evolution of a sample set is calculated by propagating each sample according to the dynamic model. So, that procedure generates the hypotheses.

4.2 Estimation of region size

The size of the search region represents a critical point. In our case, we use the *a-priori* information about the target object (the pedestrian) to solve this tedious problem. We assume an averaged height of people equal to 160 cm, ignoring the error introduced by this approximation. That means, we can estimate the region size s of the hypothetical bounding box containing the region of interest $\mathrm{r} = \{\mathrm{l}, \mathrm{s}, \mathrm{m}, \gamma\}$ by projecting the hypothetical positions from top-view plan in Fig. 4. A camera calibration step is necessary to verify the hypotheses by projecting the bounding boxes. So this automatic scale selection is an useful tool to distinguish regions. In this way for each visual tracker we can perform a realistic partitioning (bounding boxes) with consequent reduction in the computational cost. The distortion model of the camera's lenses has not been incorporated in this article. Under this approach, the processing time is dependent on the region size.



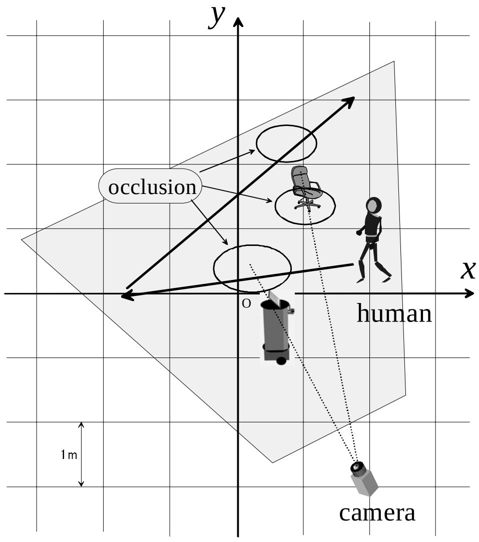
(a)

(b)

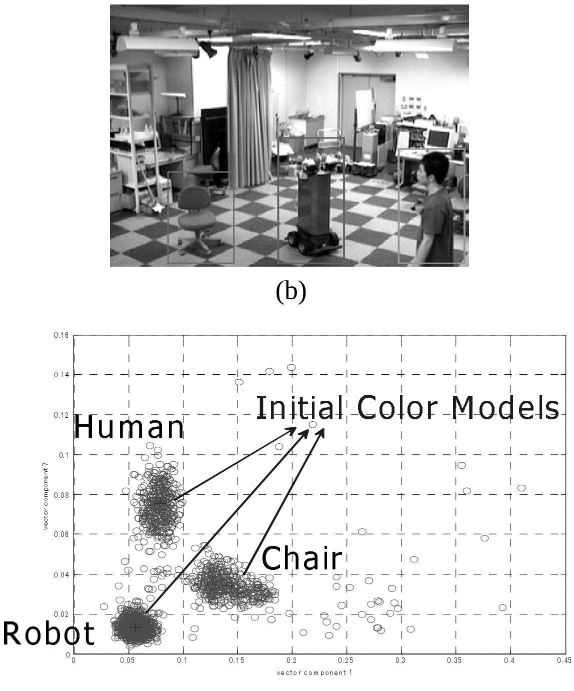
Fig. 4. (1) the approximation of Top-View plan by image plan with a monocular camera, (2) size estimation

4.3 Tracking experiments

Some experiments are performed to verify this tracking method. Fig.5 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. 5(a). Fig. 5(b) shows the clustering result of the feature vectors obtained in a given time, when three objects exist in the space as shown in Fig. 5(c).



(a)



(b)

(c)

Fig. 5. Experiment: moving area and models.

Fig. 6 shows the captured image by a camera in this experiment. Experimental result of multiple objects tracking is shown in Fig. 7. X axis and Y axis represent X and Y pixel coordinate of captured image respectively. 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. In this case, this system doesn't have the complex object models, however tracking of multiple objects was performed in low processing time.

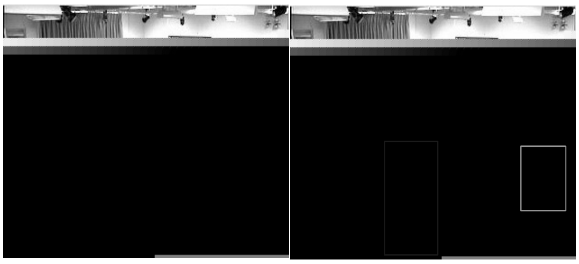


Fig. 6. Multi-Objects detection and tracking in ISpace.

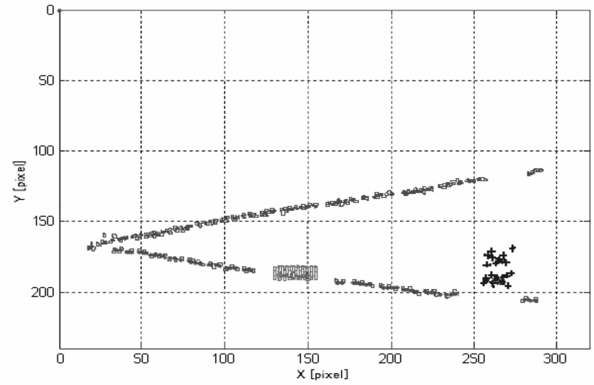


Fig. 7 Experiment results: Multi Objects tracking.

5. 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 particle filter. 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] A. W. Senior, "Tracking with Probabilistic Appearance Models," in *Proc ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems*, pp 48-55, June 2002.

[2] M. Bierlaire, G. Antonini and M.Weber, "Behavioural Dynamics for Pedestrians," *in K. Axhausen (Ed.), Moving through nets: the physical and social dimensions of travel*, pp. 1-18, Elsevier. 2003.

[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] K. Choo, and D.J. Fleet, "People tracking using hybrid Monte Carlo filtering," In *Proc. Int. Conf. Computer Vision*, vol. II, pp. 321-328, Vancouver, Canada, 2001.

[6] B. Anderson, and J. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, 1979.

[7] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte-Carlo Methods in Practice*, Springer Verlag, April 2001.

[8] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," Journal of Computational and Graphical Statistics, Vol. 5(1), pp. 1-25, 1996.

[9] K. Nummiaro, E. Koller-Meier, L.J. Van Gool, "Object Tracking with an Adaptive Color-Based Particle Filter," *DAGM-Symposium Pattern Recognition*, pp. 353-360, 2002.

[10] J. Vermaak, A. Doucet and P. Perez, "Maintaining Multi-Modality through Mixture Tracking," *International Conference on Computer Vision*, ICCV2003, Nice, France 2003.

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

[12]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.