Acquisition and Accumulation of Object Appearances using Successive Eigenspace Construction

Yasushi MAE, Tomohiro UMETANI, Tatsuo ARAI, and Kenji INOUE
Department of Systems and Human Science
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, JAPAN
mae@sys.es.osaka-u.ac.jp

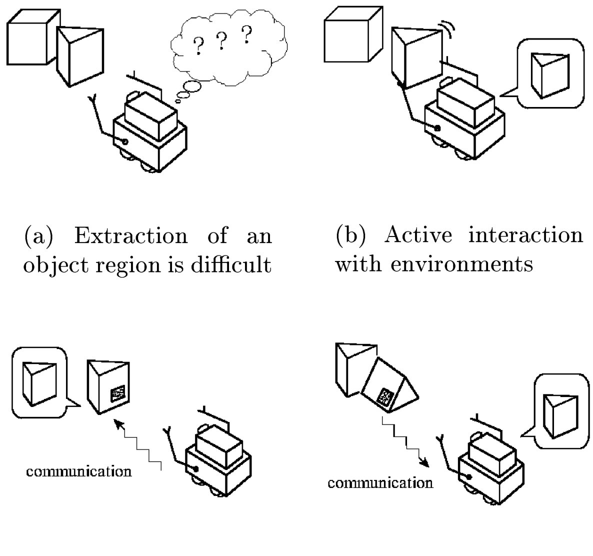
Abstract

This paper describes a method of acquisition of object appearances by active interaction with environments and accumulation of object appearances using successively constructed eigenspace. Robots directly interact with environments in order to extract an object region from a scene. Then robots acquire an appearance of an object and store it into environmental-attached storage devices such as RFID tag. Eigenspace is constructed successively every time a new appearance is obtained and various appearances are accumulated gradually. A closed sequence of the appearances is generated from the accumulated appearances which can be used for 3-D pose estimation of the object. Experimental results show effectiveness and validity of the proposed method.

1 Introduction

We hope autonomous mobile robots assist human beings in home and office. In such environments, robots have to recognize variable scenes where location and pose of objects change. Appearance of objects is one of important keys to find objects and determine their poses in a scene. We define an appearance of an object as an image taken from a viewpoint.

Autonomous robots should acquire object appearances by themselves without human assistance. It is difficult, however, to segment an object region from a scene only by vision without knowing the object appearances previously (**Fig.1(a)**). One of the other problems is to determine which appearances belong to the identical object from the appearances acquired at different times and situations. Even after robots acquired all the appearances of all objects in the environment, it needs much cost to determine which objects are in the current scene, since the object appearance in a scene must be searched for in the enormous database of objects' appearances. Distributed information management is one of the ways of solving these problems.



(a) Extraction of an object region is difficult

(b) Active interaction with environments

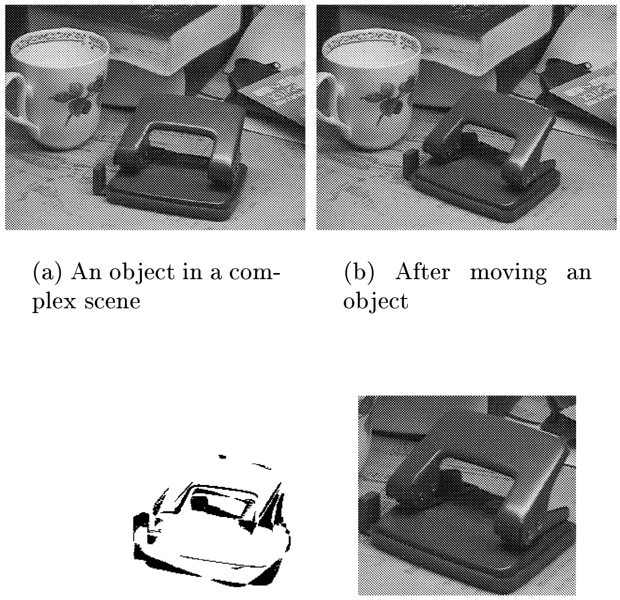
(c) Storing object appearance

(d) Obtaining object appearance

Figure 1. Object recognition using ID tag.

Recently, such as RFID(Radio Frequency Identification) tag and IDC(Intelligent Data Carrier)[1, 2] have been developed, which can be attached to environments and used for distributed information management. By using such environmental-attached storage devices, the robots can share the acquired knowledge through the devices[3, 4]. For object recognition, there has been proposed a method which stores acquired appearances into such environmental-attached storage devices [5]. In the method, robots acquire object appearances by active interaction with environments (**Fig.1(b)**). The acquired appearances are stored into the devices (**Fig.1(c)**). Then other robots can obtain the object appearances from the devices even the surrounding of the object changes (**Fig.1(d)**). The paper also proposed a method of successive construction of eigenspace from object appearances acquired successively. The paper, however, does not evaluate the successively constructed eigenspace quantitatively.

This paper describes a method of successive eigenspace construction from successively acquired object appearances, and evaluates the difference between the successively constructed eigenspace and conventional simultaneously constructed eigenspace. This paper also proposes a method of generating a closed sequence of object appearances, which can be used for estimating 3-D pose of the object. In the successive construction of eigenspace, similar appearances of objects are removed. As a result, almost the same eigenspace with the simultaneously constructed eigenspace is constructed successively with a small number of appearances. We ascertain that the pose of an object changes consecutively in the generated closed sequence. The relative 3-D pose of the object is estimated by corresponding the features in the consecutive appearances.



(a) An object in a complex scene

(b) After moving an object

(c) Extracted object regions

(d) Appearance

Figure 2. Detecting appearance of an object.

2 Object Appearance Acquisition by Active Interaction with Environments

We introduce a method which extracts an object appearances using the contact between the arm and the object in environments [5].

We suppose target objects have appropriate size and weight which robots can move and manipulate. We also suppose a robot has an arm which can move objects and can perceive the contact between the arm and the object. The robots wander the environment with fumbling motion of their arms. If the arm contacts the environment and an object moves, the moved portion in a scene is assumed to belong to an object. We also assume the object is not occluded by any other objects when the object moves.

When the arm contacts the object and the object moves a little, the portion of the image where brightness changes corresponds to the boundary of the objects and/or texture on the object surface. Then the robot can obtain the moving region on the image by subtracting the brightness of the images. The image pattern of the moving region is considered as an object appearance. The circumscribing rectangle region is extracted as an object appearance.

Next, we show an example of acquisition of an object appearance using a slight movement of an object by contact with the robot arm. In **Fig.2(a)**, it is difficult to extract a puncher region from the scene, since there are several objects in the scene. However, if the arm contacts the object and moves it a little by fumbling motion of the arm (**Fig.2(b)**), the portion with brightness change can be extracted as a moving region (**Fig.2(c)**). Thus, the image of the rectangle region including this moving region is extracted as an object appearance (**Fig.2(d)**). Though the image, object appearance, includes the background of the object, the dominant area is occupied by the object. In this way, the robot can acquire an object appearances from the current viewpoint.

If multiple robots wander the environment with fumbling their arms, and acquire appearances of objects, and store the acquired appearances into the environmental-attached storage device, then various appearances are accumulated in the devices gradually.

3 Accumulation of Appearances

3.1 Accumulation of Appearances using Eigenspace

Eigenspace is useful to store many appearances in a small memory space, since it projects images to a lower dimensional eigenspace keeping principal features. In conventional methods [7, 8, 9], eigenspace is constructed from the previously given images. In our method, however, eigenspace is constructed successively, since robots acquire object appearances successively. In the process of successive construction of eigenspace, the accumulated appearances are selected so that the various appearances of an object are accumulated on the eigenspace.

The dimension of the image must be fixed for constructing eigenspace. The vertical and the horizontal size of the images of object appearances are normalized to be predetermined size. The normalized squared image is called appearance model.

3.2 Successive Eigenspace Construction

Eigenspace is reconstructed every time robots observe an object and acquire a new appearance model. An appearance model is denoted by $\boldsymbol{X}$. The dimension of $\boldsymbol{X}$ is denoted by n: the number of pixels. The number of times of the observation by robots is denoted by t. The dimension of the eigenspace is predetermined and denoted by m.

If $t < m + 1$, the appearance model is stored into a storage device as it is.

If $t = m + 1$, the initial eigenspace is constructed from the first t models. An acquired appearance model is denoted by $\boldsymbol{X}_i$. A mean vector of $\boldsymbol{X}_i (i = 1, \ldots, t)$ is denoted by $\boldsymbol{m}$. The n-dimensional feature vectors are derived from

$$\boldsymbol{x}_i = \boldsymbol{X}_i - \boldsymbol{m}. \tag{1}$$

Eigenvectors $\boldsymbol{a}_i (i = 1, \ldots, m)$ corresponding to the largest m eigenvalues of the covariance matrix of $\boldsymbol{x}_i (i = 1, \ldots, t)$ compose $m \times n$ projection matrix A. By using the projection matrix A, n-dimensional feature vector x_i is projected to m-dimensional eigenspace. The projected vector $\boldsymbol{y}_i$ is represented by

$$\boldsymbol{y}_i = A\boldsymbol{x}_i. \tag{2}$$

Some $\boldsymbol{y}_i$ are selected and stored in accordance with the following model selection process. The m-dimensional eigenspace is divided to m-dimensional hyper-cubic cells. If multiple projected points are in a cell, the points except one nearest to the center of the cell is removed so that each cell includes a model at most. The projection matrix A and the selected projected vectors $\boldsymbol{y}_i$ on the eigenspace are stored instead of the appearance models $\boldsymbol{X}_i$. The number of stored models is denoted by N.

If $t > m + 1$, the following process is performed. appearance models are retrieved from the stored $\boldsymbol{y}_i, (i = 1, \ldots, N)$. The retrieved appearance model $\boldsymbol{X}'_i$ is represented by

$$\boldsymbol{X'}_i = A^T \boldsymbol{y}_i + \boldsymbol{m}, \tag{3}$$

where, $\boldsymbol{m}$ is the mean vector of $\boldsymbol{X}_i (i = 1, \ldots, N)$ at $t - 1$. $\boldsymbol{X}_i{}'$ approximates the original image $\boldsymbol{X}_i$ on the n-dimensional original image space. The eigenspace is reconstructed using the N retrieved appearance models $\boldsymbol{X}'_i, (i = 1, \ldots, N)$ and a newly acquired appearance model $\boldsymbol{X}_i (i = N + 1)$. If multiple appearance models are projected in a hyper-cell of the reconstructed eigenspace, the most recent model is selected and stored.

Every time robots observe the object and acquire a new appearance model, the same reconstruction process of $t > m + 1$ is repeated. As a result, eigenspace is constructed from a small number of appearance models which are acquired from various viewpoints. When every cell includes a model, the accumulated models are distributed uniformly on the eigenspace. The size of the hyper-cubic cell relates the resolution of accumulated object appearances. It should be determined based on varieties of the appearances of the object.

4 Generation of Appearance Model Sequence

We generate a closed sequence of appearance models from accumulated appearance models in order to obtain geometric relations between accumulated appearance models.

The distance $D_{i,j}$ between appearance models i, j is defined as follows:

$$D_{i,j} = \frac{1}{N} \sqrt{\sum_{k=1}^{N} \left(I_i(k) - I_j(k) \right)^2}, \tag{4}$$

where $I_i(k)$ represents the value of kth component of the model i. The distance can be measured in the space of original appearance models and/or constructed eigenspace. When the dimension of the original space is large, eigenspace is suitable for measuring distance between the models. The sequence of appearance models is determined by minimizing sum of the distances between consecutive appearances models. The evaluation function is as follows:

$$S = \sum_{i=1}^{N} D_{i,i+1}, \tag{5}$$

where N denotes the number of accumulated appearance models, and $i = 1, N+1$ indicate the identical appearance model, since the sequence is closed loop.

In a generated sequence, the difference between the consecutive appearance becomes small. This indicates that the pose of the object does not change drastically between the consecutive appearances. Thus the features in an appearance models can be easily made correspondence with the features in the consecutive appearance models. If correspondences of features are made, the change of object pose can be estimated.

5 Experiments

5.1 Evaluation of Successive Eigenspace Construction

We show the validity of successive construction of eigenspace. We use appearances of ridge line of a cube as an example of object appearances. We use 18 appearances obtained by rotation about vertical axis of a cube every 5 degree from 0 to 85 degree. **Figure 3** shows examples of appearance models which are 16×16 size image. The numbers in parentheses represent the rotation angles [deg.] of a cube. **Figure 4** shows projected points of 18 appearance models when the eigenspace is constructed simultaneously from the above 18 models. The dimension m of the eigenspace is set to 2. In the figure, the projected points corresponding consecutive appearance models are connected by line. We can see the appearances obtained from the similar viewpoints are projected on the near positions and the connected links compose an approximate circular ring. The average distance of consecutive projected points is 15.33 in the ideal eigenspace, and the minimum distance is 3.95 between the appearances from 35 [deg.] and 40 [deg.] viewpoints. We call this simultaneously constructed eigenspace the ideal eigenspace.

In order to compare the difference of constructed eigenspaces, we examine the difference of the distribution of projected points of the 18 appearance models of a cube, since the distance between appearances are measured on the eigenspace for object recognition. The distance between the eigenspaces is defined by the mean distance of corresponding projected points. The

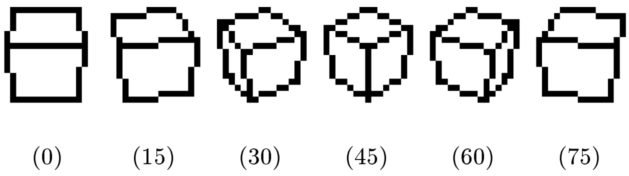


Figure 3. Edge appearances of a cube.

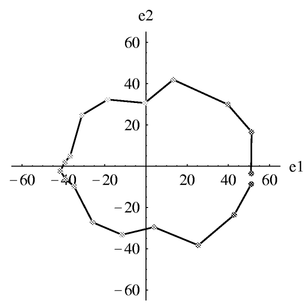
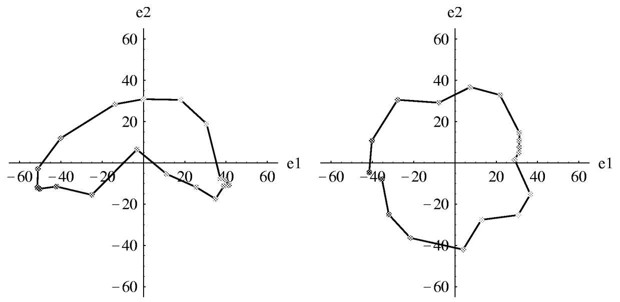


Figure 4. Ideal distribution of appearance models in eigenspace.

mean distance is calculated after translation for corresponding projected points: inversion, rotation by every 1.0 degree, and scaling by every 0.1 scale.

First, in order to show effectiveness of model selection in the process of successive eigenspace construction, we examine the distances between the ideal eigenspace and the successive constructed eigenspace in the cases with model selection and without selection. Observation order of object appearances is determined by random sampling without replacement of 18 models. Six trials are done for the two cases. Same random sampling order is given in each trial. **Table 1** shows the mean distances between the constructed eigenspace and the ideal eigenspace. The average distance in the case of construction without model selection is smaller than that of the case of construction with model selection. However, the distance of 6th trial without model selection is 12.75 which is the largest in the trials. The distribution of the projected points of 18 models in 6th trial is shown in **Fig.5**. Figure 5(a) shows an eigenspace constructed without model selection. Figure 5(b) shows an eigenspace constructed with model selection. The distribution of project points shown in Fig.5(a) is very different from that of the ideal eigenspace. This shows that there is a case that successively constructed eigenspace without model selection changes drastically depending on observation order. On the other hand, successively constructed eigenspace with model selection has no remarkable effect of obser-



(a) Successively constructed eigenspace without model selection

(b) Successively constructed eigenspace with model selection

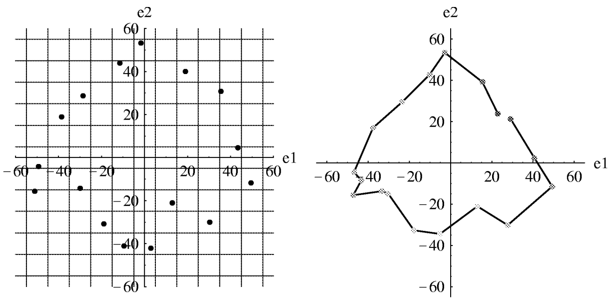
Figure 5. Successively constructed eigenspace in random sampling without replacement.

Table 1. Distances between eigenspaces constructed successively with model selection and without model selection.

random sampling	without model selection	with model selection
sampling 1	1.19	5.31
sampling 2	2.72	4.47
sampling 3	1.28	9.01
sampling 4	3.32	6.76
sampling 5	0.95	5.26
sampling 6	12.75	6.19
average	3.70	6.17

vation order comparing with Fig.4 and Fig.5(b).

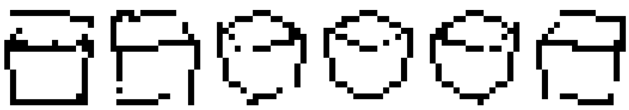
When robots wander the environment and acquire object appearances, the viewpoints are selected in random. In order to simulate this situation, we construct eigenspace successively by random sampling with replacement of the 18 appearances of the cube described in the previous subsection. The dimension m of the eigenspace is set to 2. **Figure 6(a)** shows the constructed eigenspace at $t = 54$. The black points indicate projected points on the eigenspace. We can see that 16 models are projected on different cells. Thus we can see that various appearances are stored on the eigenspace with a small number of stored models. **Figure 6**(b) shows the projected points of 18 models on the successively consructed eigenspace at $t = 54$. The



(a) Accumulated projected points of appearances on an eigenspace

(b) Successively constructed eigenspace

Figure 6. Successively constructed eigenspace in random sampling with replacement.



(0) (15) (30) (45) (60) (75)

Figure 7. Retrieved appearances from successively constructed eigenspace of a cube.

consecutive appearances are connected by line. From the figure, we can see that the consecutive appearances are projected to the close positions on the eigenspace and the connected links compose a approximate circular ring. In addition, we show the retrived appearance models from succesively constructed eigenspace. **Figure 7** shows the retrieved appearance models from the stored models at $t = 54$. The numbers in parentheses represent the rotation angles [deg.] of a cube. Though some vertical ridges of the cube are not retrieved comparing with original images of Fig.3, this is because the dimension of the eigenspace is too low; $m = 2$,$n = 256$. By comparing the retrieved images with the original images (Fig.3), we can conclude that the images retrieved from the successive constructed eigenspace keep the principal features of the cube.

As a result, the successively constructed eigenspace from random observation can be used in the same way of using conventional simultaneously constructed eigenspace.

(40) (55) (60)

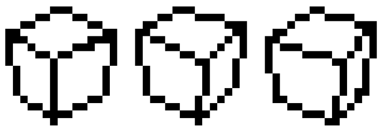


Figure 8. Three consecutive images in an appearance model sequence.

Table 2. Estimated Euler angles [deg.]

	$R_{g,h}$	$R_{h,i}$	$R_{i,g}$
α	8.75	-45.85	-6.10
β	19.41	6.07	-23.46
γ	-8.44	48.78	3.70

5.2 Generation of Appearance Model Sequence

In order to ascertain the efficiency of a method of generation of an appearance model sequence. we use the original 18 appearance models of a cube for generation of a seqecne. The evaluation function of eq.(5) is minimized by simulated annealing in the experiment. The generation algorithm can be applied to retrieved appearance models. The generated closed sequence of appearance models is (0) $\rightarrow$ (5) $\rightarrow$ (15) $\rightarrow$ (30) $\rightarrow$ (40) $\rightarrow$ (55) $\rightarrow$ (60) $\rightarrow$ (65) $\rightarrow$ (75) $\rightarrow$ (80). The numbers in the parentheses represent the rotation angles [deg.] of a cube . This shows that the pose of the cube changes consecutively in the generated sequence.

Next, we estimate a rotation matrix of a cube assuming a weak-perspective model as a projection model. **Figure 8** shows three consecutive appearance models in the generated closed sequence composed of 10 appearance models. We use seven corners of these three appearance models of a cube as features for correspondence to estimate rotation matrix $R_{i,j}$ [6]. Table 2 shows a set of estimated euler angles α, β, γ. A parameter β represents the rotation angle about Z axis. The true rotation parameters are $R_{g,h} = 15.0$, $R_{h,i} = 5.0$, and $R_{i,g} = -20.0$ [deg.]. The estimaton error is due to the low image resolution 16×16. We can see the object pose can be raughly estimated by conrresponding features of consecutive appearance models in the generated sequence.

6 Conclusions

We have proposed a method of successive acquisition of object appearances by active interaction with environment. The acquired appearances are stored into environmental-attached storage devices. Various appearances of the object are accumulated using successive eigenspace construction. We examine the different between the successively constructed eigenspace and simultaneously constructed eigenspace, and asertain the validity of the proposed method. We also proposed a method of generating a closed sequence of appearance models. The relative 3-D pose of the object is estimated using feature correspondences between the consecutive models in the sequence.

References

[1] Thomas von Numers et al., “An Intelligent Data Carrier System for Local Communication Between Cooperative Multiple Mobile Robots and Environment,” Preprints of 2nd IFAC conference on Intelligent Autonomous, Vehicles 95, pp.366 - 371, 1995.

[2] Fujii, T. et al., “Knowledge Sharing among Multiple Autonomous Mobile Robots through Indirect Communication using Intelligent Data Carrier,” *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1466-1471, 1996.

[3] Mae,Y., et al., “Relocation of Storage Devices Attached to Environment for Effective Knowledge Sharing among Multiple Robots”, Intelligent Autonomous Systems 6,pp.479-486,2000.

[4] Umetani,T., et al., “Adaptive Relocation of Environment-Attached Storage Devices by Multiple Robots”, Proc. of the 2001 IEEE International Conference on Robotics and Automation,pp.2998-3003,2001.

[5] Mae,Y., et al., “Object Recognition using Appearance Models Accumulated into Environments,” Proc. of 15th International Conference on Pattern Recognition, Vol.4, pp.845-848, 2000.

[6] Xu,G. and Sugimoto,N., ”A Linear Algorithm for Motion from Three Weak Perspective Images Using Euler Angles”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.21, No.1, pp.54-57, 1999.

[7] Turk,M.A. and Pentland,A.P., “Face recognition using eigenfaces,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.

[8] Murase,H. and Nayar,S.K., “Visual learning and recognition of 3-D objects from appearance”, International Journal of Computer Vision, Vol.14,pp.5-24,1995.

[9] Nayar,S.K., et al., “Subspace methods for robot vision”, IEEE Trans. Robotics and Automation, Vol.12, No.5, pp.750-758, 1996.