Planning of Multiple Camera Arrangement for Object Recognition in Parametric Eigenspace

Tomokazu Takahashi[1,2], Osanori Matsugano[1], Ichro Ide[1], Yoshito Mekada[3], Hiroshi Murase[1]
[1]*Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, Japan*
[2]*Japan Society for the Promotion Science*
[3]*Department of Life System Science and Technology, Chukyo University, Japan*
ttakahashi@murase.m.is.nagoya-u.ac.jp

Abstract

When objects are recognized by using multiple cameras, recognition rates strongly depend on the camera arrangement. In this paper, we propose a new method for planning a multiple camera arrangement for accurate recognition. We use a parametric eigenspace method for the recognition framework in which objects are represented as manifolds in an eigenspace. The proposed method evaluates the adequacy of camera arrangement according to the relations between the manifolds in the eigenspace. In the experiments, we defined a function that measures relations by the distances between manifolds. The experimental results show the effectiveness of the proposed method.

1. Introduction

The recognition of three-dimensional objects is an important technique in monitoring systems, human interfaces, and industrial applications. Among various approaches, we focus on appearance-based recognition because it has tolerance for high frequency noise. One widely used method is the parametric eigenspace method proposed by Murase and Nayar [1] that achieves both object recognition and parameter estimation, such as object poses and/or light source positions. Many works related to this method have been reported [2, 3, 5].

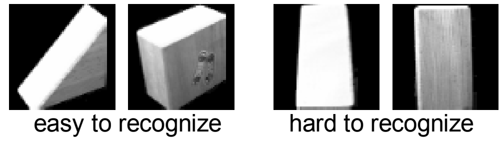
In appearance-based object recognition, two main factors cause performance degradation. The first is the quality degradation of input images depending on capturing conditions, as shown in Fig. 1. The second occurs when the viewpoint from which objects are captured is not suitable for distinguishing the object, as shown in Fig. 2. The use of multiple inputs can reduce these problems. For example, Yamaguchi et al. used multiple frames in a video sequence for robust face recognition [4]. Active recognition methods [5, 6] improved the accuracy of the results by adaptive camera repositioning through iterative recognition processes.

On the other hand, using prefixed multiple cameras for recognition is common. Selinger and Nelson used multiple cameras for object recognition in a cluttered scene [7]. Shakhnarovich recognized face and gait from multiple viewpoints [8]. In real world applications, however, since recognition accuracy generally depends on camera arrangement and object features, camera arrangement should be carefully planned. An optimal arrangement can also be found by experimentally evaluating the recognition rates for all possible arrangements. However, when searching for the arrangements, the number of recognition processes increases exponentially by the number of cameras.

Therefore, we propose a new method for planning a multiple camera arrangement for accurate object recognition. We utilize the parametric eigenspace method [1] for the recognition framework. Since the proposed method determines camera arrangement by using the relations between manifolds in an eigenspace, we can omit the recognition process that is searching for optimal camera arrangements. The remainder of this

(a) (b)

Figure 1: Degradation of image quality: (a) is clipped from a distant or wide-angle view image and (b) is zoomed image.



easy to recognize hard to recognize

Figure 2: Object appearances depending on viewpoint

paper is organized as follows. Section 2 describes a parametric eigenspace method using multiple cameras. Section 3 gives the details of the proposed method. Section 4 reports object recognition experiments with the proposed method. Results are discussed in Section 5. Then we summarize this paper in Section 6.

2. Object recognition with multiple cameras

Figure 3 shows the camera coordinate system used for this paper. Each camera position has two parameters, ϕ_m^h and ϕ_m^v, that represent the horizontal and vertical angles respectively. In this section, a parametric eigenspace method using multiple cameras is described.

2.1. Training

For each training image, feature vector $\hat{\mathbf{x}}$ with pixel values as its elements is normalized by $\mathbf{x} = \hat{\mathbf{x}} / \|\hat{\mathbf{x}}\|$, and then a matrix $\mathbf{X}$ is formed as:

$$\mathbf{X} = \left[\mathbf{x}_{1,1}^{(1)} - \mathbf{c}, \cdots, \mathbf{x}_{h,v}^{(p)} - \mathbf{c}, \cdots, \mathbf{x}_{H,V}^{(P)} - \mathbf{c}\right], \qquad (1)$$

where $p\,(=1, 2, ..., P)$ represents an object category, h and v are horizontal and vertical pose indicators, and $\mathbf{c}$ is the mean vector of $\mathbf{x}$ for all h, v, and p. A universal eigenspace is formed by eigenvectors $\mathbf{e}_i$ $(i = 1, 2, .., k)$ of $\mathbf{X}\mathbf{X}^T$ that correspond to the k-largest eigenvalues. Using Equation 2, feature vectors $\mathbf{x}_{h,v}^{(p)}$ are projected on points $\mathbf{g}_{h,v}^{(p)}$ in the eigenspace:

$$\mathbf{g}_{h,v}^{(p)} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]^T \left(\mathbf{x}_{h,v}^{(p)} - \mathbf{c}\right) \qquad (2).$$

For each object, manifold $G^{(p)}(\boldsymbol{\theta})$ is obtained from $\mathbf{g}_{h,v}^{(p)}$ by cubic spline interpolation. Here, $\boldsymbol{\theta} = \left(\theta^h, \theta^v\right)$ represents a pose parameter vector of an object: in other words, a camera position that captures the object.

In the same manner, for each object an eigenspace is formed and manifold $G_O^{(p)}(\boldsymbol{\theta})$ in the eigenspace is obtained. These manifolds are utilized in the pose estimation process.

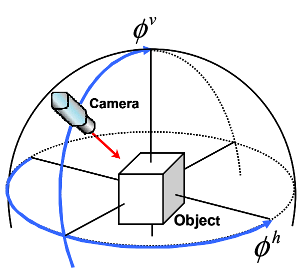


Figure 3: Camera coordinate system

2.2. Recognition with multiple cameras

For object recognition using M cameras whose positions are known, $m\,(=1, 2, ..., M)$-th camera position $\boldsymbol{\varphi}_m = \left(\phi_m^h, \phi_m^v\right)$ can be written as $\boldsymbol{\alpha}_m = \boldsymbol{\varphi}_m - \boldsymbol{\varphi}_1$, since input object pose $\tilde{\boldsymbol{\theta}}$ is unknown. Feature vector $\mathbf{y}_m$ from the m-th camera is obtained in the same manner as in the training stage and then projected on point $\mathbf{z}_m$ in the universal eigenspace as:

$$\mathbf{z}_m = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]^T (\mathbf{y}_m - \mathbf{c}) \qquad (3)$$

The distance between $\mathbf{z}_m$ and $G^{(p)}(\boldsymbol{\theta})$ is defined as:

$$d^{(p)}(\boldsymbol{\theta}) = \sum_{m=1}^{M} \left\| \mathbf{z}_m - G^{(p)}(\boldsymbol{\theta} + \boldsymbol{\alpha}_m) \right\| \qquad (4).$$

As recognition results, object category $\tilde{p}$ is obtained from the following equation:

$$\tilde{p} = \arg\min_{p} \left(\min_{\boldsymbol{\theta}} d^{(p)}(\boldsymbol{\theta}) \right) \qquad (5).$$

As pose estimation results according to the first camera position $\boldsymbol{\varphi}_1$, object pose $\tilde{\boldsymbol{\theta}}$ is obtained from Equation 6:

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} d_O^{(\tilde{p})}(\boldsymbol{\theta}), \qquad (6)$$

where $d_O^{(\tilde{p})}(\boldsymbol{\theta})$ represents a distance between projected points and the manifold in the object eigenspace of $\tilde{p}$. The distance is obtained by replacing $G^{(p)}(\boldsymbol{\theta})$ in Equation 4 by manifold $G_O^{(p)}(\boldsymbol{\theta})$.

3. Determination of multiple camera arrangements for accurate recognition

3.1. Camera arrangement determination in parametric eigenspace

For accurate object recognition, we propose a new method for determining a multiple camera arrangement. We believe that we can obtain information for planning a camera arrangement by referring to the relations between the manifolds in a universal eigenspace.

Camera arrangement $\tilde{\boldsymbol{\alpha}} = \{\tilde{\boldsymbol{\alpha}}_m\}$ represents a set of determined camera positions. The proposed method obtains $\tilde{\boldsymbol{\alpha}}$ from the following equation:

$$\tilde{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} F(G, \boldsymbol{\alpha}), \tag{7}$$

where $G = \{G^{(p)}(\boldsymbol{\theta})\}$ represents a set of manifolds and $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_m\}$ is an arbitrary camera arrangement. Therefore, we need to find an appropriate function $F(G, \boldsymbol{\alpha})$ to evaluate camera arrangement adequacy from the viewpoint of recognition accuracy.

3.2. Evaluation function based on distances between manifolds

In a universal eigenspace, some regions often exist where two different manifolds partially intersect or are located extremely close to each other. When input images are projected in such critical regions, recognition tends to fail. This is one main factor that affects performance degradation. Therefore, we define a function based on the distances between manifolds that actually effectively reduce such recognition failures; it is quite simple.

For a certain camera arrangement $\boldsymbol{\alpha}$, we define the distance between manifolds that correspond to objects p and $q\ (\neq p)$ as:

$$D^{(p,q)}(\boldsymbol{\alpha}) = \min_{\boldsymbol{\theta}} \frac{1}{M} \sum_{m=1}^{M} \left\| G^{(p)}(\boldsymbol{\theta} + \boldsymbol{\alpha}_m) - G^{(q)}(\boldsymbol{\theta} + \boldsymbol{\alpha}_m) \right\| \tag{8}.$$

Using this equation, we define the spread function of the manifolds as:

$$D(\boldsymbol{\alpha}) = \frac{1}{P} \sum_{p=1}^{P} \min_{q} D^{(p,q)}(\boldsymbol{\alpha}) \tag{9}.$$

This spread function can be considered an evaluation function, since it represents the simplicity of object distinction. Therefore, the $\tilde{\boldsymbol{\alpha}}$ that we are seeking is given by the following equation:

$$\tilde{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} D(\boldsymbol{\alpha}) \tag{10}.$$

Summarizing Equations 8, 9, and 10, we obtain the following equation:

$$\tilde{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \frac{1}{P} \sum_{p=1}^{P} \min_{q,\boldsymbol{\theta}} \frac{1}{M} \sum_{m=1}^{M} \left\| G^{(p)}(\boldsymbol{\theta} + \boldsymbol{\alpha}_m) - G^{(q)}(\boldsymbol{\theta} + \boldsymbol{\alpha}_m) \right\| \tag{11}.$$

4. Experiments

To demonstrate the proposed method's effectiveness, we evaluated recognition rates by using determined camera arrangement. In the experiments, arrangement adequacies were measured by Equation 9. We used 72 objects divided into four object sets, as shown in Fig. 4. The sets included: (1) FP with 27 football players, (2) CC with 15 cartoon characters, (3) WB with nine wooden blocks, and (4) FC with 21 human faces.

In the training stage, for each object we captured 252 training images at a size of 64 x 64 pixels by changing the camera position horizontally from 0 to 350º with 10º intervals and vertically from 0 to 90º with 15º intervals.

In the experiments, to approximately calculate the distance between manifolds needed for evaluating camera arrangement adequacy, we sampled 360 x 90 points on each manifold and measured the distances between them on two manifolds.

For the test images, camera positions were horizontally shifted 5º from the training stage. We captured test images from these camera positions at three different pixel resolutions: 64 x 64, 32 x 32, and 16 x 16, as shown in Fig. 5. For each pair of camera positions, we examined recognition rates and evaluated the camera arrangement adequacies given by Equation 9.

5. Results and discussion

Table 1 shows examples of FP camera arrangements in middle resolution cases and their recognition rates. We obtained camera arrangement (c) from the proposed method. Since the 98.9% recognition rate given by (c) was higher than the 67.3% and 87.6% given by other arrangements, (a) and (b), the proposed method improved the recognition rate. Fig. 6 shows the relation between FP recognition rates in the middle resolution case and the adequacies of camera arrangements. In Fig. 6, (a), (b), and (c) correspond to the camera arrangements in Table 1. The strong correlation shows that the proposed method worked effectively, even if such a simple function was used.

Table 2 shows the improvements of the recognition rate by the proposed method. Compared to inappropriate camera arrangements, the proposed method's results improved recognition rates up to 49.4% (from 25.0% to 74.4%) in low resolution FP cases.

FP CC WB FC

Figure 4: Examples of 72 objects used in experiments

High: 64 x 64 Middle: 32 x 32 Low: 16 x 16

Figure 5: Three levels of input image resolution

Table 1: Recognition rates of FP in middle resolution cases and their camera arrangements: (c) shows results of proposed method

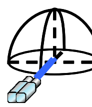
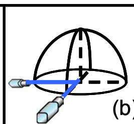
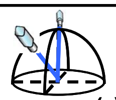
Arrangement	 (a)	 (b)	 **(c)**
Recognition rate [%]	67.3	87.6	**98.9**

Table 2: Improvement of recognition rates by using camera arrangement planned by proposed method

Resolution	Object set	Recognition rate [%]	
		Inappropriate arrangements	**Arrangement by proposed method**
High	FP	99.0	**100.0**
	CC	95.7	**99.3**
	WB	91.9	**99.4**
	FC	99.9	**100.0**
Middle	FP	67.3	**98.9**
	CC	70.6	**91.8**
	WB	69.6	**90.9**
	FC	31.9	**79.6**
Low	FP	25.0	**74.4**
	CC	31.7	**75.1**
	WB	55.4	**80.8**
	FC	12.2	**42.0**

6. Summary

For accurate object recognition using multiple cameras, we proposed a method for determining camera arrangement by using an evaluation function. We defined a simple function based on distances between manifolds. To evaluate the effectiveness of the proposed method, we conducted object recognition experiments with two cameras. The experimental results showed that the proposed method improved recognition rates up to 49.4% (from 25.0% to 74.4%) in low resolution FP cases.

Future works include the evaluation of the proposed method using more than three cameras as well as the development of better functions for evaluating camera arrangement adequacy.

Acknowledgements

Parts of this research were supported by the Grant-In-Aid for Scientific Research (16300054) and the 21st century COE program from the Ministry of Education, Culture, Sports, Science and Technology.

This work is developed based on MIST library (http://mist.suenaga.m.is.nagoya-u.ac.jp).

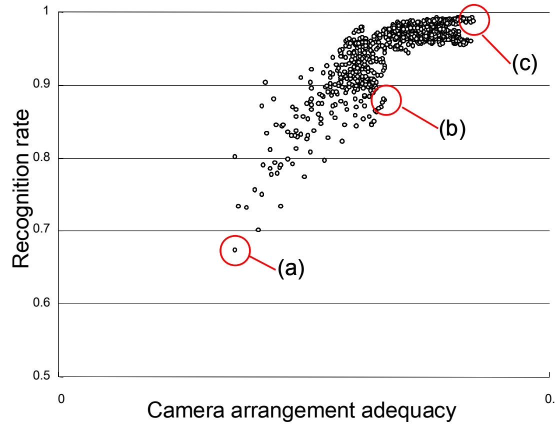


Figure 6: Relation between FP recognition rates in middle resolution case and camera arrangement adequacies: arrangements (a), (b), and (c) correspond to arrangements in Table 1

References

[1] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3-D Object Recognition from Appearance," *Internat. J. Comput. Vision*, vol. 14, pp. 5-24, 1995.

[2] H. Murase and S. K. Nayar, "Illumination Planning for Object Recognition using Parametric Eigenspaces," *IEEE Trans. PAMI*, vol. 16, no. 12, pp.1218-1227, 1994.

[3] H. Tanaka, I. Kitahara, H. Saito, H. Murase, K. Kogure, and N. Hagita, "Dynamic Visual Learning for People Identification with Sparsely Distributed Multiple Surveillance Cameras," *Proc. SCIA2005*, pp. 130-140, 2005.

[4] O. Yamaguchi and K. Fukui, "Smartface" – A Robust Face Recognition System under Varying Facial Pose and Expression," *IEICE Trans. Inf. Sys.*, vol. E86-D, no.1, pp. 37-44, 2003

[5] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Active Object Recognition in Parametric Eigenspace," *Proc. BMVC'98*, vol.2, pp. 629-638, 1998.

[6] F. G. Callari and F. P. Ferrie, "Autonomous Recognition: Driven by Ambiguity," *Proc. CVPR'96*, pp.701-707, 1996.

[7] A. Selinger and R. C. Nelson, "Appearance-based Object Recognition using Multiple Views," *Proc. CVPR2001*, vol. 1, pp. 905-911, 2001.

[8] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated Face and Gait Recognition from Multiple Views," *Proc. CVPR2001*, vol. 1, pp. 439-446, 2001.