SUPPORT VECTOR MACHINES
FOR OBJECT RECOGNITION UNDER VARYING ILLUMINATION CONDITIONS

Takahiro OKABE and Yoichi SATO

Institute of Industrial Science, University of Tokyo
4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan
{takahiro, ysato}@iis.u-tokyo.ac.jp

ABSTRACT

We propose an appearance-based method for object recognition under varying illumination conditions. It is known that images of an object under varying illumination conditions lie in a convex cone formed in the image space. In addition, variations due to changes in light intensity can be canceled by normalizing images. Based on these observations, our proposed method combines binary classifications using discriminant hyperplanes in the normalized image space. For obtaining these hyperplanes, we compared Support Vector Machine (SVM), which has been used successfully for object recognition under varying poses, and Fisher's linear discriminant (FLD). We have conducted a number of experiments by using the Yale Face Database B and confirmed that SVMs are effective also for object recognition under varying illumination conditions.

1. INTRODUCTION

The appearance of an object depends upon poses of the object, viewpoints, and illumination conditions. Therefore, object recognition based on computer vision suffers from three problems: variations due to changes in pose, viewpoint, and illumination. In this work, we focus on the problem of object recognition under varying illumination conditions, provided that both the pose and the viewpoint are fixed.

A gray image with D pixels can be represented by a D-dimensional vector, namely, a point in a D-dimensional space. In this paper, we refer to the D-dimensional space as the *image space*. An approach for object recognition that considers the image space itself as the feature space and applies algorithms of pattern recognition is called *appearance-based methods*.

It has been shown that images of an object under varying illumination conditions lie in a convex cone, called the *illumination cone*, formed in the image space when the superposition property of illumination is assumed [3]. The superposition property means that an image of an object taken under two light sources is represented by the addition of two images taken under each of the two light sources. However, previously proposed appearance-based methods [25, 14, 2, 23] do not sufficiently take into account that images of an object under varying illumination conditions lie in a convex cone.

In contrast to these methods, we propose an appearance-based method that utilizes properties of illumination cone in advance. Assuming that there is no intersection among illumination cones of objects, illumination cones of two arbitrary objects are separated by a hyperplane passing through the origin of the image space [2, 3]. In addition, variations due to changes in light intensity can be canceled by normalizing images. Based on these observations, our proposed method combines binary classifications using discriminant hyperplanes in the $(D-1)$-dimensional *normalized image space*.

For obtaining these hyperplanes, we compared two opposite paradigms experimentally by using the Yale Face Database B [8]. One is FLD [7] that takes the whole distribution of training patterns into consideration. Another paradigm is SVM [26] that pays attention only to training patterns near the discriminant boundary and is compatible with the superposition property of illumination in a geometrical sense. SVMs have been applied successfully to object/face recognition [18, 17, 9, 11]. As far as we know, however, no detailed study focusing on illumination variations has been reported until now. The present work is complementary to the work by Pontil and Verri [18] and by Heisele *et al.* [11] that demonstrates the effectiveness of SVMs for object recognition under varying poses. The main contribution of the present study is to demonstrate experimentally that SVMs are effective also for object recognition under varying illumination conditions.

2. PREVIOUS WORK

With regard to the problem of object recognition under varying illumination conditions, three different approaches have been developed: *feature-based methods*, appearance-based methods, and *generative methods*.

The first approach is based on features that are insensitive to changes in illumination. However, features such as edges and corners cannot always be extracted stably. Moreover, information essential for recognition may be lost by utilizing a part of images [5]. Recently, Chen *et al.* [6] have shown that the direction of image gradient is insensitive to changes in illumination direction.

On the other hand, the second approach applies algorithms of pattern recognition in which all pixel values are used as inputs. For instance, eigenfaces proposed by Turk and Pentland [25] compresses images by applying Principal Component Analysis (PCA). Murase and Nayar [14] represent a set of images of an object due to parametric changes in pose and/or illumination by using a manifold in the eigen-space. Belhumeur *et al*. [2] proposed a method called Fisherfaces in which images of objects under varying illumination conditions are projected to the low-dimensional subspace derived from Multiple Discriminant Analysis (MDA), and confirmed experimentally that the subspace derived from MDA is more effective than that derived from PCA.

However, these appearance-based methods do not sufficiently take into account that images of an object under varying illumination conditions lie in a convex cone. Moreover, in order to recognize an object under a certain imaging condition, appearance-based methods need training images of the object taken under the similar imaging conditions. Although the concept of virtual eigenspace proposed by Shakunaga and Shigenari [23] alleviates the latter limitation, their method does not take account of the convexity of the illumination cone.

The third approach called generative methods [24, 3, 8, 1] has been proposed recently in which a set of images of an object under varying illumination conditions is generated from a small number of training images of the object, assuming the Lambertian model. It has been shown that they are effective even when illumination conditions of test images differ greatly from those of training images.

However, these generative methods also have limitations. Because generative methods assume the Lambertian model, they cannot generate specular reflection components. Moreover, they cannot represent cast shadows accurately because cast shadows are high-frequency components. Therefore, they are applicable only when diffuse reflection components described by the Lambertian model and attached shadows are dominant in test images. To alleviate these limitations, Okabe and Sato [16] proposed a method that is robust against outliers such as specular reflection components and cast shadows.

In this paper, we propose an appearance-based method for object recognition under varying illumination conditions that takes account of properties of illumination cone. Our proposed method has distinct advantages over previously proposed methods. Compared with the gradient angle method [6], our method does not require a probability distribution of image gradient. Unlike the previously proposed appearance-based methods [25, 14, 2, 23], our method utilizes the information about the distribution of patterns in the feature space beforehand. Therefore, our method performs better and should be effective to some extent even when illumination conditions of test images differ greatly from those of training images. Furthermore, different from the generative methods [24, 3, 8, 1, 16], our method does not need sophisticated but complex modeling of reflection components under varying illumination conditions. Another important merit is that we do not assume the Lambertian model. Therefore, our method should be applicable to objects whose reflectance properties cannot be approximated by the Lambertian model.

3. PROPOSED METHOD

3.1. Illumination cone

To begin with, we describe variations due to changes in illumination on condition that both the pose and the viewpoint are fixed. Let $\boldsymbol{X}$ denote a D-dimensional vector that represents a gray image with D pixels. Let $\mathcal{C}$ denote a set of images of an object under arbitrary illumination conditions. When we denote images of the object taken under two arbitrary illumination conditions by $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ respectively, the superposition property of illumination means that $\boldsymbol{X}_1 + \boldsymbol{X}_2 \in \mathcal{C}$. Thus, the following relationships are derived.

$$s\boldsymbol{X}_1 \in \mathcal{C}, \quad \mathrm{t}\boldsymbol{X}_1 + (1-\mathrm{t})\boldsymbol{X}_2 \in \mathcal{C}, \tag{1}$$

where $s \geq 0$ and $0 \leq t \leq 1$. Therefore, $\mathcal{C}$ forms a convex cone, called the illumination cone, whose apex is at the origin of the image space [3].

It is known that the dimension of the illumination cone of a convex Lambertian object is equal to the number of distinct surface normals of the object [3]. However, it has been shown experimentally that face images under varying illumination conditions are approximately represented as linear combinations of a small number of basis images [10]. Moreover, it has been shown recently, based on the analysis in the frequency domain using spherical harmonics, that diffuse reflection components described by the Lambertian model can be approximated by a 4 to 9-dimensional subspace [1, 19]. Furthermore, specular reflection components described by, for instance, the Torrance-Sparrow model with large surface roughness also can be represented by a relatively low-dimensional subspace [20].

Therefore, it is expected that there is no intersection among illumination cones of objects, provided that they lie in low-dimensional subspaces and distribute independently in the high-dimensional image space. Even though they have intersections, the volume of the intersections is expected to be small. Henceforth, we assume that there is no intersection among illumination cones of objects. It has been shown that the illumination cone of a mirror-like or a highly-specular object cannot be represented approximately by a low-dimensional subspace [20]. Therefore, mirror-like and highly-specular objects are beyond the scope of our proposed method.

3.2. Image normalization

Equation (1) means that the norm of the vector $\boldsymbol{X}$ is proportional to light intensity. Actually, the norm is determined by the product of light intensity and surface albedo. Therefore, if intensity of light source is constant, it may be possible to utilize the norm of the vector as a clue for object recognition because the norm is proportional only to the surface albedo in this case. However, for object recognition under arbitrary illumination conditions, that is, when both the direction and intensity of light source change, it would be reasonable to cancel variations due to changes in light intensity by normalizing images.

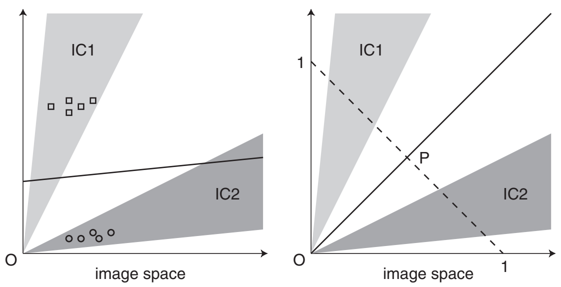

Fig. 1. Discriminant hyperplane determined by training patterns (left) and ideal one considering properties of convex cone (right).

In the present paper, we adopt the normalization by L_1 norm, $\tilde{\boldsymbol{X}} = \boldsymbol{X}/(\boldsymbol{X}^T\mathbf{1})$, where $\mathbf{1}$ is a D-dimensional vector in which all elements are equal to 1. The $(D-1)$-dimensional hyperplane defined by the above normalization is called the normalized image space [23]. It is worth noting that usual preprocess such as histogram equalization or zero-mean (and unit-variance) normalization does not preserve the separability of illumination cones in the image space.

3.3. Recognition based on superposition property

First, let us consider the problem of object recognition under varying illumination conditions in the D-dimensional image space. When there is no intersection among illumination cones of objects, illumination cones of two arbitrary objects, IC1 and IC2, are separated by a hyperplane passing through the origin of the image space (the line OP in the two-dimensional example of Figure 1) because they are convex cones whose apexes are at the origin of the image space [2, 3]. On the other hand, a certain illumination cone is not necessarily linearly separable from all other illumination cones. Therefore, the problem of object recognition under varying illumination conditions results in a combination of binary classifications using hyperplanes passing through the origin of the D-dimensional image space.

Secondly, we consider the same problem in the $(D-1)$-dimensional normalized image space. Because illumination cone is a convex cone, the cross section of the normalized image space through a certain illumination cone is also convex. Therefore, cross sections corresponding to two arbitrary objects are separated by a $(D-2)$-dimensional hyperplane in the normalized image space (the point P in the example of Figure 1). Consequently, the problem of object recognition under varying illumination conditions results in a combination of binary classifications using hyperplanes in the $(D-1)$-dimensional normalized image space.

4. BINARY CLASSIFICATION

4.1. Preliminaries

Let us consider patterns belonging to either of two classes, A and B. Let a vector $\boldsymbol{x}$ denote a pattern in the d-dimensional feature space. We represent a linear discriminant function as

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b, \tag{2}$$

where $\boldsymbol{w}$ and b are parameters called *weight vector* and *bias* respectively. Then, we classify $\boldsymbol{x}$ into A if $f(\boldsymbol{x}) > 0$, and classify it into B if $f(\boldsymbol{x}) < 0$. Therefore, the hyperplane defined by $f(\boldsymbol{x}) = 0$ becomes the discriminant hyperplane for the two classes. We denote the class to which a training pattern $\boldsymbol{x}_i$ $(i = 1, 2, \cdots, l)$ belongs by a label y_i $(i = 1, 2, \cdots, l)$, and assign $y_i = 1$ when $\boldsymbol{x}_i$ belongs to A, and $y_i = -1$ otherwise.

4.2. Fisher's linear discriminant

We denote the means of training patterns belonging to each class by $\boldsymbol{m}_A$, $\boldsymbol{m}_B$, and denote the within-class scatter matrix and the between-class scatter matrix by S_W, S_B respectively. FLD projects patterns to the one-dimensional axis $\boldsymbol{w}$ that maximizes the ratio of the between-class scatter to the within-class scatter

$$J(\boldsymbol{w}) = (\boldsymbol{w}^T S_B \boldsymbol{w})/(\boldsymbol{w}^T S_W \boldsymbol{w}), \tag{3}$$

and then classifies them on this axis by using the bias b. It is known that the maximization of the function $J(\boldsymbol{w})$ results in an eigenvalue problem and the solution is given by $\boldsymbol{w} = S_W^{-1}(\boldsymbol{m}_A - \boldsymbol{m}_B)$.

However, when $d \geq l$, that is, when the dimension of the feature space is larger than the number of training patterns, S_W becomes singular. In addition, it is known that a hyperplane separating all training patterns can be obtained by $1/2$ or more probability when $d \geq (l/2 - 1)$. Namely, the discriminant hyperplane is not overdetermined. The number of training patterns we used in experiments is much smaller than the dimension of the image space. Accordingly, in the experiments, we compressed d-dimensional feature space into d'-dimensional one by using PCA before applying FLD. The dimension d' is set to $([l/2 - 1] - 1)$ from our preliminary experiments[1]. We also set the bias as $b = -\boldsymbol{w}^T(\boldsymbol{m}_A + \boldsymbol{m}_B)/2$.

4.3. Support vector machine

We define a margin γ_i $(i = 1, 2, \cdots, l)$ of a training pattern $\boldsymbol{x}_i$ with respect to a hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ by $\gamma_i = y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)/||\boldsymbol{w}||$. When the training pattern $\boldsymbol{x}_i$ is correctly classified by the hyperplane, $\gamma_i > 0$ and the margin is equal to the distance between the pattern and the hyperplane. Then, we denote the minimum of these margins by γ and refer to it as the margin of the hyperplane with respect to all training patterns. SVM finds the discriminant hyperplane that maximizes the margin γ among an infinite number of separating hyperplanes.

In order to remove the redundancy that a hyperplane does not change when parameters $\boldsymbol{w}$ and b are multiplied by a certain scalar, we add the following constraints.

$$\min_{i=1,\cdots,l} |\boldsymbol{w}^T\boldsymbol{x}_i + b| = 1. \tag{4}$$

[1] [ ] is Gauss' notation: $[x] = n$ when $n \leq x < n+1$, where x and n are a real number and an integer respectively.

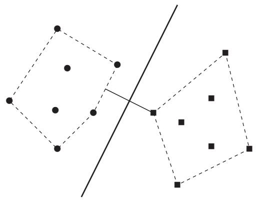

Fig. 2. Training patterns in the normalized image space and the hyperplane bisecting the closest points in two convex hulls.

Then, the maximization of the margin γ results in the optimization problem: minimizing $(\boldsymbol{w}^T\boldsymbol{w})$ subject to $y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1$.

We can transform this problem into the problem of finding the saddle point of the Lagrangian

$$L = \boldsymbol{w}^T\boldsymbol{w}/2 - \sum_{i=1}^{l} \alpha_i\{y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) - 1\}, \qquad (5)$$

by using the Lagrangian multipliers $\alpha_i \geq 0$ $(i = 1, 2, \cdots, l)$. Because $\partial L/\partial \boldsymbol{w} = \mathbf{0}$ and $\partial L/\partial b = 0$ at the saddle point,

$$\boldsymbol{w} = \sum_{i=1}^{l} y_i\alpha_i\boldsymbol{x}_i, \quad \sum_{i=1}^{l} y_i\alpha_i = 0 \qquad (6)$$

are derived. Substituting these equations into equation (5), the learning algorithm of SVM is summarized as the quadratic programming problem:

$$\begin{aligned} &maximize \quad && \sum_{i=1}^{l} \alpha_i - \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \boldsymbol{x}_i^T \boldsymbol{x}_j/2, \\ &subject\ to \quad && \sum_{i=1}^{l} y_i\alpha_i = 0, \quad \alpha_i \geq 0. \end{aligned} \qquad (7)$$

Therefore, the discriminant hyperplane is determined by the optimization calculation without falling into local maximum. In addition, the solution of the above problem has to satisfy the complementarity conditions

$$\alpha_i\{y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) - 1\} = 0. \qquad (8)$$

From equations (4) and (8), $\alpha_i \neq 0$ only for training patterns that are closest to the discriminant hyperplane. Therefore, from equation (6), the discriminant hyperplane is determined only by these closest training patterns called *support vectors*.

It is known that the performance of SVM is characterized by the number of support vectors. Consequently, it is expected that SVM has outstanding generalization ability even when the dimension of the feature space is much larger than the number of training patterns. From our preliminary experiments, we confirmed that SVMs are effective even in high-dimensional feature spaces. Accordingly, we applied SVMs without dimensionality reduction in our experiments.

It is worth noting that SVM is compatible with the superposition property of illumination in a geometrical sense. Assuming the superposition property, convex combinations of training images of an object are also images of the object. Therefore, discriminant hyperplanes should be determined by not training patterns but convex combinations, that is, a *convex hull* of training patterns (Figure 2). Among an infinite number of separating hyperplanes, let us consider the hyperplane bisecting the closest points in two convex hulls (Figure 2). This hyperplane is appropriate for classification from a geometrical point of view because it is robust against perturbations of the training patterns. Interestingly, it has been shown, based on the duality, that the hyperplane bisecting the closest points is identical to the optimal hyperplane determined by SVM [4].

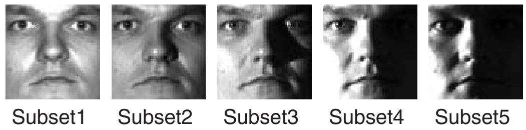

Fig. 3. Images of an individual belonging to each subset: the angle θ between the light source direction and the optical axis lie $[0°, 12°)$, $[20°, 25°]$, $[35°, 52°)$, $(60°, 77°)$, and $(85°, 128°)$ respectively.

5. EXPERIMENTAL RESULTS

In order to demonstrate the effectiveness of SVMs for object recognition under varying illumination conditions, we have conducted a number of experiments by using the Yale Face Database B [8]. It is known that face recognition under varying illumination conditions is a very difficult task because variations due to changes in illumination are usually larger than those due to changes in face identity [13]. The performance of generative methods for face recognition [8, 1] implies that the assumption that there is no intersection among illumination cones is approximately satisfied.

5.1. Face image database

The database consists of images of 10 individuals in 9 poses acquired under 64 different point light sources and an ambient light: 5850 images in total. The coordinates of the left eye, right eye, and mouth are appended for images in the frontal pose, and the coordinate of the face center is appended for images in other poses. Each image is assigned to one of 5 subsets according to the angle θ between the direction of the light source and the optical axis of a camera.

In Section 5.2 and 5.3, we conducted experiments by using 650 images in the frontal pose. In Section 5.4, we used 3250 images in 5 poses at 12 degrees from the optical axis. In the experiments, these images were cropped and down-sampled to 64×64 pixels by averaging. Actually, in order to remove any bias due to the scale and position of a face in each image from the recognition performance, they were aligned so that the locations of the eyes or the face center were the same. In addition, the image taken under the ambient light was subtracted from them.

In Figure 3, we show images of an individual belonging to each subset. One can confirm that images vary significantly depending on the direction of the light source.

5.2. Extrapolation of illumination direction

We compared four algorithms: Nearest-Neighbor rule (NN), EigenFaces (EF), FLD, and SVM. In EF, we compressed the feature space in the same way as FLD, and then applied nearest neighbor rule. For combining binary classifications in FLD and SVM, we adopted the tournament rule. In addition, we used the library for SVM [22]. We conducted two experiments for each algorithm: one using patterns in the $4096(= 64\times64)$-dimensional image space and the other using patterns in the normalized image space.

First, we used 7 images belonging to subset 1 ($\theta < 12^\circ$) as training images of each individual, and then tested other images ($\theta \geq 20^\circ$). Namely, we conducted experiments on extrapolation of illumination direction. In Table 1, we show the recognition error rates of eight methods for each subset. The index attached to each algorithm represents the feature space used in the method; 1 and 2 correspond to the image space and the normalized image space respectively. Therefore, FLD-2 and SVM-2 are methods that take account of both the convex and cone properties, while other methods consider only the convex property (FLD-1 and SVM-1), or only the cone property (NN-2 and EF-2), or neither of them (NN-1 and EF-1).

Overall, these appearance-based methods break down to images in subset 5 because error rates are close to the error rate obtained by chance (90%). Then, comparing FLD-1 and FLD-2, or SVM-1 and SVM-2, one can find that the recognition performance is significantly improved by using normalized images. This is because FLD-1 and SVM-1 cannot necessarily find separating hyperplanes passing through the origin of the image space while FLD-2 and SVM-2 utilize the properties of a convex cone in advance. Moreover, SVM-2 is superior to FLD-2 and other methods. Therefore, we can conclude that a combination of SVMs and the normalized image space is effective for extrapolation of illumination direction.

5.3. Interpolation of illumination direction

Secondly, we used 26 images belonging to subset 1 ($\theta < 12^\circ$) and 5 ($\theta > 85^\circ$) as training images of each individual, and then tested other images ($20^\circ \leq \theta < 77^\circ$). Namely, we conducted experiments on interpolation of illumination direction. We show the recognition error rates in Table 2. Also in this case, we can confirm that SVM-2 taking account of the properties of a convex cone outperforms other methods. Furthermore, the performance of SVM-2 is comparable to that of generative methods [8]. This shows that complex modeling of reflection components is not always necessary when a variety of training images are prepared. Lee *et al*. [12] and Sato *et al.* [21] discussed a set of lighting directions suitable for taking training images.

It is interesting that the performance of EF and FLD becomes worse by using normalized images, different from the case of the extrapolation experiments. The reason is considered as follows. Because images in the database are taken under light sources with almost the same intensity, the norm of each image is closely related to the surface albedo of the face. Therefore, there is a possibility that the difference in norm is emphasized by dimensionality reduction, and then work effectively for recognition.

Table 1. Extrapolation of illumination direction on the condition of fixed frontal pose.

Error rate (%): extrapolation, frontal pose					
Method	Dimension	2	3	4	5
NN-1	4096	5.1	50.8	81.2	85.2
NN-2	4095	0	7.6	56.5	77.2
EF-1	$4096 \to 5$	25.4	72.9	84.8	86.2
EF-2	$4095 \to 5$	7.6	47.5	73.2	78.8
FLD-1	$4096 \to 5$	4.2	37.3	71.7	85.7
FLD-2	$4095 \to 5$	0	13.6	60.9	88.9
SVM-1	4096	2.5	22.9	75.4	87.3
SVM-2	4095	0	0	36.2	88.4

Table 2. Interpolation of illumination direction on the condition of fixed frontal pose.

Error rate (%): interpolation, frontal pose				
Method	Dimension	2	3	4
NN-1	4096	5.1	44.9	16.7
NN-2	4095	0	8.5	13.8
EF-1	$4096 \to 24$	6.8	55.1	32.6
EF-2	$4095 \to 24$	18.6	67.8	51.4
FLD-1	$4096 \to 24$	1.7	15.3	17.4
FLD-2	$4095 \to 24$	5.1	23.7	40.6
SVM-1	4096	0.8	11.0	8.0
SVM-2	4095	0	0	4.3

5.4. Variations in illumination and pose

So far, we assume that images in a certain pose are cropped correctly in advance. However, this assumption may not be hold strictly in practice. Therefore, we conducted experiments by using images taken under varying illumination conditions and poses, and confirmed the robustness of our proposed method against variations due to slight changes in pose. In this section, we used images in 5 poses shown in Figure 4 instead of images in the frontal pose, and conducted extrapolation and interpolation experiments.

A set of images of an object under varying illumination conditions and poses can be represented by a union of illumination cones because variations in illumination for each pose are represented by an illumination cone. Therefore, the image set is not necessarily convex although it is still a cone whose apex is at the origin of the image space.

We show the results of interpolation experiment in Table 3. Comparing FLD-1 and FLD-2, or SVM-1 and SVM-2, one can confirm that FLD-2 and SVM-2 are superior to FLD-1 and SVM-1, that is, discriminant hyperplanes in the normalized image space are more effective than those in the image space. This is because the optimal hyperplane in the image space does not necessarily pass through the origin of the image space. In addition, SVM-2 is superior to all other methods, and its performance is almost equal to that in the case of fixed pose. We consider that this is because the reflectance property of faces can be approximated by the Lambertian model. Namely, when an illumination cone of an object is represented approximately by a low-dimensional subspace, a set of images of the object under varying illumination conditions and poses

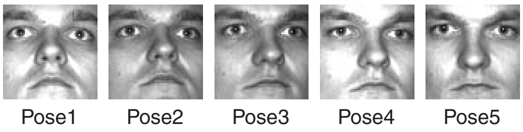

Fig. 4. Images of an individual in 5 poses.

Table 3. Interpolation of illumination direction under varying poses.

Error rate (%): interpolation, 5 poses				
Method	Dimension	2	3	4
NN-1	4096	3.5	32.2	13.2
NN-2	4095	0	14.5	15.1
EF-1	4096 $\rightarrow$ 128	3.8	34.5	17.4
EF-2	4095 $\rightarrow$ 128	0.2	27.8	22.0
FLD-1	4096 $\rightarrow$ 128	3.0	21.5	12.8
FLD-2	4095 $\rightarrow$ 128	0.5	2.5	8.7
SVM-1	4096	2.3	15.7	14.5
SVM-2	4095	0	0.5	4.2

can also be approximated by a relatively low-dimensional subspace. Therefore, it is expected that image sets of objects are separable in the high-dimensional image space.

In the case of extrapolation experiments, we obtained the similar results (see [15]).

6. CONCLUSIONS AND FUTURE WORK

We discussed the problem of object recognition under varying illumination conditions. The contribution of the present study is summarized in the following four points. We showed that (i) the problem results in a combination of binary classifications using hyperplanes in the $(D-1)$-dimensional normalized image space. We confirmed experimentally that (ii) the use of hyperplanes in the normalized image space is effective, and that (iii) SVMs are better suited than FLDs for obtaining discriminant hyperplanes. In addition, we showed experimentally that (iv) a combination of linear SVMs and the normalized image space works well even when variations due to slight changes in pose accompany those due to changes in illumination.

In the present study, we have conducted experiments by using face images. However, our proposed method should be applicable to non-Lambertian objects when illumination cones of objects are approximated by low-dimensional subspaces. Therefore, in the future work, we plan to confirm the effectiveness of our method for objects with various reflectance properties.

Acknowledgements
A part of this work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 13224051).

7. REFERENCES

[1] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces", In *Proc. IEEE ICCV 2001*, pp.383–390, 2001.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. PAMI*, 19(7), pp.711–720, 1997.

[3] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions?", *Int'l. J. Computer Vision*, 28(3), pp.245–260, 1998.

[4] K. Bennett and E. Bredensteiner, "Duality and geometry in SVM classifiers", In *Proc. Int'l. Conf. Machine Learning (ICML 2000)*, pp.65–72, 2000.

[5] R. Brunelli and T. Poggio, "Face recognition: features versus templates", *IEEE Trans. PAMI*, 15(10), pp.1042–1052, 1993.

[6] H. Chen, P. Belhumeur, and D. Jacobs, "In search of illumination invariants", In *Proc. IEEE CVPR 2000*, pp.254–261, 2000.

[7] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.

[8] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose", *IEEE Trans. PAMI*, 23(6), pp.643–660, 2001.

[9] G. Guo, S. Li, and K. Chan, "Face recognition by support vector machines", In *Proc. IEEE FG 2000*, pp.195–201, 2000.

[10] P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions", In *Proc. IEEE CVPR '94*, pp.995–999, 1994.

[11] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: global versus component-based approach", In *Proc. IEEE ICCV 2001*, pp.688–694, 2001.

[12] K.-C. Lee, J. Ho, and D. Kriegman, "Nine points of light: acquiring subspaces for face recognition under variable lighting", In *Proc. IEEE CVPR 2001*, 1, pp.519–526, 2001.

[13] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction", In *Proc. ECCV '94*, pp.286–296, 1994.

[14] H. Murase and S. Nayar, "Visual learning and recognition of 3-D objects from appearance", *Int'l. J. Computer Vision*, 14(1), pp.5–24, 1995.

[15] T. Okabe and Y. Sato, "Support vector machines for object recognition under varying illumination", *IPSJ Trans. Computer Vision and Image Media*, 44(SIG 5), pp.22–29, 2003. (in Japanese)

[16] T. Okabe and Y. Sato, "Object recognition based on photometric alignment using RANSAC", In *Proc. IEEE CVPR 2003*, 1, pp.221–228, 2003.

[17] P. Phillips, "Support vector machines applied to face recognition", *Advances in Neural Information Processing Systems 11*, pp.803–809, 1998.

[18] M. Pontil and A. Verri, "Support vector machines for 3D object recognition", *IEEE Trans. PAMI*, 20(6), pp.637–646, 1998.

[19] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object", *J. Opt. Soc. Am. A*, 18(10), pp.2448–2459, 2001.

[20] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering", In *Proc. ACM SIGGRAPH 2001*, pp.117–128, 2001.

[21] I. Sato, T. Okabe, Y. Sato, and K. Ikeuchi, "Appearance sampling for obtaining a set of basis images for variable illumination", In *Proc. IEEE ICCV 2003*, pp.800–807, 2003.

[22] C. Saunders, M. Stitson, J. Weston, L. Bottou, B. Schoelkopf, and A. Smola, "Support vector machine reference manual", Technical Report CSD-TR-98-03, Department of Computer Science, Royal Holloway, University of London, Egham, UK, 1998.

[23] T. Shakunaga and K. Shigenari, "Decomposed eigenface for face recognition under various lighting conditions", In *Proc. IEEE CVPR 2001*, 1, pp.864–871, 2001.

[24] A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image", *Int'l. J. Computer Vision*, 21(1/2), pp.99–122, 1997.

[25] M. Turk and A. Pentland, "Face recognition using eigenfaces", In *Proc. IEEE CVPR '91*, pp.586–591, 1991.

[26] V. Vapnik, *Statistical learning theory*, John Wiley & Sons, New York, 1998.