Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-5-2003

Subspace-based face recognition: outlier detection and a new distance criterion

P. Chen and D. Suter

MONASH UNIVERSITY

Subspace-based face recognition: outlier detection and a new distance criterion

Pei Chen and David Suter
Dept ECSE

Abstract: Illumination effects, including shadows and varying lighting, makes the problem of face recognition challenging. Experimental and theoretical results show that the face images under different illumination conditions lie in a low-dimensional subspace, hence principal component analysis (PCA) or low-dimensional subspace techniques have been used. Following this spirit, we propose new techniques for the face recognition problem, including an *outlier detection strategy* (mainly for those points not following the Lambertian reflectance model), and a new Bayesian-based error criterion for the recognition algorithm. Experiments using the Yale-B face database show the effectiveness of the new strategies.

Keywords: Face recognition, Linear subspace, Principal component analysis, Illumination effect.

1. Introduction

Illumination effects make the problem of face recognition very challenging [4, 5, 9, 10, 13, 21]. It has been observed that "the variations between the images of the same face due to illumination and viewing direction are almost larger than image variations due to change in face identity"[15]. A robust recognition approach should be able to overcome this difficulty.

In order to tackle this issue [23], PCA has been utilized to model the lighting variation in images. It has been proved, experimentally [8, 12, 16, 17, 18] and theoretically [1, 2, 19, 20], that the possible images approximately concentrate in a low-dimensional subspace. although the dimension of the image set for an object is actually "equal to the number of distinct surface normals" [5]. Experimental observations [8, 12, 25] have helped firmly establish that the images of the same face, produced under different lighting conditions, approximately lie in a low-dimensional subspace. Another influential example of the power of such an approach was the SLAM system [16, 17, 18], which captured the variations due to pose and illumination by a 20-dimensional or less subspace – extending the applications to object recognition and pose determination etc. Recently, it was proved, by using spherical harmonics, that "all Lambertian reflectance functions obtained with arbitrary distant sources lie in close to a 9D linear subspace": Basri and Jacobs [1, 2] and Ramamoorthi and Hanrahan [19, 20].

Another variant of PCA-based face recognition is the linear subspace approach (section 2). For a Lambertian object, its images approximately lie in a 3D linear subspace if there is neither attached shadow nor cast shadow. This fact implies that, given three images for a Lambertian object, its any image can be a linear combination of these three generic images. This is the well-known photometric stereo method [21]. Moreover, it also suggests a simple, but effective, approach to the problem of face recognition. This subspace-based approach consists of two steps: the off-line training stage and the on-line recognition stage [4, 9]. In the training stage, obtain three basis images from three or more images; and in the recognition stage, simply compute the distance of the new image to each face basis and choose the identity that has the minimal distance.

In addition to the traditional work on the PCA-based face recognition [8, 12, 23], two other studies should be noted: those involving the illumination cone [5, 9, 10] and the related theoretical development [1, 2, 19, 20]. The illumination cone has been proved to be effective in modeling the lighting and pose effect, on the Yale-B face database. The theory of the low-dimensional subspace of the Lambertian objects not only explains the observed phenomenon [1, 2, 19, 20], but also provides new insights on the problem of face recognition [14].

The contributions of this paper are: (1) in section 3, we propose a new Bayesian distance for the subspace-based recognition problem, this is based on a new theory about the subspace's *learning* capacity [7]. (2), in order to remove points not following the reflectance model, we employ the iterative reweighted least square (IRLS) technique (section 4) to detect the pixels that do not obey the dimension-3 constraint, such as eyeballs. The experiments on the Yale-B face database show the effectiveness of the new techniques.

2. Lambertian Reflection

2.1. Lambertian reflectance and 3-dimensional subspace

The images of a Lambertian object can be approximately modeled by a 3-dimensional subspace if the light source lies at infinity and there is neither attached shadow nor cast shadow [9, 21]. Following [5], for any point p on a Lambertian surface, illuminated by an infinite light source, its intensity can be described by

$$I(p) = a(p)\mathbf{n}(p)^T \mathbf{s} = \mathbf{b}(p)^T \mathbf{s} \qquad (1)$$

where $a(p)$ (a scalar) is the albedo at position p, $\mathbf{n}(p)$ (a 3-vector) is the inward normal of the surface at position p, and $\mathbf{s}$ (a 3-vector) is the direction of the light. Let $\mathbf{B} \in R^{n,3}$ be a matrix where each row is $\mathbf{b}(p)^T$. The illumination subspace can be generated by:

$$L = \{x \mid x = \mathbf{Bs}, \forall \mathbf{s} \in R^3\} \qquad (2)$$

The images without shadows are a subset of L. The set of all images, the non-negative orthant, is defined as:

$$L_0 = \{x \mid x = \max(\mathbf{Bs},0), \forall \mathbf{s} \in R^3\} \qquad (3)$$

A general subspace-based algorithm for the face recognition is [9]:

(a) Training stage. Arrange the training samples as the training matrix, each column of which is an image. By SVD [11], the 3 basis images are calculated as the 3 singular vectors that correspond to the 3 largest singular values.

(b) Recognition stage. Calculate the distance of the test image to the 3-dimensional subspace that is spanned by the 3 basis images. The target is selected as that who has the shortest distance.

2.2. Attached shadow and low-dimensional subspace

The 3-dimensional constraint does not hold when there is a shadow. Intrinsically, the dimension of the image set for an object is “equal to the number of distinct surface normals”. However, it has been proved, experimentally and theoretically, that the image set approximately lies in a low-dimensional subspace [8, 12, 23].

However, an important theoretical proof exists that shows that the images of a Lambertian object can be approximately modeled by a 9-dimensional subspace if there is no cast shadow [1, 2, 19, 20]. Moreover, it has been experimentally proved that images with shadow can be approximately modeled by 5±2 eigenimages [8]. Based on this 9-dimension theory, 9 points of light for face recognition were optimally determined [14].

2.3. Generation of the image basis from synthetic images

One does not want to use more images then necessary in constructing a training set. It has been shown [14] that for a single face approximately 9 well-chosen lighting directions are optimal. However, the result in [14] was not good enough in practice. A reliable approach to obtain the image basis is to calculate them from a large amount of training images, for example 80-120 training images [9]. Although a large set of images is unwieldy, a possible solution to this problem is to use the synthetic images, as in [9]. In this paper, we also employ this strategy to obtain the image basis.

The procedure of generating the image basis in the training stage is (taking the Yale-B face database as an example):

(a). Obtain the illumination subspace L in (2) from more than 2 images that have no shadow. Here, we use the 7 images in “subset 1” as the training samples.

(b). Generate the synthetic images that are illuminated by a light at infinity, as in (3).

(c). Calculate the approximate low-dimensional subspace from the simulated images, by SVD [11].

As an important variation, we employ the iterative least squares procedure as an “outlier” detection strategy when we calculate the illumination subspace in sub-step (a), because not all the pixels of a face can be approximately Lambertian, for example the eyeballs and eyebrows. This *outlier detection strategy* will be presented in section 4.

In the simulation of the possible images in sub-step (b), we only consider the effect of the attached shadow, as the “Cones-attached” in [9]. However, we don’t need to reconstruct the face to a generalized Bas-Relief (GBR) transformation [3]. Instead, we “randomly” generate the synthetic images, because an arbitrary “linear” combination of the three basis images can be an image, illuminated by a light with unknown direction [9]. It should be noted that the negative pixels in the synthetic images have to be set as zeroes. The attached shadow can be modeled this way, while the cast shadow cannot be modeled. Although the direction of the light $\mathbf{s}$ is randomly generated, we set the energy of first basis image is half as that of the other two basis images, in order to model the shadow effect better.

In calculating the basis images in sub-step (c), which accounts for the attached shadow, we find that the 7-dimensional subspace performs slightly better than the 9-dimensional subspace. Crucially, in our approach, the "outliers" detected in sub-step (a) are not included in calculating the distances of the test image to the 7-dimensional subspaces. Consequently, we compare the angles between the test image and the 7-dimensional subspaces, because different identities have different outliers.

3. Learning capacity of low-dimensional subspace and a new Bayesian distance

Few people have properly estimated where the noise in the learning and recognition processes reside. In [7], based on the matrix perturbation theory [22, 24], the learning capacity of the low-dimensional linear subspace has been studied. The theory states that the distance of a new test vector to the estimated low-dimensional subspace comes from two sources: one is from the learning samples and another from the test vector. More formally, it is described by the following result [7].

Result (*Learning capacity of LSA***)**: *For a rank-r LSA-based recognition system, the "error measure" (the SSD) comes from two independent sources: the noise in the basis images (eg. quantization and model error) and the noise in the test image. Specifically, the SSD is (see eq. 38, [7]):*

$$(m-r)\sigma_t^2 + (m-r)\sigma_l^2 \sum_{i=1}^{r} \frac{f_i^2}{\kappa_i^2} \qquad (4)$$

where m is the dimension of the object, σ_t and σ_l are the noise levels for the test image and the learning samples respectively, and κ_i $(1 \le i \le r)$ is the i^{th} singular value of the training matrix and f_i $(1 \le i \le r)$ is the magnitude of the i^{th} component of the test image.

From the result, we can see that some error is introduced by the noise in the training samples. Thus, this part of the error in (4) should be subtracted in the recognition stage. More formally, suppose the new test image has a distance of d to the r-dimensional subspace. From (4), we take the following distance as the criterion for the classification:

$$\sqrt{\max(d^2 - (m-r)\sigma_l^2 \sum_{i=1}^{r} \frac{f_i^2}{\kappa_i^2},\ 0)} \qquad (5)$$

The estimation of the noise level σ_l in the learning samples will be discussed in section 4.

4. Non-Lambertian pixel detection

Although the human faces can be approximately modeled as Lambertian, some part are obviously non-Lambertian, for example the eyeballs and eyebrows. Moreover, some parts of the true training samples that are in the shadow don't obey the 3-dimension constraint. In order to obtain an accurate 3-dimensional illumination subspace, we should exclude these abnormal pixels.

Here, we employ a variant of the iterative reweighted least square (IRLS) as the "outlier" detection strategy: the weight is either 1 or 0. More specifically, we retain those data whose residual is less than 3 times of the noise scale and prune the other data. Thus, a general 1/0 IRLS iteratively works this way:
(i) to estimate the scale from the residual of the retained data.
(ii) if there is some "outliers", whose residual is larger than 3 times of the scale, to prune these data and go to sub-step (i); else, terminate the iteration.

Because we work on low-dimensional subspaces, the general 1/0 IRLS can not be directly applied to detect the non-Lambertian pixels. Particularly, we have to define the residual for a pixel, which in fact is a n-dimensional vector if we work on n training images.

Suppose an $m \times n$ training matrix consists of n training images, each of which has m pixels. First, calculate the r-dimensional subspace by SVD [11]. Second, calculate the residual matrix, by subtracting the r largest components from each column. Third, calculate the 2-norm of all row vectors and regard them as the residual for the corresponding pixels. The scale can be estimated as the root mean square of the residuals of the retained pixels. The detected mask for the non-Lambertian pixels are displayed in fig 3, where the black pixels denote the non-Lambertian ones.

4.1. Noise level estimation

In this section, we explain how to estimate the noise level σ_l in (5). In our synthetic generation of the image basis, we have to first estimate the noise level in the actual images. Then, to use these estimates to calculate the estimates for the noise levels in the synthetic images (generated by a linear combination of basis images, as per (b) step in section 2.3).

From [6], the maximal likelihood (ML) estimate of the noise level in an r-dimensional illumination subspace is as follows:

$$\sqrt{\frac{1}{m(m-r)}\sum_{i=r+1}^{m}\kappa_i^2} \tag{6}$$

where κ_i is the i^{th} singular value of the actual training matrix. It should be noted that the estimate in (6) is calculated from the outlier-detected training matrix.

We calculate the total noise energy of the synthetic training matrix and regard the root mean of the energy as the noise level.

5. Experimental results

In this section, we report our results, comparing with that in [9, 14]. As in [9], we also do the face recognition experiments on the Yale-B face database, which consists of 10 people. We follow [9] in cropping, centering and resizing the images. The 10 people are shown in fig. 1.

Here, we only study the illumination effect on the recognition problem. Thus, only the frontal pose, at which 64 pictures were taken for each people, is used. These 64 images are divided into 5 subsets of 7/12/12/14/19 pictures. Two images are shown in fig. 2 for each subset. From “subset 1” to “subset 5”, there is more and more shadow the pictures. In fact, the pictures in “subset 5” are almost indiscernible, as shown in fig, and no result has been reported on this subset. By employing the new strategies, we obtain a good performance on this subset, up to 92.1% of correct rate.

It should be noted that, the misclassification rate of *9PL* [14] for "subset 4" of the database was 5.6%. Because 7 training images were from "subset 4", we should not include these 7 images in the test image set when we calculate the misclassification rate.

Table 1: Comparison of the error classification rate on Yale-B face database.

Method	Subset 1-3	Subset 4	Subset 5
Linear subspace [9]	**0**	**15**	/
Cones-attached [9]	**0**	**8.6**	/
Cones-cast [9]	**0**	**0**	/
9PL [14]	**0**	**2.8(5.6)**	/
Proposed	**0**	**0**	**7.9**

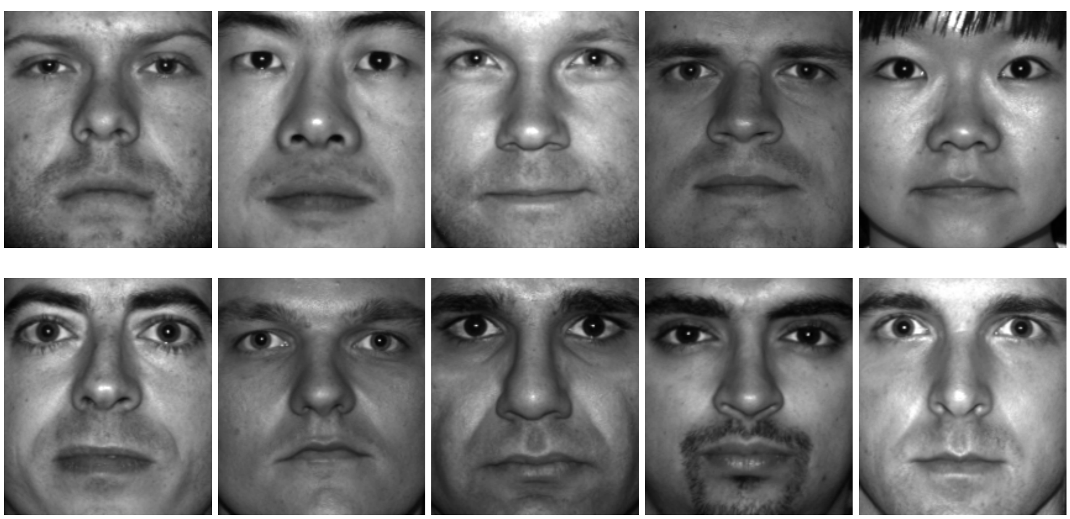

Figure 1: 10 people in Yale-B face database.

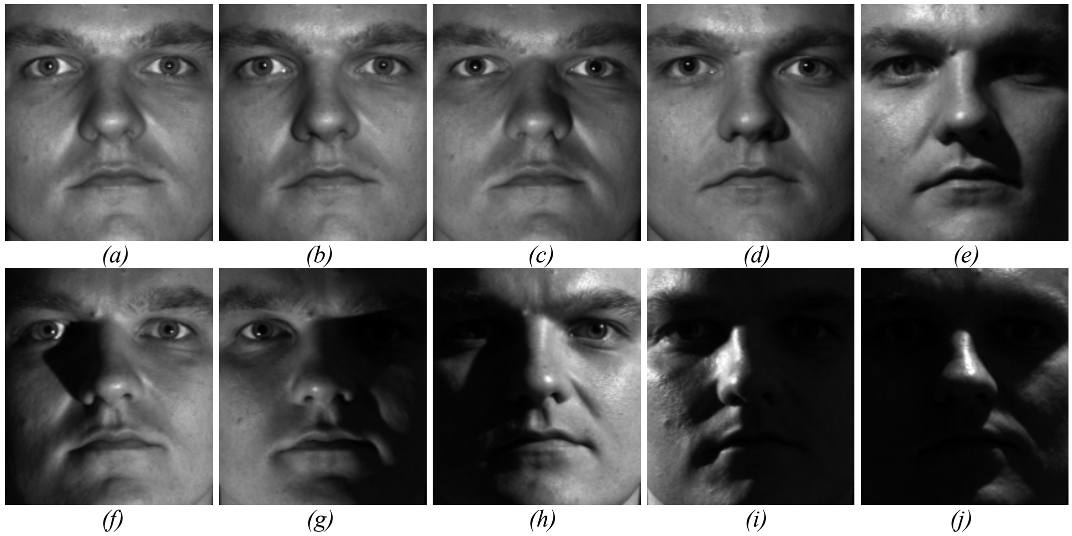

Figure 2: Different images under different illumination conditions, for people 7 in fig 1. (a) and (b) from subset 1; (c) and (d) from subset 2; (e) and (f) from subset 3; (g) and (h) from subset 4; and (i) and (j) from subset 5.

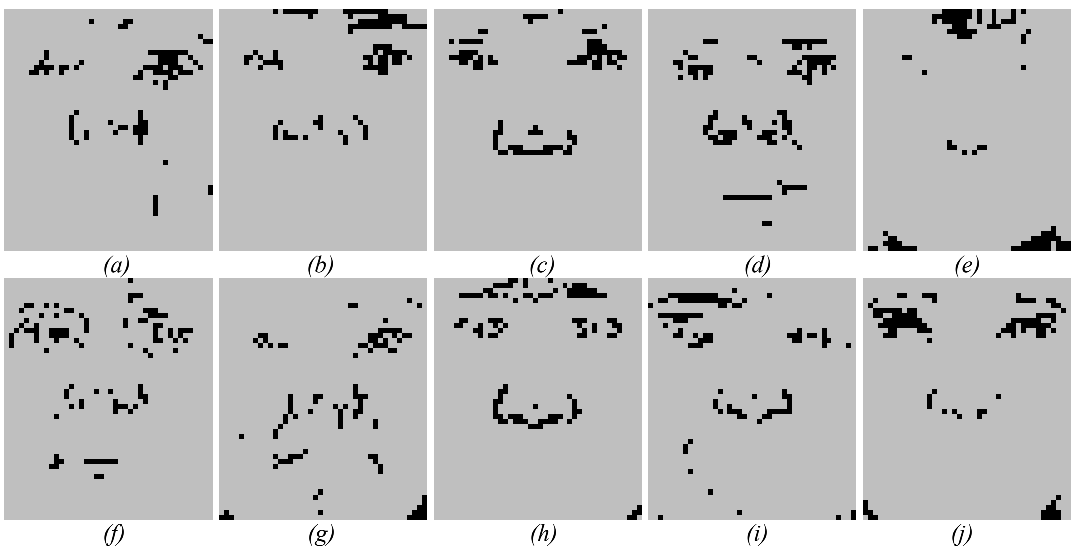

Figure 3: Mask for the outliers, which do not obey the 3-dimensional constraint. The black pixels denote the outliers. From (a) to (j), the masks correspond to people 1 to people 10 in fig. 1.

6. Conclusion

In this paper, we introduce two new techniques into the subspace-based face recognition: outlier detection and a new distance-based criterion for the classification. Without constructing 3D scene, the standard subspace approach, augmented with the new techniques described here, proves to be comparable to Cones-cast, where the cast shadow has to be detected and consequently demands the GBR reconstruction [3]. Moreover, by the new techniques, a good performance can be obtained on “subset 5”, which is the most challenging in the Yale-B face database and on which no performance has been reported.

References:

[1] R. Basri and D.W. Jacobs. *Lambertian Reflectance and Linear Subspaces*. in *ICCV*, pp. 383-390, 1999.

[2] R. Basri and D.W. Jacobs, *Lambertian Reflectance and Linear Subspaces.* IEEE PAMI, vol. **25**, no. 2, pp. 218-233, 2003.

[3] P. Belhumeur, D.J. Kriegman, and A. Yuille, *The Bas-Relief Ambiguity.* IJCV, vol. **35**, no. 1, pp. 33-44, 1999.

[4] P.N. Belhumeur, J. Hespanha, and D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection.* IEEE PAMI, vol. **19**, no. 7, pp. 711-720, 1997.

[5] P.N. Belhumeur and D. Kriegman, *What is the Set of Images of an Object under all Possible Illumination Conditions?* IJCV, vol. **28**, no. 3, pp. 245-260, 1998.

[6] C.M. Bishop. *Bayesian PCA*. in *NIPS*: MIT Press, pp. 382–388, 1998.

[7] P. Chen and D. Suter, *An analysis of linear subspace approaches for computer vision and pattern recognition.* submitted to SIAM Journal on Applied Mathematics, no., 2003.

[8] R. Eipstein, P. Hallinan, and A. Yuille. *5±2 eigenimages suffices: An empirical investigation of low-dimensional lighting models*. in *Proc. IEEE Workshop Physics-Based vision*, pp. 108-116, 1995.

[9] A. Georghiades, P.N. Belhumeur, and D. Kriegman, *From Few to Many: Generative Models of Object Recognition.* IEEE PAMI, vol. **23**, no. 6, pp. 643-660, 2001.

[10] A. Georghiades, D. Kriegman, and P.N. Belhumeur. *Illumination cones for recognition under variable lighting: faces*. in *CVPR*, pp. 52-58, 1998.

[11] G.H. Golub and C.F.V. Loan, *Matrix Computations,*. 2nd ed. 1989, Baltimore: Johns Hopkins University Press.

[12] P. Hallinan. *A low-dimensional representation of human faces for arbitrary lighting conditions*. in *CVPR*, pp. 995-999, 1994.

[13] D. Jacobs, P.N. Belhumeur, and R. Basri. *Comparing Images Under Variable Illumination*. in *CVPR*, pp. 610-617, 1998.

[14] K. Lee, J. Ho, and D. Kriegman. *Nine points of light acquiring subspaces for face recognition under variable lighting*. in *CVPR*, pp. 519-526, 2001.

[15] Y. Moses, Y. Adini, and S. Ullman. *Face recognition: The problem of compensating for changes in illumination direction*. in *ECCV*, pp. 286-296, 1994.

[16] H. Murase and S.K. Nayar, *Illumination planning for object recognition using parametric eigenspaces.* IEEE PAMI, vol. **16**, no. 12, pp. 1219-1227, 1994.

[17] H. Murase and S.K. Nayar, *Visaul learning and recognition of 3-D objects from appearance.* IJCV, vol. **14**, no., pp. 5-24, 1995.

[18] S.A. Nene, S.K. Nayar, and H. Murase. *Software library for appearance matching (SLAM)*. in *ARPA Image understanding workshop*, pp., 1994.

[19] R. Ramamoorthi, *Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertial Object.* IEEE PAMI, vol. **24**, no. 10, pp. 1322-1333, 2002.

[20] R. Ramamoorthi and P. Hanrahan, *On the relationship between radiance and irrandiance: Determining the illumination from images of a convex Lambertian object.* Journal of the Optical Society of America (JOSA A), vol. **18**, no. 10, pp. 2448-2459, 2001.

[21] A. Shashua, *On photometric issues in 3D visual recognition from a single 2D image.* IJCV, vol. **21**, no. 1/2, pp. 99-122, 1997.

[22] G.W. Stewart and J.G. Sun, *Matrix perturbation theory*. 1990: Acedemic press.

[23] M. Turk and A. Pentland, *Eigenfaces for recognition.* J. Cognitive Neuroscience, vol. **3**, no. 1, pp. 71-96, 1991.

[24] J.H. Wilkinson, *The algebraic eigenvalue problem*. 1965, Oxford: Clarendon Press.

[25] A.L. Yuille, D. Snow, R. Epstein, and P.N. Belhumeur, *Determining generative models of objects under varying illumination: shape and albedo from multiple images using SVD and integrability.* IJCV, vol. **35**, no. 3, pp. 203-222, 1999.