Recognition of Digital Images of the Human Face at Ultra Low Resolution Via Illumination Spaces[*]

Jen-Mei Chang[1], Michael Kirby[1], Holger Kley[1], Chris Peterson[1],
Bruce Draper[2], and J. Ross Beveridge[2]

[1] Department of Mathematics, Colorado State University, Fort Collins, CO
80523-1874 U.S.A.
{chang,kirby,kley,peterson}@math.colostate.edu
[2] Department of Computer Science, Colorado State University, Fort Collins, CO
80523-1873 U.S.A.
{draper,ross}@math.colostate.edu

Abstract. Recent work has established that digital images of a human face, collected under various illumination conditions, contain discriminatory information that can be used in classification. In this paper we demonstrate that sufficient discriminatory information persists at ultra-low resolution to enable a computer to recognize specific human faces in settings beyond human capabilities. For instance, we utilized the Haar wavelet to modify a collection of images to emulate pictures from a 25-pixel camera. From these modified images, a low-resolution illumination space was constructed for each individual in the CMU-PIE database. Each illumination space was then interpreted as a point on a Grassmann manifold. Classification that exploited the geometry on this manifold yielded error-free classification rates for this data set. This suggests the general utility of a low-resolution illumination camera for set-based image recognition problems.

1 Introduction

The face recognition problem has attracted substantial interest in recent years. As an academic discipline, face recognition has progressed by generating large galleries of images collected with various experimental protocols and by assessing the efficacy of new algorithms in this context. A number of proposed face recognition algorithms have been shown to be effective under controlled conditions. However, in the field, where data acquisition is essentially uncontrolled,

---

the performance of these algorithms typically degrades. In particular, variations in the illumination of subjects can significantly reduce the accuracy of even the best face recognition algorithms.

A traditional approach in the face recognition literature has been to normalize illumination variations out of the problem using techniques such as nonlinear histogram equalization or image quotient methods [1]. While such approaches do indeed improve recognition, as demonstrated on the FERET, Yale and PIE databases, they do not exploit the fact that the response of a given subject to variation in illumination is *idiosyncratic* [2] and hence can be used for discrimination. This work builds on the observation that in the vector space generated by all possible digital images collected by a digital camera at a fixed resolution, the images of a fixed, Lambertian object under varying illuminations lie in a convex cone [3] which is well approximated by a relatively low dimensional linear subspace [4,5].

In our framework, we associate to a set of images of an individual their linear span, which is in turn represented, or encoded, by a point on a Grassmann manifold. This approach appears to be useful for the general problem of comparing sets of images [6]. In the context of face recognition our objective is to compare a set of images associated with subject 1 to a set of images associated with subject 2 or to a different set of images of subject 1. The comparison of any two sets of images is accomplished by constructing a subspace in the linear span of each that optimizes the ability to discriminate between the sets. As described in [2], a sequence of nested subspaces may be constructed for this purpose using principal vectors computed to reveal the geometric relationship between the linear spans of each subject. This approach provides an immediate pseudo-metric for set-to-set comparison of images. In an application to the images in the CMU-PIE Database [7] and Yale Face Database B [4], we have previously established that the data are *Grassmann separable* [2], i.e., when distances are computed between sets of images using their encoding as points on an appropriately determined Grassmann manifold the subjects are all correctly identified.

The CMU-PIE database consists of images of 67 individuals. While the Grassmann separability of a database of this size is a significant, positive result, it is important to understand the general robustness of this approach. For example, the application of the methodology to a larger data set is of critical interest. In the absence of such data, however, we propose to explore a related question: as we reduce the effective resolution of the images of the 67 individuals which make up the CMU-PIE database, does Grassmann separability persist? The use of multiresolution analysis to artificially reduce resolution introduces another form of nested approximation into the problem that is distinct from that described above.

We observe that facial imagery at ultra low resolutions is typically not recognizable or classifiable by human operators. Thus, if Grassmann separability persists at ultra low resolution, we can envision large *private* databases of facial imagery, stored at a resolution that is sufficiently low to prevent recognition by a human operator yet sufficiently high to enable machine recognition and classification via the Grassmann methods described in Section 2.

Accordingly, the purpose of this paper is to explore the idiosyncratic nature of digital images of a face under variable illumination conditions at extremely low resolutions. In Section 2 we discuss the notion of classification on a Grassmann variety and a natural pseudo-metric that arises in the context of Schubert varieties. In Section 3 we extend these ideas to the context of a sequence of nested subspaces generated by a multiresolution analysis. Results of this approach applied to the CMU-PIE database are presented in Section 4. We contrast our approach with other methods in Section 5 and discuss future research directions in Section 6.

2   Classification on Grassmannians

The general approach to the pattern classification problem is to compare labeled instances of data to new, unlabeled exemplars. Implementation in practice depends on the nature of the data and the method by which features are extracted from the data and used to create a representation optimized for classification.

We consider the case that an observation of a pattern produces a set of digital images at some resolution. This consideration is a practical one, since the accuracy of a recognition scheme that uses a single input image is significantly reduced when images are subject to variations, such as occlusion and illumination [8]. Now, the linear span of the images is a vector subspace of the space of all possible images at the given resolution, and thus, corresponds to a point on a Grassmann manifold.

More precisely, let k (generally independent) images of a given subject be grouped together to form a data matrix X with each image stored as a column of X. If the column space of X, $\mathcal{R}(X)$, has rank k and if n denotes the image resolution, then $\mathcal{R}(X)$ is a k-dimensional vector subspace of $\mathbb{R}^n$, which is a point on the Grassmann manifold $G(k,n)$. See Fig. 1 for a graphical illustration of this correspondence. Specifically, the real Grassmannian (Grassmann manifold), $G(k,n)$, parameterizes k-dimensional vector subspaces of the n-dimensional vector space $\mathbb{R}^n$. Naturally, this parameter space is suitable for subspace-based algorithms. For example, the Grassmann manifold is used in [9] when searching for a geometrically invariant subspace of a matrix under full rank updates. An optimization over the Grassmann manifold is proposed in [10] to solve a general object recognition problem. In the case of face recognition, by realizing sets of images as points on the Grassmann manifold, we can exploit the geometries imposed by individual metrics (drawn from a large class of metrics) in computing distances between these sets of images.

With respect to the natural structure of a Riemannian manifold that the Grassmannian inherits as a quotient space of the orthogonal group, the geodesic distance between two points $A, B \in G(k,n)$ (i.e., two k-dimensional subspaces of $\mathbb{R}^n$) is given by $d_k(A, B) = \|(\theta_1, \ldots, \theta_k)\|_2$, where $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ are the principal angles between the subspaces A and B. The principal angles are readily computed using an SVD-based algorithm [11].
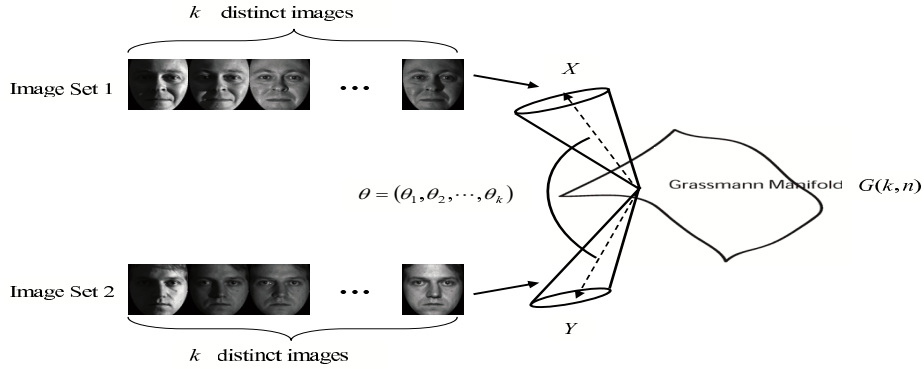
Fig. 1. Illustration of the Grassmann method, where each set of images may be viewed as a point on the Grassmann manifold by computing an orthonormal basis associated with the set

Principal angles between subspaces are defined regardless of the dimensions of the subspaces. Thus, inspired by the Riemannian geometry of the Grassmannian, we may define, for any vector subspaces A, B of $\mathbb{R}^n$,

$$d_\ell(A, B) = \|(\theta_1, \ldots, \theta_\ell)\|_2,$$

for any $\ell \leq \min\{\dim A, \dim B\}$. While d_ℓ is not, strictly speaking, a metric (for example, if $\dim A \cap B \geq \ell$, then $d_\ell(A, B) = 0$), it nevertheless provides an efficient and useful tool for analyzing configurations in $\cup_{k \geq \ell} G(k, n)$. For points on a fixed Grassmannian, $G(k, n)$, the geometry driving these distance measures is captured by a type of Schubert variety $\bar{\Omega}_\ell(W) \subseteq G(k, n)$. More specifically, let W be a subspace of $\mathbb{R}^n$, then we define

$$\bar{\Omega}_\ell(W) = \{E \in G(k, n) \,|\, \dim(E \cap W) \geq \ell\}.$$

With this notation, $d_\ell(A, B)$ simply measures the geodesic distance between A and $\bar{\Omega}_\ell(B)$, i.e. $d(A, \bar{\Omega}_\ell(B)) = min\{d_k(A, C) | C \in \bar{\Omega}_\ell(B)\}$ (it is worth noting that under this interpretation, $d_\ell(A, B) = d_\ell(B, A)$).

3   Resolution Reduction

3.1   Multiresolution Analysis and the Nested Grassmannians

Multiresolution analysis (MRA) works by projecting data in a space V onto a sequence of nested subspaces

$$\cdots \subset V_{j+1} \subset V_j \subset V_{j-1} \subset \cdots \subset V_0 = V.$$

The subspaces V_j represent the data at decreasing resolutions and are called *scaling subspaces* or *approximation subspaces*. The orthogonal complements W_j to

V_j in V_{j-1} are the *wavelet subspaces* and encapsulate the error of approximation at each level of decreased resolution. For each j, we have an isomorphism

$$\phi^j \colon V_{j-1} \xrightarrow{\sim} V_j \oplus W_j.$$

Let $\pi^j \colon V_j \oplus W_j \to V_j$ denote projection onto the first factor and let $\psi^j = \pi^j \circ \phi^j$ (thus $\psi^j \colon V_{j-1} \to V_j$). This single level of subspace decomposition is represented by the commutative diagram in Fig. 2(a).

Let $G(k, V)$ denote the Grassmannian of k-dimensional subspaces of a vector space V. Suppose that V, V' are vector spaces, and that $f \colon V \to V'$ is a linear map. Let $\ker(f)$ denote the kernel of f, let $\dim(A)$ denote the dimension of the vector space A and let

$$G(k, V)^{\circ} = \{ A \subset V \mid \dim(A) = k \text{ and } A \cap \ker(f) = 0 \}.$$

If $k + \dim \ker(f) \leq \dim V$, then $G(k, V)^{\circ}$ is a dense open subset of $G(k, V)$ and almost all points in $G(k, V)$ are in $G(k, V)^{\circ}$. Now if $A \cap \ker(f) = 0$, then $\dim f(A) = \dim A$, so f induces a map

$$f_k^{\circ} \colon G(k, V)^{\circ} \to G(k, V').$$

Furthermore, if f is surjective, then so is f°. The linear maps of the MRA shown in (a) of Fig. 2 thus induce the maps between Grassmannians shown in (b) of the same figure.

Finally, we observe that if A, B are vector subspaces of V, then $\dim(A \cap B) = \dim(f(A) \cap f(B))$ if and only if $(A + B) \cap \ker(f) = 0$. In particular, when $(A + B) \cap \ker(f) = 0$ and $\ell \leq \min\{\dim A, \dim B\}$, then $d_\ell(A, B) = 0$ if and only if $d_\ell(f(A), f(B)) = 0$.

From this vantage point, we consider the space spanned by a linearly independent set of k images in their original space on the one hand, and the space spanned in their reduced resolution projections on the other hand, as points on corresponding Grassmann manifolds. Distances between pairs of sets of k linearly independent images or their low-resolution emulations can then be computed using the pseudo-metrics d_ℓ on these Grassmann manifolds. The preceding observation suggests the possibility that for resolution-reducing projections, spaces which were separable by d_ℓ remain separable after resolution reduction. Of course, taken to an extreme, this statement can no longer hold true. It is therefore of interest to understand the point at which separability fails.

3.2   Image Resolution Reduction

In a 2-dimensional Discrete Wavelet Transform (DWT), columns and rows of an image I each undergo a 1-dimensional wavelet transform. After a single level of a 2-dimensional DWT on an image I of size m-by-n, one obtains four sub-images of dimension $\lceil \frac{m}{2} \rceil$-by-$\lceil \frac{n}{2} \rceil$. If we consider each row and column of I as a 1-dimensional signal, then the *approximation* component of I is obtained by a low-pass filter on the columns then a low-pass filter on the rows and sampled

$$V_j \oplus W_j \xrightarrow{\;\pi^j\;} V_j \qquad G(k, V_j \oplus W_j)^{\circ} \xrightarrow{\;(\pi^j)^{\circ}_k\;} G(k, V_j)$$

$$\phi^j \uparrow \quad \nearrow \psi^j \qquad\qquad (\phi^j)^{\circ}_k \uparrow \quad \nearrow (\psi^j)^{\circ}_k$$

$$V_{j-1} \qquad\qquad\qquad G(k, V_{j-1})^{\circ}$$
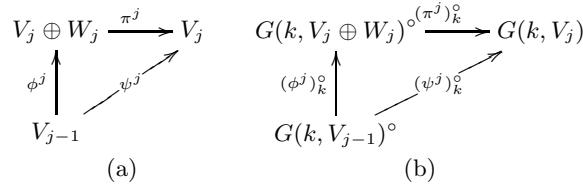
$$(a) \qquad\qquad\qquad\qquad (b)$$

Fig. 2. (a) Projection maps between scaling and wavelet subspaces for a single level of wavelet decomposition. (b) Projection maps between nested Grassmannians for a single level of decomposition.

on a dyadic grid. The other 3 sub-images are obtained in a similar fashion and collectively, they are called the *detail* component of I. The approximation component of an image after a single level of wavelet decomposition with the Haar wavelet is equivalent to averaging the columns, then the rows. See Fig. 3 for an illustration of the sub-images obtained from a single level of Haar wavelet analysis.

To use wavelets to compress a signal, we sample the approximation and detail components on a dyadic grid. That is, keeping only one out of two wavelet coefficients at each step of the analysis. The approximation component of the signal, A_j, after j iterations of decomposition and down-sampling, will serve as the same image in level j with resolution $\lceil \frac{m}{2^j} \rceil$-by-$\lceil \frac{n}{2^j} \rceil$.

In the subsequent discussions, we present results obtained by using the *approximation* subspaces. However, similar results obtained by using the *wavelet* subspaces are also observed.
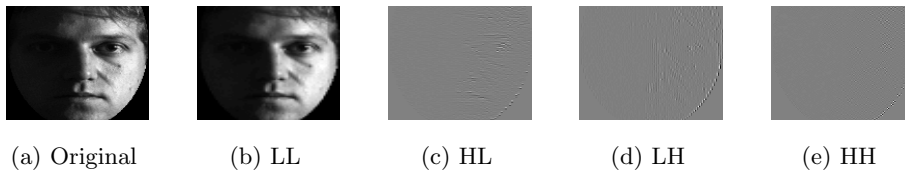

(a) Original   (b) LL   (c) HL   (d) LH   (e) HH

Fig. 3. An illustration of the sub-images from a single level of Haar wavelet analysis on an image in CMU-PIE. From left to right: original image, approximation, horizontal, vertical, and diagonal detail.

4   Results: A 25-Pixel Camera

The experiment presented here follows the protocols set out in [2], where it was established that CMU-PIE is *Grassmann separable*. This means that using one of the distances d_ℓ on the Grassmannian, the distance between an estimated illumination space of a subject and another estimated illumination space of the same subject is always less than the distance to an estimated illumination space of

any different subject. In this new experiment we address the question of whether this idiosyncratic nature of the illumination spaces persists at significantly reduced resolutions. As described below, we empirically test this hypothesis by calculating distances between pairs of scaling subspaces.

The PIE database consists of digital imagery of 67 people under different poses, illumination conditions, and expressions. The work presented here concerns only illumination variations, thus only frontal images are used. For each of the 67 subjects in the PIE database, 21 facial images were taken under lighting from distinct point light sources, both with ambient lights on and off. The results of the experiments performed on the ambient lights off data is summarized in Fig. 4. The results obtained by running the same experiment on illumination data collected under the presence of ambient lighting were not significantly different.

For each of the 67 subjects, we randomly select two disjoint sets of 10 images to produce two 10-dimensional estimates of the illumination space for the subject. Two estimated spaces for the same subject are called matching subspaces, while estimated subspaces for two distinct subjects are called non-matching subspaces. The process of random selection is repeated 10 times to generate a total of 670 matching subspaces and 44,220 non-matching subspaces. We mathematically reduce the resolution of the images using the Haar wavelet, effectively emulating a camera with a reduced number of pixels at each step. As seen in Fig. 5, variations in illumination appear to be retained at each level of resolution, suggesting that the idiosyncratic nature of the illumination subspaces might be preserved. At the fifth level of the MRA the data corresponds to that which would have been captured by a camera with 5×5 pixels. We observe that at this resolution the human eye can no longer match an image with its subject.

The separability of CMU-PIE at ultra low resolution is verified by comparing the distances between the matching to the non-matching subspaces as points on a Grassmann manifold. When the largest distance between any two matching subspaces is less than the smallest distance between any two non-matching subspaces, the data is called Grassmann separable. This phenomenon can be observed in Fig. 4. The three lines of the box in the box whisker plot shown in Fig. 4 represent the lowest quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data and outliers are data with values beyond the ends of the whiskers.

Using d_1, i.e., a distance based on only one principal angle, we observe a significant separation gap between the largest and smallest distance of the matching and non-matching subspaces throughout all levels of MRA. Specifically, the separation gap between matching and non-matching subspaces is approximately $16°$, $18°$, $17°$, $14°$, $8°$, and $0.17°$ when subspaces are realized as points in $G(10, 22080)$, $G(10, 5520)$, $G(10, 1400)$, $G(10, 360)$, $G(10, 90)$, and $G(10, 25)$, respectively. Note that the non-decreasing trend of the separation gap is due to the random selection of the illumination subspaces.
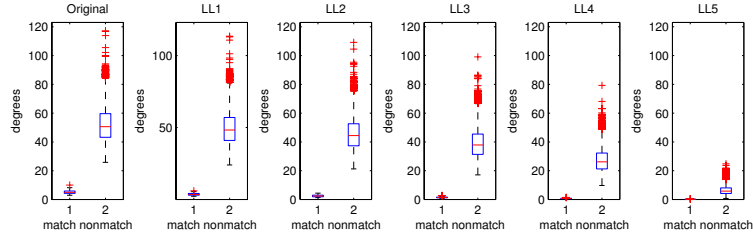
Fig. 4. Box whisker plot of the minimal principal angles of the matching and non-matching subspaces. Left to right: original (resolution 160×138), level 1 Haar wavelet approximation (80×69), level 2 (40×35), level 3 (20×18), level 4 (10×9), level 5 (5×5). Perfect separation of the matching and non-matching subspaces is observed throughout all levels of MRA.

As expected, the separation gap given by the minimal principal angle between the matching and non-matching subspaces decreases as we reduce resolution, but never to the level where points on the Grassmann manifold are misclassified. In other words, individuals can be recognized at ultra-low resolutions provided they are represented by multiple image sets taken under a variety of illumination conditions.

It is curious to see if similar outcomes can be observed when using unstructured projections, e.g., random projections, to embed subject illumination subspaces into spaces of significantly reduced dimensions. To test this, we repeated
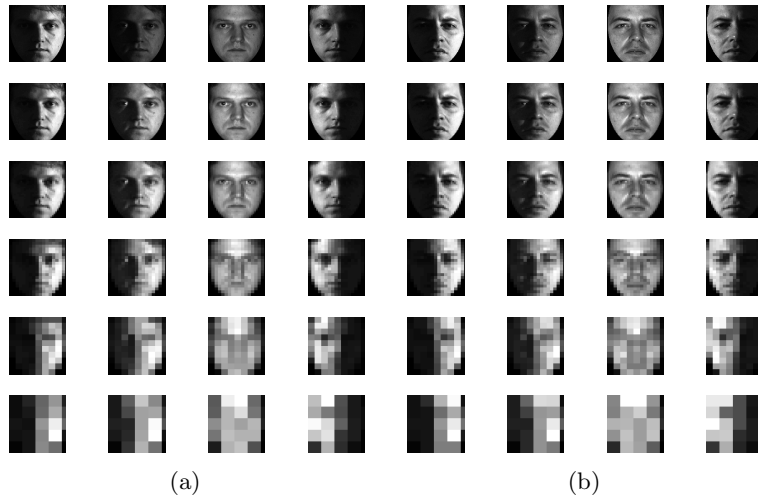

(a)   (b)

Fig. 5. Top to bottom: 4 distinct illumination images of subjects 04006 (a) and 04007 (b) in CMU-PIE; level 1 to level 5 approximation obtained from applying 2D discrete Haar wavelet transform to the top row

the experiments described above in this new setting. Subject illumination subspaces in their original level of resolution were projected onto low dimensional spaces via randomly determined linear transformations. Error statistics were collected by repeating the experiment 100 times. Perfect separation between matching and non-matching subspaces occurred when subject illumination subspaces were projected onto random 35-dimensional subspaces. This validates the use of digital images at ultra low resolution and emphasizes the importance of illumination variations in the problem of face recognition. Furthermore, while unstructured projections perform surprisingly well in the retention of idiosyncratic information, structured projections that exploit similarities of neighboring pixels allow perfect recognition results at even lower resolutions.

We remark that the idiosyncratic nature of the illumination subspaces can be found not only in the scaling subspaces, but also in the wavelet subspaces. Indeed, we observed perfect separation using the minimal principal angle in almost all scales of the wavelet subspaces.

5   Related Work

A variety of studies consider the roles of data resolution and face recognition, including [12,13,14,15,16]. A common feature of these studies is the practice of using single to single image comparison in the recognition stage (with the exception of [16]). Among the techniques used to train the algorithms are PCA, LDA, ICA, Neural Network, and Radial Basis Functions. Some of the classifiers used are correlation, similarity score, nearest neighbor, neural network, tangent distance, and multiresolution tangent distance. If variation in illumination is present in the data set, it is removed by either histogram equalization [17] or morphological nonlinear filtering [18]. Except in [16], the variation of illumination was treated as noise and eliminated in the preprocessing stage before the classification takes place.

In a more related study, Vasconcelos and Lippman proposed the use of transformation invariant tangent distance embedded in the multiresolution framework [16]. Their method, based on the (2-sided) tangent distance between manifolds, is referred to as the multiresolution tangent distance (MRTD) and is similar to our approach in that it requires a set-to-set image comparison. It is also postulated that the use of a multiresolution framework preserves the global minima that are needed in the minimization problems associated with computing tangent distances. The results of [16], however, are that when the only variation in the data is illumination, the performance of MRTD is inferior to that of the normal tangent distance and Euclidean distance. Hence, it appears that the framework of [16] does not sufficiently detect the idiosyncratic nature of illumination at low resolutions.

In summary, we have presented an algorithm for classification of image sets that requires no training and retains its high performance rates even at extremely low resolution. To our knowledge, no other algorithm has claimed to have achieved perfect separability of the CMU-PIE database at ultra low resolution.

6   Discussion

We have shown that a mathematically emulated ultra low-resolution illumination space is sufficient to classify the CMU-PIE database when a data point is a set of images under varying illuminations, represented by a point on a Grassmann manifold. We assert that this is only possible because the idiosyncratic nature of the response of a face to varying illumination, as captured in digital images, persists at ultra low resolutions. This is perhaps not so surprising given that the configuration space of a 25-pixel camera consists of 256^{25} different images and we are comparing only 67 subjects using some 20 total instances of illumination. The representation space is very large compared to the amount of data being stored. Furthermore, the reduction of resolution that was utilized takes advantage of similarities of neighboring pixels. The algorithm introduced here is computationally fast and can be implemented efficiently. In fact, on a 2.8GHz AMD Opteron processor, it takes approximately 0.000218 seconds to compute the distance between a pair of 25-pixel 10-dimensional illumination subspaces.

The work presented here provides a blueprint for a low-resolution illumination camera to capture images and a framework in which to match them with low-resolution sets in a database. Future work will focus on evaluating this approach on a much larger data set that contains more subjects and more variations. The Grassmann method has shown promising results in a variety of face recognition problems [6,19,2], we intend to examine the effect of resolution reduction on the accuracy of the algorithm with a range of variations, such as viewpoint and expressions.

References

1. Riklin-Raviv, T., Shashua, A.: The quotient image: Class based re-rendering and recognition with varying illuminations. PAMI 23(2), 129–139 (2001)
2. Chang, J.M., Beveridge, J., Draper, B., Kirby, M., Kley, H., Peterson, C.: Illumination face spaces are idiosyncratic. In: International Conference on Image Procesing & Computer Vision, vol. 2, pp. 390–396 (June 2006)
3. Belhumeur, P., Kriegman, D.: What is the set of images of an object under all possible illumination conditions. IJCV 28(3), 245–260 (1998)
4. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. PAMI 23(6), 643–660 (2001)
5. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. PAMI 25(2), 218–233 (2003)
6. Chang, J.M., Kirby, M., Kley, H., Beveridge, J., Peterson, C., Draper, B.: Examples of set-to-set image classification. In: Seventh International Conference on Mathematics in Signal Processing Conference Digest, The Royal Agricultural College, Cirencester, Institute for Mathematics and its Applications, pp. 102–105 (December 2006)
7. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. PAMI 25(12), 1615–1618 (2003)

8. Yamaguchi, O., Fukui, K., Maeda, K.: Face recognition using temporal image sequence. In: AFGR, pp. 318–323 (1998)
9. Smith, S.: Subspace tracking with full rank updates. In: The 31st Asilomar Conference on Sinals, Systems & Computers, vol. 1, pp. 793–797 (November 1997)
10. Lui, X., Srivastava, A., Gallivan, K.: Optimal linear representations of images for object recognition. PAMI 26, 662–666 (2004)
11. Golub, G.H., Loan, C.F.V.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
12. Kouzani, A.Z., He, F., Sammut, K.: Wavelet packet face representation and recognition. In: IEEE Int'l Conf. on Systems, Man and Cybernetics, Orlando, vol. 2, pp. 1614–1619. IEEE Computer Society Press, Los Alamitos (1997)
13. Feng, G.C., Yuen, P.C., Dai, D.Q.: Human face recognition using PCA on wavelet subband. SPIE J. Electronic Imaging 9(2), 226–233 (2000)
14. Nastar, C., Moghaddam, B., Pentland, A.: Flexible images: Matching and recognition using learned deformations. Computer Vision and Image Understanding 65(2), 179–191 (1997)
15. Nastar, C.: The image shape spectrum for image retrieval. Technical Report RR-3206, INRIA (1997)
16. Vasconcelos, N., Lippman, A.: A multiresolution manifold distance for invariant image similarity. IEEE Trans. Multimedia 7(1), 127–142 (2005)
17. Ekenel, H.K., Sankur, B.: Multiresolution face recognition. Image Vision Computing 23(5), 469–477 (2005)
18. Foltyniewicz, R.: Automatic face recognition via wavelets and mathematical morphology. In: Proc. of the 13th Int'l Conf. on Pattern Recognition, vol. 2, pp. 13–17 (1996)
19. Chang, J.M., Kirby, M., Peterson, C.: Set-to-set face recognition under variations in pose and illumination. In: 2007 Biometrics Symposium at the Biometric Consortium Conference, Baltimore, MD, U.S.A. (September 2007)