Computer Vision and Image Understanding 114 (2010) 135–145

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Comparing and combining lighting insensitive approaches for face recognition

Raghuraman Gopalan [a,*], David Jacobs [b]

[a] Dept. of ECE, University of Maryland, College Park 20742, USA
[b] Dept. of Computer Science, University of Maryland, College Park 20742, USA

ARTICLE INFO

Article history:
Received 19 April 2009
Accepted 27 July 2009
Available online 12 August 2009

Keywords:
Face recognition
Gradient direction
Lighting
Classifier comparison and combination

ABSTRACT

Face recognition under changing lighting conditions is a challenging problem in computer vision. In this paper, we analyze the relative strengths of different lighting insensitive representations, and propose efficient classifier combination schemes that result in better recognition rates. We consider two experimental settings, wherein we study the performance of different algorithms with (and without) prior information on the different illumination conditions present in the scene. In both settings, we focus on the problem of having just one exemplar per person in the gallery. Based on these observations, we design algorithms for integrating the individual classifiers to capture the significant aspects of each representation. We then illustrate the performance improvement obtained through our classifier combination algorithms on the illumination subset of the PIE dataset, and on the extended Yale-B dataset. Throughout, we consider galleries with both homogenous and heterogeneous lighting conditions.

1. Introduction

There are many algorithms in the literature that address the problem of lighting insensitive 2D face recognition. This is a challenging problem because lighting drastically affects the appearance of a face. In this paper, we attempt to address this problem by understanding the relative merits of different lighting insensitive representations. We make two main contributions. First, we compare a number of algorithms (both class-based, and class-independent) from the perspective of how well they capture different properties of the human face such as, changes in albedo, and changes in surface normal orientations. After analyzing the relative strengths of these algorithms, we propose effective classifier combination schemes that encode such information to produce better recognition performance.

1.1. Relation to prior work

There are quite a few works in the literature that provide a comparative study of lighting invariant face recognition algorithms. For instance, Ruiz-del-Solar and Quinteros [1] investigate a set of illumination compensation and normalization approaches in an eigenspace-based face recognition setup. They compare the algorithms based on the modeling stages required, simplicity, speed and recognition rates. France and Nanni [2] compare the recognition rates of a set of image based and 3D model based algorithms, and then propose a simple fusion algorithm based on the sum rule to highlight the advantage of classifier fusion.

We mainly differ from the existing surveys in two aspects. First, we study how robust different representations are, in capturing face properties (such as changes in albedo, and surface normal orientation) under lighting variations. Next, we are specifically interested in performing recognition when there is only one exemplar for each person in the gallery. On top of this, we consider galleries with both homogenous and heterogeneous lighting across different subjects. This setting, though restrictive, applies to many real-life conditions wherein we may have only one sample picture of a person (with arbitrary lighting condition) for recognition. This makes the problem much more challenging. We consider two experimental settings. One (in Section 3), when there is no prior training information on the effect of typical lighting changes on faces, where we analyze the performance of five class-independent representations. And the other (in Section 4), which provides some training data showing the possible lighting conditions, wherein we also include four class-based algorithms in the analysis, since they can use the prior lighting information to learn to perform classification.

1.2. Contributions of this paper

Given this experimental setup, we make the following three observations to enable better understanding of lighting insensitive face recognition. First, after reviewing nine algorithms we consider in Section 2 (spanning both class-based and class-independent approaches), in Section 3.1 (and in Section 4.1) we compare their performance on the PIE data set [3]. We find that two very simple

* Corresponding author.
E-mail addresses: raghuram@umiacs.umd.edu (R. Gopalan), djacobs@umiacs.umd.edu (D. Jacobs).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.07.005

methods perform the best. Overall a very simple comparison method using the direction of the image gradient performs better than a number of more recent approaches.

Second, we note that a face contains quite different sources of information, including albedo changes (e.g., eyebrows), regions of rapid change in surface orientation (e.g., nose) and smooth regions (e.g., cheeks). By looking at individual regions, we can get a better understanding of how well each algorithm makes use of each source of information. In Section 3.2 we show experimentally that the relative performance of different class-independent algorithms varies between different regions of the face. To gain intuition, we then consider very simple idealizations of different face regions, and highlight extreme differences of performance for different surfaces.

Finally, these results suggest that we may be able to achieve better performance by combining different representations, benefiting from their different strengths. We show that this is indeed true, demonstrating performance gains with a very simple combination scheme (in Section 3.3) that adaptively integrates information from different class-independent representations on the various facial sub-regions, and then (in Section 4.2) by combining information from both class-based and class-independent methods using an SVM that automatically learns the relative importance of these algorithms.

It will be an interesting topic of future research to determine how best to integrate these representations into recognition algorithms that allow for small changes in pose and facial expression, such as those seen in the recent FRGC data set. However, in this paper we wish to isolate the effect that lighting change alone has, and to understand this effect thoroughly. For this reason, we experiment using the illumination portion of the CMU-PIE data set [3] (shown in Fig. 1), and the extended Yale-B dataset [4] (shown in Fig. 2), which controls other sources of variation. We then evaluate the scalability of these representations on images with more controlled lighting conditions, but with other image variations, using the ORL face database [5]. In addition to a standard experimental set-up, in which all gallery images are created with identical lighting, we also consider the more challenging case, in which every gallery image is produced by randomly chosen lighting. This simulates some of the challenges of real-world data sets.

2. Description of algorithms

We compared nine algorithms, including both class-based approaches and class-independent approaches, in our experiments. Although this set of algorithms is certainly not exhaustive, it does give a good sample of different approaches to lighting insensitive face recognition. A brief description of these approaches is given below.

Fig. 1. Sample images from the CMU-PIE dataset [3].

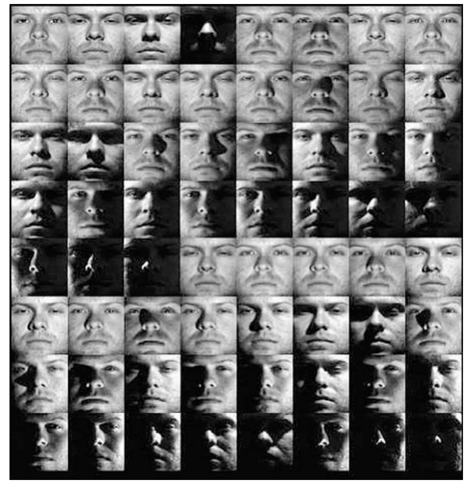

Fig. 2. Sample images from the extended Yale-B dataset [4].

Eigenfaces [6] is a standard benchmark for face recognition. It projects face images into a low-dimensional linear subspace found using principal component analysis. Although not especially well suited to handling lighting variation, it provides a useful point of comparison.

The *Fisherfaces* algorithm [7] (see also [8]) projects images into a direction that not only separates different classes, but also minimizes the within-class scatter. This was explicitly proposed as an effective way to capture variations due to lighting.

Bayesian face recognition [9] models variations between images from the same or different individuals using mixtures of Gaussians. The similarity measure is computed based on the maximum-a-posteriori rule, as opposed to the Euclidean norm. In principle, it can model changes due to lighting.

Correlation filters [10] introduce the use of spatial frequency domain methods for lighting insensitive face recognition. A separate filter is trained for every subject (using their 2D Fourier transform representation) such that it produces sharp correlation peaks for the images belonging to that subject, and low values otherwise.

Instead of modeling the illumination variations using face-specific information (as in [10]), the *image preprocessing* algorithm [11] estimates the luminance map present in the image in order to compensate for it, and thereby produces the reflectance map that contains the true information about the facial features of the subjects. This preprocessed image can then be fed into any classifier. We used Eigenfaces [6] to perform classification, as suggested in [11].

Along similar lines, the *self-quotient* image [12] estimates the reflectance of the image by convolving the image with a smoothing kernel and then dividing the original image by the smoothed image (which mostly contains the low frequency components that correspond to illumination effects), and has shown very good performance on PIE data. In this work we used a much simpler isotropic smoothing instead of anisotropic smoothing (as suggested in [12]). In this form, the algorithm amounts to smoothing the image with a Gaussian, and then pixel-wise dividing the original image by the smoothed image. We obtained the same results given by the authors for the original algorithm, but the results could be different on other datasets.

Another algorithm that displays insensitivity to illumination is the *Eigenphases* [13] method. This algorithm uses the phase information from the frequency domain representation of the image for classification. It is known that the phase information retains most of the intelligibility of the image when compared to the magnitude information of the spectral components, and the authors demonstrate this for the task of face recognition.

The *Whitening* approach described in [14] is specialized for smooth regions wherein the albedo and the surface normal of the neighboring pixels are highly correlated. This means that the pixel independence assumption made implicitly in computing the sum of squared distances (SSD) is not optimal. This algorithm tries to increase the dissimilarity between the images of different objects by decorrelating the image intensities by applying a Whitening operator. We use the simple Laplacian of Gaussian operator for whitening, as suggested in [14].

Finally, classification based on the *gradient* direction of the images has also been shown to work well on surfaces, including faces, having properties that change much more rapidly in one direction than in the other (e.g., [14,15] reviews many papers that use this method, going back to the early 1990s). We implement this method by computing the SSD between the gradient directions in two images. There are other methods that perform well for lighting invariant recognition such as, Gabor Jets [16] and Normalized correlation [17] using small windows. However these two methods have been shown to be quite similar to gradient direction in [14] and hence they are not included in our experiments

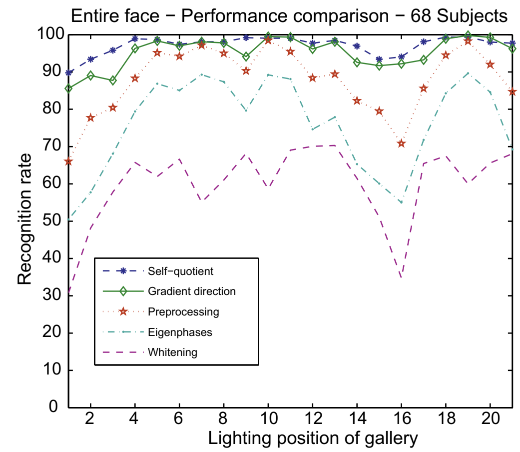

Fig. 3. Performance comparison of all the class-independent algorithms on the entire face (without training). Note: all the graphs are best viewed in color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Setting 1: no training set (on the possible lighting conditions)

In this section, we analyze the performance of the algorithms in the absence of any prior information on the lighting conditions present in the scene. Under such conditions, since the class-specific algorithms do not have sufficient exemplars to learn the lighting variations present in the scene, we consider only the class-independent representations. We then divide the face into several regions to study the relative performance of these algorithms. We provide intuitive explanations for the variations in their performance, and then use this information to design an effective classifier combination algorithm.

3.1. Initial comparisons

We compare the five class-independent algorithms using a standard experimental protocol for PIE data. Each image in this dataset contains one of 68 individuals viewed from the frontal pose and illuminated by a point source of light from one of 21 different directions, without the ambient lighting conditions as shown in Fig. 2. For all the experiments we used properly cropped faces (by removing the scene background present in the dataset images and retaining only the facial region).

In many applications we do not have access to multiple images of a person with the same pose under different illumination conditions. Hence algorithms that perform well with a minimum number of images of a person are normally preferred. Therefore for all the 68 subjects, we use one illumination condition as the gallery (which contains sample images of the subjects) and the remaining 20 illumination conditions as the probes (which will be compared with all the images in gallery). This is a standard set-up, adopted in many previous papers.

The result of this experiment is given in Fig. 3, which shows recognition rates when each lighting condition is used as the gallery. It can be observed that whitening [14] performs much worse than the other class-independent algorithms. But we retain [14] for the experiments involving sub-regions of the face because it is supposed to perform better in smooth regions.

One difficulty with these results is that the performance of the best algorithms is perfect in many cases, making it difficult to distinguish between them. To address this, we also consider a much more challenging recognition task, in which each individual's gallery image is randomly chosen. This makes recognition much more difficult, since it is likely that faces of different individuals taken under similar lighting will appear to be more similar than faces of the same individual taken under very different lighting. However, this difficulty reflects the challenges of many real-world problems, such as sorting personal photos, in which gallery images taken under controlled conditions are not available. Results, averaged over twenty different trials, are given in the Table 1.

Overall, the self-quotient and gradient direction produce the best performance with a homogenous gallery, with gradient direction performing much better with a gallery formed from heterogeneous lighting. This is rather surprising, since gradient direction is very simple, and is the earliest of these approaches. These results suggest that gradient direction would be an appropriate benchmark algorithm when new methods are proposed.

3.2. Facial sub-regions

Next, we explore the performance of these algorithms in more detail. As noted, the face provides different sorts of information due to variations in albedo and shape. To get an idea of how different algorithms make use of this information, we divide the face coarsely into seven regions (Fig. 4): eyes, nose, lips, two cheek regions and two chin regions. We experiment with the algorithms in these regions, and then provide simple models, to gain a better intuitive understanding of the results.

Some existing works on studying the contributions of different face regions include Nanni and Maio [18], and Nanni and Lumini [19]. In [18], features are extracted from different sub-windows of a face using a bank of Gabor filters and Karhunen–Loeve transform. The features obtained by each pattern are used to train a Parzen window classifier to perform face recognition. On the other hand [19], combines wavelet coefficients from selected sub-bands of several wavelet families and performs face authentication.

Table 1
Performance of algorithms on homogenous gallery and heterogeneous gallery.

Algorithm	Homogenous gallery	Heterogeneous gallery
Self-quotient	97	64
Gradient direction	95	78
Preprocessing	88	66
Eigenphases	74	60
Whitening	60	40

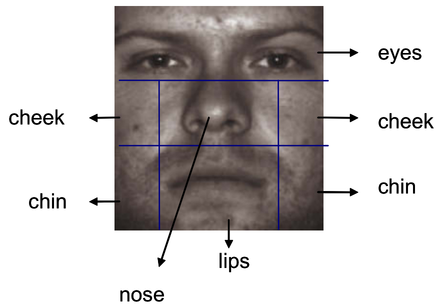

Fig. 4. Face sub-regions.

3.2.1. Experiments

For each region, both the gallery and the probe contain the same facial features cropped from the face. For all 68 subjects, one illumination condition was used as the gallery and the remaining lighting conditions were used as probes.

The results of recognition experiments on different facial features are provided in Fig. 5. We show results for all 68 individuals using gallery images that contain the same lighting. This avoids the need to average over random trials, and still provides sufficient difficulty to evaluate the methods without a ceiling effect, because recognition using a single face region is quite difficult. We performed recognition just as in the last section, but using isolated facial regions.

We find that the relative performance of the different algorithms varies in the different facial sub-regions. Some of the most noticeable effects are: the self-quotient image algorithm [12] performs the best in all regions except for the nose region; gradient direction performs well everywhere except for the cheek region; Whitening [14] performs poorly, but relatively better in the cheek region.

3.2.2. Analysis of simple models

To analyze these results, we model the effects of lighting variation on three simple types of scenes. These are related to important facial characteristics. We make the following observations. First, the face contains albedo variations, especially in the eyes, eyebrows and lips. Second, we consider regions of very high curvature or discontinuity in surface normals, especially at the nose. Finally, the remainder of the face contains regions of smooth variation in shape with little change in albedo. We model these three types of regions with very simple, synthetic models, for which it is easier to understand algorithm performance. We do not expect results with these simple models to perfectly match experiments on faces, since any one face region contains a mix of all three effects. However, we do see that our models explain some of the general trends of our experiments.

3.2.2.1. Planar models with albedo variations. Through this model, we would like to characterize planar objects that exhibit very large variations in albedo. Towards this end, we create images containing an outer rectangular box of fixed size and an inner rectangular box of variable size (Fig. 6).

This representation has some degree of correlation with the eye region of the face, which has large variations in albedo, due to the eye and eyebrow, while it has much smaller variation in shape. Specifically, the inner rectangle can be related to the human eye while the outer rectangle corresponds to the region surrounding the eyes. We assume that the rectangular surface is lambertian and that the point light source is at a far distance from the object. The illumination conditions of the two rectangles are varied by changing the position of the point light source. To capture variations between individuals, the position and size of the inner rectangle are changed by small amounts for all possible illumination conditions. Ninety different illumination conditions were generated for 400 possible positions and sizes of the inner rectangle. Based on this synthetic dataset, the following results were obtained (Table 2). A recognition setup, like the one discussed in Section 3.2.1, was adopted.

As in the case of the human face, self-quotient and gradient direction based methods perform very well in these synthetic conditions. The gradient direction method works very well due to the presence of rich information of the gradient angle change in the boundary between the two rectangles.

The self-quotient image algorithm works well because there is a no change in the surface normal and there is a sizable change in the albedo. In these conditions [12], points out that self-quotient is invariant to lighting changes. This algorithm is shown to capture the albedo changes very well. Whitening does badly as the albedo is not smooth, and is not whitened by the filter we use.

3.2.2.2. Shape variations in smooth objects. In this model, we attempt to simulate the case wherein the object is predominantly smooth, with gradual variations in its shape. Such a variation can be captured by a small piece of a smooth cylinder (Fig. 7). We construct this model by considering cylinders of different radii (accounting for the different subjects) and varying the position of the point light source for each cylinder. This representation correlates with the human cheeks where different human cheeks vary in curvature, without discontinuities in shape or much variation in albedo. Again we assume that the cylindrical surface is lambertian and that the point light source is distant from the object.

The dataset contains cylinders of 11 different radii with 9 illumination conditions and the results are given in Table 3. We see that the gradient direction based method performs very poorly, matching the fact that it is also the least effective method on human cheeks. Even though gradient direction is invariant to lighting for a cylinder, there is no variation in direction of gradient between subjects, while the gradient direction does not capture the changes in curvature. The self-quotient algorithm works well because the Gaussian kernel which is used to filter the image attenuates different frequencies in different ways. The intensity is basically a sine wave, and when we divide it by the smoothed sine, we get a constant function whose magnitude encodes the cylinder's curvature. The intensity of the resulting representation therefore captures the dominant frequency of the initial image. The algorithm uses this criterion to classify these images and is thereby invariant to changes in illumination. Whitening's good recognition rates are in line with the prediction in [14] that it will perform well on smooth surfaces. Eigenphases performs well because the phase spectrum of the signal will be a function of the frequency information present in the signal. This frequency information helps this algorithm to classify the query images properly and thereby give good recognition rates.

3.2.2.3. Shape variations in objects with discontinuities. Through this model, we capture the variations in the shape of an object that has some discontinuities. The motivation behind this model is to obtain an approximate representation of the human nose, which can be modeled as a prism. We consider the two sides visible from the frontal view of a prism to represent the nose (Fig. 8). This model, however, does not exactly represent the nose because, we do not consider the effects caused by the dark holes in the bottom region of the nose. A human nose may be a combination of our simple prism model, and a model of albedo variation, such as our eye model.

The shape of this pyramidal surface is changed to represent different individuals and the position of the light source is moved to create different lighting conditions. The experiment consisted of 12

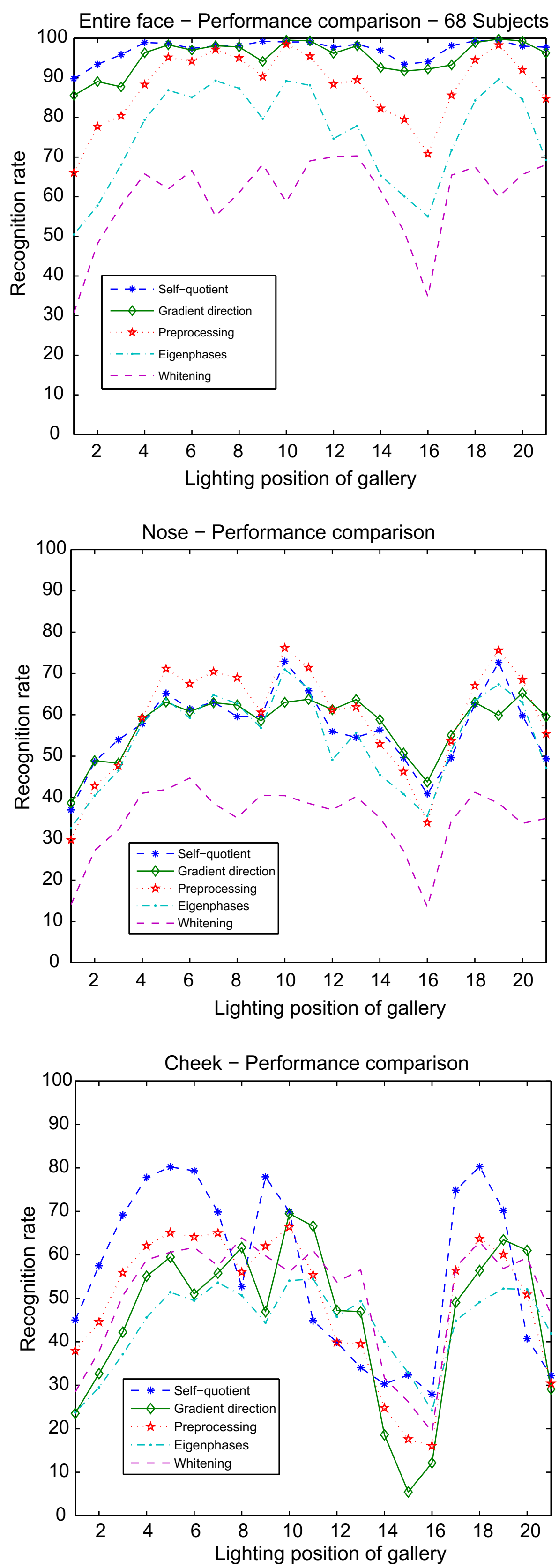

Fig. 5. Performance comparison of class-independent algorithms on different regions without training. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subjects with 10 illumination conditions each and the results are given in Table 4.

The self-quotient image does not perform well due to the change in the orientation of the surface normal between the differ-

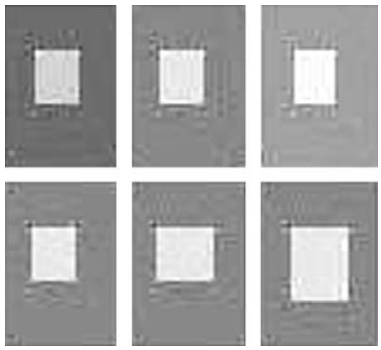

Fig. 6. Rectangle model. Top: variation in illumination. Bottom: variation in albedo.

Table 2
Performance of algorithms on the rectangle model.

Algorithm	Recognition rate (Rectangle model)	Recognition rate (Human eyes)
Self-quotient	100	90.3
Gradient direction	100	88
Preprocessing	49.4	73
Eigenphases	49.4	71.4
Whitening	17.6	58.2

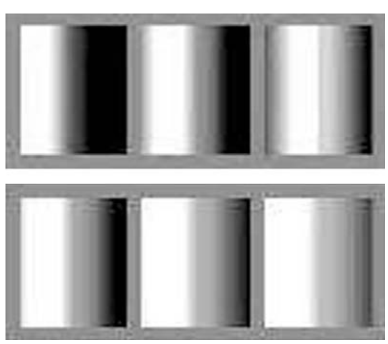

Fig. 7. Cylinder model. Top: variation in illumination. Bottom: variation in curvature (see that as the curvature increases from left image to right image in the bottom row, the change in the lighting pattern gets slower.

Table 3
Performance of algorithms on the cylinder model.

Algorithm	Recognition rate (Cylinder model)	Recognition rate (Human cheek)
Self-quotient	100	56.5
Gradient direction	9.1	41.5
Preprocessing	100	49.2
Eigenphases	100	44.1
Whitening	100	50.8

ent regions in the triangular model. Lighting variations can change the ratio of the intensity in two regions of the prism, and the self-quotient cannot undo this. Thus we find that the self-quotient algorithm is not very effective in capturing shape variations, as predicted in [12]. This matches the fact that the nose is the only region in which self-quotient is not the best. Gradient direction works well because it captures the variation in the shape of the triangles. Whitening does not perform well due to the absence of smooth variations in the surface. The preprocessing algorithm is formulated in such a way that, it controls the illumination variations both in regions where the luminance changes smoothly and

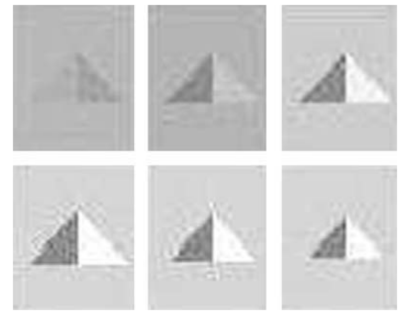

Fig. 8. Triangle model. Top: variation in illumination. Bottom: variation in shape.

Table 4
Performance of algorithms on the triangle model.

Algorithm	Recognition rate (Triangle model)	Recognition rate (Human nose)
Self-quotient	42.7	57
Gradient direction	100	57.6
Preprocessing	57.8	59.1
Eigenphases	46.7	54.3
Whitening	28.7	34.8

in regions where there are discontinuities. So, this algorithm performs relatively well in all the regions.

Our results are related to, but also differ somewhat from the discussion in [15] and [14]. They point out that representations related to the direction of the gradient are insensitive to lighting variation for surfaces that change rapidly in shape or albedo in one direction but not another, while whitening approaches are better suited for smooth surfaces that vary slowly in both directions. First, we show that algorithm performance can vary depending on whether variations occur in shape or in albedo. Second, we show with our cylinder example that variations within a class must also be considered when determining the effectiveness of a representation. In some cases, gradient direction may not discriminate within a class, while features such as curvature do.

3.3. Classifier combination

The fact that different approaches perform well on different parts of the face suggests that we can improve overall performance by combining methods. To demonstrate this, we experimented with a simple method for combining representations.

First, the outputs of the two top performing algorithms for every feature (including the entire face) are combined by normalizing the SSD for every gallery-probe combination and then adding the normalized results of the top two algorithms. For example, self-quotient image and gradient direction algorithms were combined for the entire face and eyes, self-quotient image and preprocessing algorithms were combined for the nose region and so on. We show the results of adaptively integrating different representations on various facial regions in Table 5 and Fig. 9, for the task of recognition with a homogenous gallery.

It can be seen that our classifier combination algorithm results in a substantial improvement in situations, like the nose, where the best individual algorithm doesn't perform that well. For regions such as the entire face, the performance improvement is only moderate since the best individual algorithm by itself has recognition rates close to the ceiling. These results, in effect, drive home the point that an effective classifier combination algorithm should take into account the relative strengths of the individual classifiers in capturing different characteristics of the object of interest.

Table 5
Performance comparison of combined classifier with the best individual algorithms.

Region	Recognition rate of the combined classifier	Recognition rate of the best individual algorithm
Entire face	99.1	97.1
Eyes	95.7	90.3
Lips	83.3	80.6
Nose	69.3	59.1
Chin	64.9	59.5
Cheek	62.8	56.5

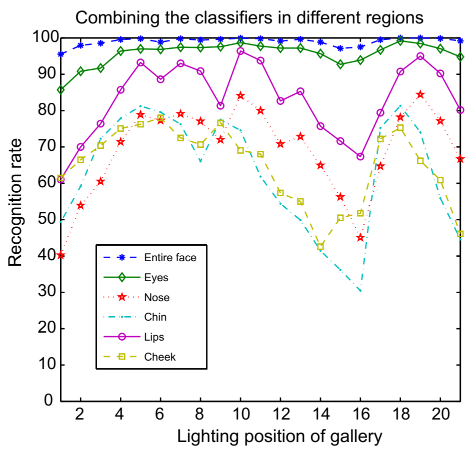

Fig. 9. Performance comparison of combined classifier on all facial regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

With this encouraging result in hand, we would like to formulate a combination algorithm that automatically learns the relative strengths of the individual algorithms, rather than having a user specifying which algorithms to combine based on observation. Towards this end, in Section 4, we consider the setup of having some prior information on the different lighting conditions present in the scene such that one can get a feel of the relative performance of different classifiers and thereby learn the ideal combination strategy before testing it out on the subjects of interest. Since we have a representative training set, we now include the four class-based algorithms (discussed in Section 2) into our analysis.

4. Setting 2: with prior training on different lighting conditions

In this section we first analyze the performance of different algorithms in the presence of training data that contains representative lighting conditions present in the scene. We then combine the most informative algorithms using a support vector machine (SVM) framework, which learns the combination parameters automatically.

4.1. Initial comparisons

We compare the five class-independent algorithms, along with the four class-based algorithms on the PIE dataset. We use all 21 illumination conditions of the first 34 subjects for training. The algorithms were then tested on the remaining 34 subjects, with one homogenous exemplar lighting condition (for all the subjects) in the gallery. This test is done mainly to analyze the effect of training on the different class-based algorithms. The class-independent algorithms, of course, were tested directly on the second half of the 34 subjects.

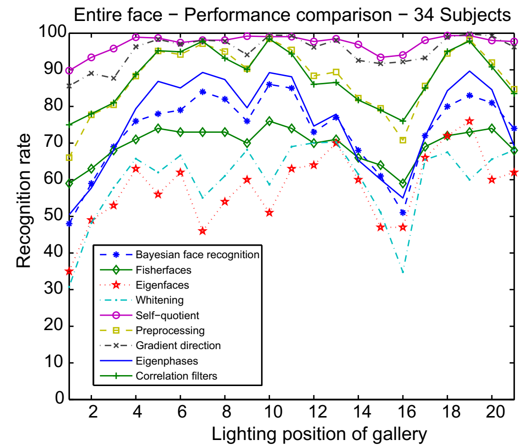

Fig. 10. Performance comparison of class-based, and class-independent algorithms on entire face. The same gallery lighting condition was used for all subjects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The result of this experiment is given in Fig. 10, which shows the recognition rates when each lighting condition was used as the gallery. Similar to the experiments conducted without training data (in Section 3.1), the algorithms based on self-quotient image [12], and the direction of image gradients [15] perform the best. Yet another observation is that the three class-based methods (Fisherfaces [7], Bayesian face recognition [9], Eigenfaces [6]), and whitening [14] perform worse than other algorithms. So, we exclude [7], [9], and [6] from the experiments for classifier combination. However, we retain [14] since it adds considerable value in the cheek region (as shown in Section 3.2.1). For the correlation filters algorithm [10] (and for the class-based algorithms in general), we do not perform the analysis on different facial sub-regions because it is a learning algorithm, and it is difficult to give intuitive explanation of its performance variations (if any) on different facial regions. Now that we have representative algorithms from both class-based and class-independent streams, we discuss our proposed combination strategy in the next sub-section. Specifically, we consider the recognition setup wherein the gallery and the probes have heterogeneous lighting conditions, in order to overcome the ceiling effect in the recognition rates of certain algorithms, and also because this setup simulates a more representative real-world setting.

4.2. Classifier combination

Along similar lines with the discussion in Section 3.3, we expect that we can achieve better performance by using learning to determine the best way of combining information. We do this by training a support vector machine (SVM) [20] to perform a verification task, as done previously by [21], for instance. Given a pair of images, the SVM is trained to determine whether they come from the same or different individuals. The radial basis function (RBF) kernel was used to map the inputs to a higher dimensional space. The SVM was trained using intra-personal pairs and extra-personal pairs from the first 34 subjects of the PIE dataset, and tested with randomly generated pairs from the remaining subjects. The lighting conditions used for training and testing were also disjoint. The input to the SVM is the (absolute) difference between the two images after processing them to create six different representations based on gradient direction, self-quotient, eigenphases, whitening, image preprocessing and correlation filters. We contrast the performance of an SVM that uses all six representations with six SVMs that each use just one of the representations. [22]

have used a similar approach, training an SVM with just differences in gradient direction.

The result of the SVM combination is given in Fig. 11 in the form of Correct Accept Rate (CAR) vs. Correct Reject Rate (CRR) curves; It can be seen that the combination results in a good improvement in verification accuracy. For example, the combined method has an Equal Error rate of 7%, compared to 10% for the best individual algorithm (using the gradient direction). In order to check the generalizability of these results, we experimented with the extended Yale-B dataset [4]. This dataset has cropped faces of 38 subjects under 64 different lighting conditions. All the other variations such as pose, and expressions are fixed. We then performed a similar verification experiment, by training the SVM using the lighting conditions corresponding to the first 18 individuals, and tested it using the pair-wise differences obtained from the remaining 20 subjects. The CAR–CRR curves for the SVM combination, as well as the individual algorithms are given in Fig. 12. Once again, the representation based on the direction of image gradient is the best individual algorithm, followed by the correlation filters. The proposed classifier combination algorithm again results in a substantial improvement in performance over the individual algorithms. These results again reinforce our observation that, combining different representations by learning their relative strengths is crucial to obtain good performance improvement.

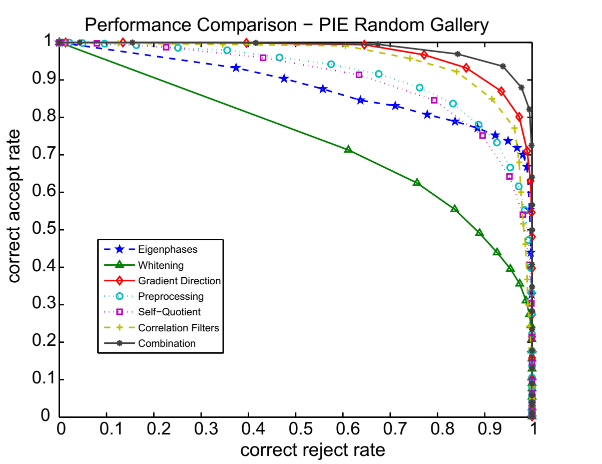

Fig. 11. CAR–CRR curves for PIE heterogeneous gallery experiments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

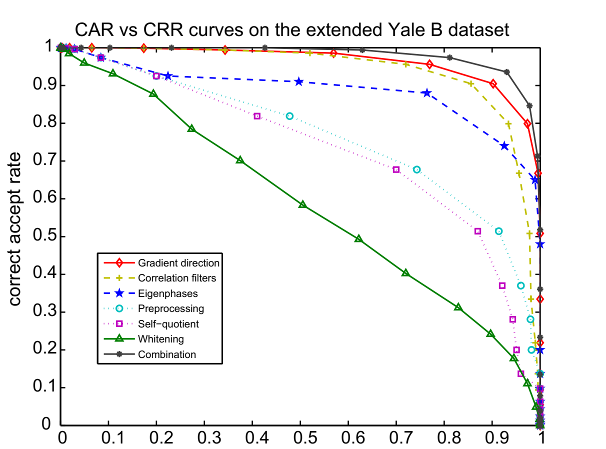

Fig. 12. CAR–CRR curves for the extended Yale-B heterogeneous gallery experiments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Experiments on faces with more controlled lighting

Our main focus is on understanding the role of different representations when lighting changes. However, it is also important to determine the relative sensitivity of these methods to other image variations. If a representation is insensitive to lighting variation, but highly sensitive to changes in expression, for example, it may be less useful in a general face recognition system.

To evaluate this we have experimented using the ORL dataset [5], which has large variations in pose and expression, but small variations in lighting. Fisher discriminant analysis (LDA) [7] has been shown to be effective on this data [23], helping to compensate for the correspondence problem that this data gives rise to. Therefore we use LDA as the base method, in combination with the lighting insensitive representations. Due to the challenges of this data (see Fig. 13), we adopt the widely used leave-one-out testing protocol.

Fisher discriminant analysis [7] was performed on the training images using the six different lighting insensitive representations used in Section 4.2. An optimal set of parameters were determined for each representation to learn the inter-class and intra-class variations. For the combined classifier, learning was done by concatenating all the six representations. The test images were then projected onto the learned subspace to perform recognition. This setup was repeated ten times, by taking one of the possible ten images per person in the test set. The average recognition rate over these ten trials are reported in Table 6 below. We also compare our results with the previously reported results on this dataset, from [24] and [23], which primarily uses the intensity image of the face to learn the classifier.

It can be seen that some of the lighting invariant representations, like gradient direction [15] perform well under general imaging conditions, and the combined representation does provide improvements in the recognition rate. At the same time it seems that the self-quotient image [12] is particularly sensitive to non-lighting variations. But we would like to make a point here regarding the amount of training data used. All our previous experiments (until Section 4.2) were done with just single image per person in the training set. Our study mainly focuses on how lighting invariant a representation can be, given it sees just one image of the person in arbitrary illumination. But when there are other sources of variations in the dataset, such as expression, registration, scale and pose, we need to have multiple images of a person in the training set in order to learn a good classifier. It is an interesting future work to design classifiers capturing the contributions of different representations, to perform robust face recognition (with very few training examples) under multiple sources of variations.

5. Discussion

Besides the performance gains obtained through classifier combination, another interesting observation of this work is that a single classifier based on the direction of image gradient works very well. Throughout the experiments discussed here, the gradient direction algorithm clearly performs the best with just one exemplar per person in the gallery (with both homogenous and heterogeneous gallery lighting conditions). In order to emphasize the significance of this observation, we compare our results with two recently reported algorithms from the literature.

Fig. 13. Sample images from ORL face database [5].

5.1. Comparison with the work of Tan and Triggs [25]

First we consider the work by Tan and Triggs [25], which proposes enhanced local texture feature sets for illumination robust face recognition. They introduce Local Ternary Patterns (LTP), a generalization of the Local Binary Pattern (LBP) texture descriptor [26], and show it to be more discriminant and less sensitive to noise. They then couple this descriptor with a preprocessing step that compensates for lighting, and use a distance transform based similarity metric to obtain good recognition results. We now compare the results of gradient direction with those reported by [25].

For the experiments on the extended Yale-B dataset [4], the frontal face images with most neutral lighting sources ('A+000E+00') were used as the gallery. The probe was divided into five subsets, according to the angle between the light source direction and the central camera axis (12°, 25°, 50°, 77°, 90°), containing frontal images of all 38 subjects. The results obtained by the Tan and Triggs algorithm [25], and by using gradient direction based classifier (with L1-Norm as the distance measure) [15] are given in Table 7.

It can be seen that we obtain slightly better results using the gradient direction algorithm [15]. We then compare our results on the PIE dataset [3], wherein again, images of all 68 subjects with neutral lighting sources were used as gallery, and the remaining images were used as the probe. In this setup, we obtain the maximum possible recognition rates like [25], as shown in Table 8. Through these experiments, we observe that a simple classifier based on the image gradient orientation offers similar (and in some cases, better) recognition performance.

5.2. Comparison with the algorithm for face recognition using sparse representations [27]

Next, we consider a more recent work by Wright et al [27], using the theory of sparse representations for face recognition. The main motivation behind this work is to represent a test face image as a sparse combination of the 'most identical' images present in the training set, so that the occlusions present in the test data can be effectively factored out. The authors also illustrate the potential applications of such an approach for handling variations in illumination. They provide results for lighting invariant face recognition on the extended Yale-B dataset [4] by using sparse representation-based classification (SRC) on different sets of features including, Eigenfaces [6], Fisherfaces [7], Laplacianfaces [28], Randomfaces (obtained by performing random projections on the input faces), and downsampled faces. The gallery for their experiments contained half of the available lighting conditions (i.e. 32 per subject), with the lighting chosen randomly for different subjects. The results obtained using their best image representa-

Table 6
Recognition rates on the ORL face database [5].

Algorithm	Recognition rates
Gradient direction [15]	95.75
Eigenphases [13]	90.75
Preprocessing [11]	81.75
Self-quotient [12]	80
Whitening [14]	94.25
Correlation filters [10]	96.25
Combination	98.5
Fisherfaces [24]	98.5
ICA [24]	93.8
Eigenfaces [24]	97.5
Kernel Eigenfaces [24]	98
2DPCA [23]	98.3

Table 7
Comparing the overall recognition rates of Tan and Triggs algorithm [25] with that of Gradient direction algorithm [15] on the extended Yale-B dataset [4].

Algorithm	Subset # (number of probes)				
	1 (263) (%)	2 (456) (%)	3 (455) (%)	4 (526) (%)	5 (714) (%)
Tan and Triggs [25]	100	100	100	99.2	97.2
Gradient direction [15]	100	100	100	100	99.73

Table 8
Comparing the overall recognition rates of Tan and Triggs algorithm [25] with that of Gradient direction algorithm [15] on PIE dataset [3].

Algorithm	Recognition rates
Tan and Triggs [25]	100
Gradient direction [15]	100

Table 9
Performance of SRC based face recognition algorithm [27] on the extended Yale-B dataset.

Dimension of the face image	Recognition rate of E-random faces [27]
30	90.72
56	94.12
120	96.35
504	98.26

Table 10
Performance of gradient direction algorithm [15] on the extended Yale-B dataset.

# Heterogeneous gallery lighting per subject	Recognition rate of gradient direction algorithm [15]
1	59.1
2	75.8
4	93.5
6	98.6

tion (E-random faces), with different dimensions for the face image, is reproduced in the Table 9 below.

We now compare these results with that obtained using the direction of image gradient [15]. We used the L1-Norm to compute the distance (since it gave better performance than the L2-Norm, of about 5% improvement in the recognition rate). We varied the number of (random) lighting conditions for every subject in the gallery, and the results averaged over multiple trials are given in Table 10. The input image dimensions used for our experiment is 1920 (i.e. $48^* 40$).

The important result, as we see, from the Tables 9 and 10 is, although the input image dimensions of our experiment is higher than that of [27], the simple classifier based on the direction of image gradients [15] performs better than [27] with just six lighting conditions in the gallery (when compared with 32 in the case of [27]). Overall, the message we would like to convey from the comparisons given in Section 5.1, and Section 5.2 is that the gradient orientations [15] retain most of the person-specific information even under very challenging lighting conditions, and it is interesting to see how this information can be better utilized in dealing with more challenging face recognition settings.

6. Conclusions

We have compared a number of approaches to illumination insensitive face recognition, both experimentally and using an analysis of simple idealizations of face features. Based on all the results obtained, we make the following observations: (1) Gradient direction works very well under both homogenous gallery and heterogeneous gallery settings. We suggest that it should be a baseline algorithm for future methods, especially since it is so simple to implement. (2) The self-quotient image and gradient direction based algorithms work extremely well under homogenous gallery lighting. (3) Not all the methods that use training data perform better than simpler methods that use general image processing. This suggests that these methods do not get as much out of training as might be possible. An exception to this is the correlation filters algorithm, which offers better recognition rates than most of the class-independent algorithms, but still is not as good as the direction of gradient (even when the training data has a very good representation of different lighting conditions). (4) Different representations work well in different parts of the face. For example, the self-quotient image is less effective in the nose region, while the gradient direction performs poorly in the cheek region. We are able to explain these results using a simple idealization of facial features. (5) Consequently, it is possible to improve performance by combining different representations. We demonstrate this using two classifier combination algorithms. The first algorithm adaptively integrates information from individual classifiers on various facial regions, whereas the other learns the best combination strategy using a SVM framework. It remains an interesting topic for future work to characterize the strengths and weaknesses of these approaches when both lighting and pose or facial expression vary.

Acknowledgments

We gratefully acknowledge financial support from a Honda Research Initiation Grant and from the Office of Naval Research under the MURI Grant N00014-08-10638.

References

[1] J. Ruiz-del Solar, J. Quinteros, Illumination compensation and normalization in eigenspace-based face recognition: a comparative study of different preprocessing approaches, Pattern Recognition Letters 29 (2008) 1966–1979.
[2] A. Franco, L. Nanni, Fusion of classifiers for illumination robust face recognition, Expert Systems with Applications 36 (2009) 8946–8954.
[3] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination and expression database, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2004) 1615–1618.
[4] K. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 684–698.
[5] F. Samaria, A. Harter, Parameterisation of a stochastic model for human faceidentification, in: Proceedings of the Second IEEE Workshop on applications of Computer Vision, 1994, pp. 138–142.
[6] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, IEEE Conference on Computer Vision and Pattern Recognition (1991) 586–591.
[7] P.N. Belhumeur, J.P. Hespanah, D.J. Kreigman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 711–720.
[8] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, Journal of the Optical Society of America A 14 (1997) 1724–1733.
[9] B. Moghaddam, T. Jebara, A. Pentland, Bayesian Face Recognition, Mitsubishi Electric Research Laboratories, TR2000-42, 2002.
[10] B.V.K.V. Kumar, M. Savvides, C. Xie, Correlation pattern recognition for face recognition, Proceedings of IEEE 94 (2006) 1963–1976.
[11] R. Gross, V. Brajovic, An image preprocessing algorithm for illumination invariant face recognition, International Conference on Audio- and Video-Based Biometric Person Authentication (2003) 10–18.
[12] H. Wang, S.Z. Li, Y. Wang, Face recognition under varying lighting conditions using self quotient image, IEEE International Conference on Automatic Face and Gesture Recognition (2004) 819–824.
[13] M. Savvides, B.V.K.V. Kumar, P.K. Khosla, Eigenphases vs. eigenfaces, International Conference on Pattern Recognition (2004) 810–813.
[14] M. Osadchy, D. Jacobs, M. Lindenbaum, Surface dependent representations for illumination insensitive image comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007) 98–111.
[15] H.F. Chen, P.N. Belhumeur, D.W. Jacobs, In search of illumination invariants, IEEE Conference on Computer Vision and Pattern Recognition 1 (2000) 254–261.
[16] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Wurtz, W. Konen, Distortion invariant object recognition in dynamic link architecture, IEEE Transactions on Computer 42 (1993) 300–311.
[17] B. Jahne, Digital Image Processing, Springer-Verlag, 1995.
[18] L. Nanni, D. Maio, Weighted sub-Gabor for face recognition, Pattern Recognition Letters 28 (2007) 487–492.
[19] L. Nanni, A. Lumini, Wavelet decomposition tree selection for palm and face authentication, Pattern Recognition Letters (2007) 343–353.
[20] C. Chang, C. Lin, LIBSVM: a library for support vector machines, 2001.
[21] P. Phillips, Support vector machines applied to face recognition, Advances in Neural Information Processing Systems 11 (1998) 803–809.
[22] H. Ling, S. Soatto, N. Ramanathan, D.W. Jacobs, A study of face recognition as people age, IEEE International Conference on Computer Vision (2007) 1–8.

[23] A. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 131–137.
[24] M. Yang, Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods, in: fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings 2002, pp. 215–220.
[25] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Lecture Notes in Computer Science 4778 (2007) 168–182.
[26] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, Lecture Notes in Computer Science (2004) 469–481.
[27] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 210–227.
[28] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 328–340.