Nearest-Subspace Patch Matching for Face Recognition Under Varying Pose and Illumination*

Zihan Zhou, Arvind Ganesh, John Wright, Shen-Fu Tsai, and Yi Ma
Dept. of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
{zzhou7, abalasu2, jnwright, stsai8, yima}@uiuc.edu

Abstract

We consider the problem of recognizing human faces despite variations in both pose and illumination, using only frontal training images. We propose a very simple algorithm, called Nearest-Subspace Patch Matching, which combines a local translational model for deformation due to pose with a linear subspace model for lighting variations. This algorithm gives surprisingly competitive performance for moderate variations in both pose and illumination, a domain that encompasses most face recognition applications, such as access control. The results also provide a baseline for justifying the use of more complicated face models or more advanced learning methods to handle more extreme situations. Extensive experiments on publicly available databases verify the efficacy of the proposed method and clarify its operating range.

1. Introduction

Real-world automatic face recognition systems are confronted with a number of sources of within-class variation, including pose, expression, and illumination, as well as occlusion or disguise. Several decades of intense study within the pattern recognition community have produced numerous methods for handling each of these factors individually. For example, for frontal (fixed-pose) faces, linear subspace models have proven extremely effective for modeling variations in illumination [2, 7]. On the other hand, under ambient illumination, correspondence techniques such as SIFT [13, 15] or local deformation models [16] have proven effective in dealing with varying pose and expression.

The above are only a small sample of the vast library of techniques for handling individual sources of variation in automatic face recognition. In comparison, there has been much less work on simultaneously handling multiple modes of variability, according a recent survey [24]. Although many methods developed for dealing with one type of variability are also stable w.r.t. small variation in other types, very few work well with change in both lighting and pose even if the change is moderate.

Several techniques have been proposed to mitigate the negative effect of pose variations on 2D image-based algorithms. For example, [8, 25] assume that only a finite (small) set of possible poses can occur in the test image and treats pose determination as a classification problem. In a similar spirit, [10] utilizes probabilistic models of the representation error at each pose. This approach can also be applied at the patch level [14]. Neither [10] nor [14] explicitly deals with varying illumination, however. The method of [21] utilizes training images of different subjects to synthesize new, virtual views of the test subject at a discrete set of poses, but also does not model illumination. Moreover, these approaches are fundamentally limited by the assumption that only a discrete set of poses occur in the testing. While very fine discretization in pose could conceivably allow good recognition across a continuum of poses, the amount of training data required is prohibitive. One pure 2D method that handles moderate pose without discretization is *elastic bunch graph* matching [22], which allows small 2D deformations of a set of feature point locations. That approach is related to, but more complicated than, the local translation models that are currently state-of-the-art in recognizing handwritten digits [11].

Several methods have been proposed for incorporating limited 3D information to help cope with pose variations without resorting to a full 3D model. For example, [5] utilizes a simplified (cylindrical) model of face shape and applies correspondence techniques to cope with variations in pose. There, an objective function based on normalized cross-correlation is used to reduce the effect of illumination. Active appearance models [6, 9] and related techniques [4] model possible variations in face shape and texture within a low dimensional subspace. The algorithm of [4] finds an

*This work was partially supported by the grants NSF CAREER IIS-0347456, NSF CRS-EHS-0509151, NSF CCF-TF-0514955, ONR YIP N00014-05-1-0633, NSF ECCS07-01676, and NSF IIS 07-03756.

optimal approximation within this subspace, and also produces estimates of pose and illumination, but requires detection of 6-8 feature points per image, perhaps due to the difficulty of this global optimization problem. The feature points can be detected automatically using the algorithm proposed in [18].

Finally, at the extreme end of the 2D-3D spectrum, [7] utilizes a full 3D model of each subject's face, obtained by photometric stereo. Despite this method's success in coping with both pose and illumination on limited datasets, 2D image-based approaches are still desirable for scalability and ease of enrolling new subjects into the database.

In this paper, we investigate to what extent accurate recognition is possible using only 2D frontal images. More specifically, we address the following problem:

> *Given only frontal images taken under several illuminations, recognize faces despite large variation in both pose and illumination.*

Our algorithm will apply to test images with significantly different illumination conditions, and pose variations upto $\pm 45^\circ$.[1] In this setting, a typical test image would have an arbitrary pose in the given range and also an illumination not present in the training. Our approach is simple but effective: for each small patch of the test image, we find a corresponding location and illumination condition that best approximate it. The quality of this match is used directly as a statistic for classification, and results from multiple patches are aggregated by voting. For comparison, we also propose a second scheme that performs pose-invariant recognition between the test image and a training image synthesized from the recovered illumination conditions, by matching deformation-resistant features such as SIFT keys.

The contribution of this paper is twofold. First, we show that for certain range of pose and illumination changes, surprisingly good recognition performance can be achieved with a simple method that matches a sufficient number of 2D image patches. This range, though not too extreme, covers the kind of changes one would encounter in many applications of face recognition. Second, although more sophisticated models for 3D shape and illumination may eventually yield better performance on face recognition under varying pose and illumination, the proposed method offers a simple and easily reproducible baseline for comparison. Such a baseline system can help clarify when more complicated models and systems are truly needed and verify their performance gain.

2. Compound Effect of Illumination and Pose

In this section, we discuss the difficulties associated with variations in pose and illumination, and why state-of-the-art methods that are quite effective at handling one of these modes of variability tend to fail when both are present simultaneously.

Illumination. Images of the same face, taken at the same pose but under varying illumination lie near a low-dimensional linear subspace known as the *harmonic plane* [2] or *illumination cone* [7]. Given rectified training images $I_1 \dots I_k$ and a test image I' taken at the same pose but different illumination, we can therefore write

$$I' \approx \alpha_1 I_1 + \cdots + \alpha_k I_k. \tag{1}$$

The coefficients α can be recovered by projecting I' onto the subspace spanned by $I_1 \dots I_k$. Local deviations from linearity due to self-shadowing, non-Lambertian effects and even occlusion can be accommodated by robustly solving (1) for α, as in [23].[2] Over public face datasets, such as the Extended Yale B [7], this method achieves nearly perfect recognition performance with the frontal views.

Pose. Now if a test image I' is taken at a different pose but under the same lighting condition as a training image I, we may assume that there is some deformation of the image plane $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$I'(x) \approx I(\phi(x)). \tag{2}$$

Although ϕ is in general a highly nonlinear map, over most of the image it is smooth and hence can be locally approximated by a linear (or affine) map. In this situation, one can match local features of I' with those of I using invariant descriptors such as SIFT keys [13]. For instance, based on matching the SIFT keys, for pose variation up to $\pm 45^\circ$, one can achieve nearly perfect recognition on the PIE face database [20] under the same (ambient) illumination.

Combined effect of pose and illumination. Test images taken under different pose *and* different illumination than the training images present a difficult, coupled problem. When applied individually, both of the above approaches fail. This difficulty is illustrated in Figure 1. In Figure 1 (top), the most salient SIFT keypoints detected in a test image taken at different illumination do not match well at all with those detected in the training images; while in Figure 1 (bottom) because the test image is taken at a different pose, the coefficients α found by a global linear projection do not faithfully reproduce its illumination. In the next section, we will address this coupling by simultaneously matching and solving for the illumination.

[1] Arguably, beyond 45°, 3D information becomes more critical, due to extreme occlusion and deformation of the image domain.

[2] That work introduces an error term E, so that $I' = \alpha_1 I_1 + \cdots + \alpha_k I_k + E$, and jointly solves for E and α by minimizing an ℓ^1-norm.

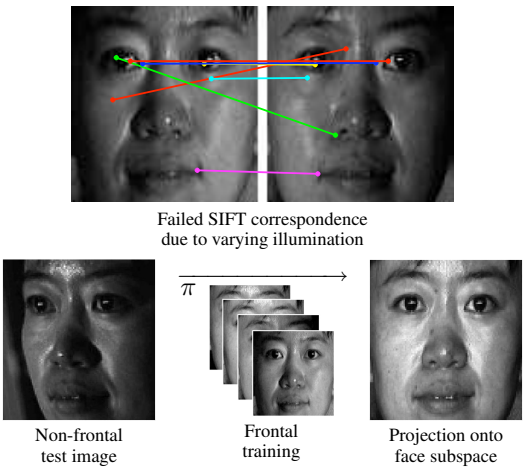


Figure 1. *Difficulty of recognition under varying pose and illumination.* Top: under varying illumination, matching techniques such as SIFT fail. Bottom: projecting a 45° test image onto the span of frontal training samples does not produce an image with similar illumination.

3. Recognition by Matching Image Patches

We assume that the face is well-localized in the test image, i.e., any gross misalignment or scaling has already been compensated for. In practice, this is accomplished by applying an affine or projective transformation to approximately map eye and mouth corners of the test image to eye and mouth corners of training image. We will denote the so-aligned test image by $I' \in \mathbb{R}^{w \times h}$. We will assume access to a gallery of registered frontal training images $I_1^{(i)} \ldots I_k^{(i)} \in \mathbb{R}^{m \times h}$ for each subject i, taken at k distinct illuminations.[3]

Matching image patches. We first select a set of feature point locations $x'_1 \ldots x'_M$ in the test image for matching. In Section 4, we will see that the choice of feature points is not essential – a sufficient number of randomly selected or evenly distributed feature points work as well as the scale-space extrema popular in the image matching literature [12, 13]. Around each feature point $x'_m \in \mathbb{R}^2$ in the test image, we select a square window $I'(W(x'_m))$ of pixels for matching. If this feature is also visible in the training (frontal) view, there is a corresponding point $x_m \in \mathbb{R}^2$ in each training image $I_j^{(i)}$. If all images were taken under ambient illumination, we would expect the corresponding patches to be quite similar in appearance: $I'(W(x'_m)) \approx I_1^{(i)}(W(x_m))$. However, if the test image is taken at more extreme illumination, it is more appropriate to approximate the patch $I'(W(x'_m))$ by a linear combination of training patches:

$$I'(W(x'_m)) \approx \alpha_1 I_1^{(i)}(W(x_m)) + \cdots + \alpha_k I_k^{(i)}(W(x_m)).$$

Now, since gross misalignments and scalings have already been compensated for, we may assume that the corresponding point lies in a small neighborhood of x'_m: $x_m \in N(x'_m) \subset \mathbb{R}^2$. For each training subject i, we search for the best match within this neighborhood, seeking a point $x_m^{(i)}$ where the subspace spanned by the training patches offers the best approximation to $I'(W(x'_m))$:

$$(x_m^{(i)}, \alpha_m^{(i)}) = \underset{\substack{x \in N(x'_m) \\ \alpha \in \mathbb{R}^k}}{\operatorname{argmin}} \left\| I'(W(x'_m)) - \sum_{j=1}^{k} \alpha_j I_j^{(i)}(W(x)) \right\|_2^2. \tag{3}$$

Thus, for each feature x'_m, $x_m^{(i)}$ denotes the location of the best match to a linear combination of training patches for subject i, while $\alpha_m^{(i)}$ denotes the corresponding best coefficients.

Classification based on Nearest Subspace matching. We compare two methods for determining the identity of the test image from the correspondences obtained from (3). The first, which we refer to as Nearest Subspace Patch Matching (NSPM), uses the approximation error in (3) as a statistic for classification. The m-th test patch $I'(W(x'_m))$ is classified to the subject i whose coefficients $\alpha_m^{(i)}$ best approximate it. The classifications of these patches are then aggregated into a single classification of the test image by voting.[4] This process is summarized as Algorithm 1 above. The algorithm is computationally efficient and scalable – the complexity is linear in the number of subjects. A similar scheme has achieved state-of-the-art performance for recognizing handwritten digits [11]. In Section 4, we will see that this simple approach also performs quite well for face recognition with moderate variations in pose and illumination.

Classification based on SIFT matching. For purposes of comparison, we also outline a second approach which uses the computed α to compensate for global illumination variations, and then applies standard matching techniques. This approach finds for each training subject i, a single illumination condition (expressed as coefficients $\alpha^{(i)}$) that best reproduces the test illumination. These coefficients are used to synthesize a set of "equivalent" frontal training images $\hat{I}^{(i)}$. In the next paragraph, we describe how the $\hat{I}^{(i)}$ can be robustly computed from noisy data. With illumination compensated for in this manner, classical deformation-invariant

[3] Generally, between $k = 3$ and $k = 9$ training images per subject suffice, depending on how extreme the test illumination may be. We will investigate this issue further in Section 4.

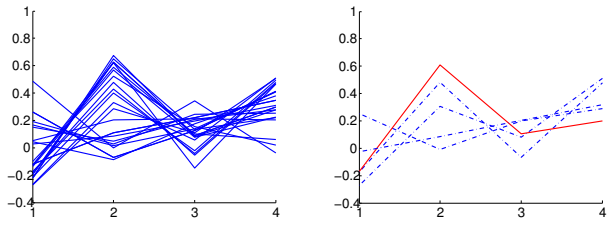
[4] Notice that instead of voting, one could also classify to the subject that minimizes the sum (over all patches) of the residuals. However, empirically, we find the voting approach is more robust to local deviations from linearity, and gives better overall performance.

Algorithm 1 (Nearest-Subspace Patch Matching).

1: **Input:** test image I', frontal training images $I_1^{(i)} \dots I_k^{(i)}$ taken at k different illuminations, for each subject i.
2: Select feature points $x'_1 \dots x'_M$ in I'.
3: **for** each feature point x'_m **do**
4: for each subject i, compute the corresponding location $x_m^{(i)}$ and coefficients $\alpha_m^{(i)}$ from (3);
5: set $\text{Identity}(x'_m)$ to be:

$$\arg\min_i \|I'(W(x'_m)) - \sum \alpha^{(i)}(j) I_j(W(x_m^{(i)}))\|_2^2;$$

6: **end for**
7: **Output:** class i that maximizes the number of x'_m with $\text{Identity}(x'_m) = i$.



(a) (b)

Figure 2. K-means clustering: (a) coefficient vectors α recovered by matching a test image at -45° pose and with illumination f13 to a linear combination of four frontal training images (f10, f13, f20, f8) from the PIE database. (b) 5 cluster centers, the one chosen for fitting the global illumination model is plotted in solid line.

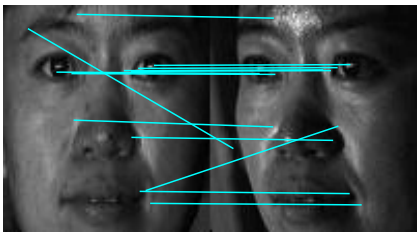


Figure 3. Matching SIFT keypoints between the same test image as in Figure 1 with a front view generated with the estimated α. In this case, there are only two mismatched SIFT keypoints.

matching algorithms can then be applied. In practice, we find that once illumination has been corrected, the SIFT algorithm performs quite successfully in matching the test image I' to the synthetic image $\hat{I}^{(i)}$ for the correct subject. The number of SIFT correspondences between I' and each of the $\hat{I}^{(i)}$ provides a simple statistic for classification.

Equation (3) gives, for each training subject i, a set of putative correspondences $x'_1 \leftrightarrow x_1^{(i)} \dots x'_M \leftrightarrow x_M^{(i)}$, and coefficient vectors $\hat{\alpha}_1^{(i)} \dots \hat{\alpha}_M^{(i)}$. For each subject i, we will synthesize a frontal training image $\hat{I}^{(i)}$ whose illumination best approximates the illumination in the test image. That is, from each set of coefficient vectors $A^{(i)} = \{\alpha_1^{(i)} \dots \alpha_M^{(i)}\}$ we will compute a single $\alpha^{(i)}$ and set $\hat{I}^{(i)} = \sum_{j=1}^k \alpha^{(i)}(j)\, I_j^{(i)}$. If the correspondences were perfect, and there were no deviations from linearity, one could simply set $\alpha^{(i)}$ to be the average of $A^{(i)}$. However, due to self-shadowing or specularity, this set is likely to contain many gross outliers (see Figure 2 for a typical set of α estimated from matching patches between two images). We therefore first remove these outliers by K-means clustering (K = 5 in all our experiments) and calculate the mean of each cluster. We then synthesize K frontal views of the face with illumination given by the K cluster means. For each synthesized image, we count its SIFT correspondences with the testing image, and only retain the cluster mean with the largest number of correspondences as our estimate of $\alpha^{(i)}$. Notice that this use of K-means for outlier rejection differs from the use in [3] where clustering is used to fit local illumination models to different parts of the face. Here we use clustering to remove outliers in fitting a single, global illumination model. With illumination compensated for, conventional feature matching techniques can be effectively applied to the test image and the synthesized training image. Figure 3 visualizes these matches for one typical example.

4. Experiments

In this section, we first investigate the effect of patch size and keypoint selection on our algorithms using small-scale experiments on the CMU PIE database [20]. Using the parameters suggested by these small experiments, we then conduct extensive large-scale recognition experiments using three publicly available databases: CMU PIE [20], Extended Yale B [7], and FERET [17]. The running time of our algorithm is proportional to the number of subjects in the database. For CMU PIE with 68 subjects, our unoptimized Matlab implementation takes 30s per test image.

We prealign all images by manually selecting the outer eye and mouth corners and then applying an affine or projective transform to map these to a canonical configuration. The resulting image is always converted to a resolution of 105×100 (see Figure 6 for an example). While this transformation accounts for global misalignment and scaling, significant distortion due to pose remains. At this resolution, upto $\pm 45^\circ$ of pose, feature points are generally displaced by less than ± 5 pixels. We therefore use a search window of size 11×11 throughout our experiments.

Effect of keypoint selection. We first investigate the sensitivity of Nearest Subspace Patch Matching to the location of the keypoints used for matching. We test three schemes: the first selects keypoints as local extrema of a scale-space Difference of Gaussians (DoG) [13, 12], the second selects keypoints on a regular grid, while the third chooses key-

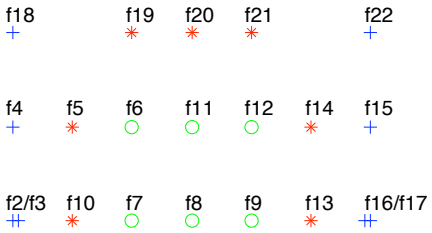


Figure 4. Grouping and relative locations of the 21 flashes for the PIE database. Group 1: o. Group 2: *. Group 3: + .

	DoG extrema	Grid	Random
Avg. rec. rate	86.5%	**96.5%**	89.4%

Table 1. Recognition rate of NSPM for various choices of keypoints.

points uniformly at random. We test these approaches using the CMU PIE database, which contains 21 illumination conditions. The layout of these light sources is visualized in Figure 4. For this experiment, the training consists of six frontal images per subject, taken with illuminations f3, f8, f10, f13, f16, and f20. The testing consists of images taken at 45° pose, with 11 different illuminations (f6-13 and f19-21). For these images, the average number of DoG extrema is 45. For the grid scheme, we choose an arrangement of $7 \times 6 = 42$ evenly spaced patches, each of size 22×22 pixels. For the random sampling scheme, we choose 40 keypoints from a uniform distribution over the image. Table 1 gives the recognition rates for Algorithm 1 for each of these three schemes. The grid scheme achieves the highest recognition rate, 96.5%, more than 7% better than the others, suggesting that even coverage of the face is more important than choosing particularly distinctive keypoints.

Effect of patch size. There is clearly a trade-off between expressiveness of the patch and quality of the local deformation model: with larger patches, the solution for α is better-conditioned, while for smaller patches the deformation is closer to a pure translation. We investigate this effect using the same six training illuminations as in the previous experiment. The testing consists of images taken at 45° pose, with the same 11 illuminations as in the previous experiment. We use 80 patches arranged on a grid, and vary the size of the patches from 6×6 to 34×34, in increments of 4. Figure 5 plots the recognition rate as a function of patch size. The highest recognition rate, 98%, is achieved with 18×18 and 22×22 patches, each of which covers roughly 4% of the face. As expected, the performance degrades when the patch is too small or large.

Settings for large-scale recognition experiments. From the above settings, we always select feature points on the grid, with patch size 22×22. The size of the search window for NSPM is 11×11. The number of patches we select is between 80 and 100, depending on the dataset.

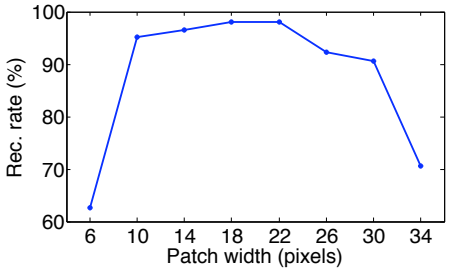


Figure 5. Recognition rate of NSPM as a function of patch size.

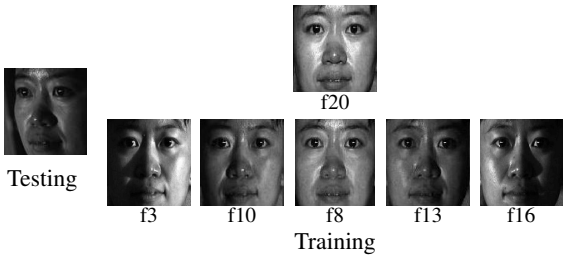


Figure 6. Experimental setup. Left: test image at -45° pose. Right: training images under 6 different illuminations.

Experiment 1: PIE database. We next test our algorithm using all 21 illuminations in the PIE database (f2 to f22) and all poses within $\pm 45^\circ$ (c27, c29, c11, c05, c37, c07 and c09). For this experiment, we use 80 patches on a 10×8 grid. The most important remaining issue is the choice of training illuminations. Estimates of the number of training illuminations needed to span the illumination subspace vary between 3 [19] and 9 [2, 7]. For our algorithm, we find that using a set of 6 illuminations (see Fig. 6) is sufficiently expressive while avoiding overfitting. Overfitting is a potential issue here because we are recovering the illumination from small patches; if too many training illuminations are present, the problem becomes ill-conditioned.

The 21 illuminations in the PIE database are visualized in Figure 4. For our training set, we select frontal images taken at illuminations f3, f8, f10, f13, f16, and f20 (see Fig. 6). We divide the 21 flashes into 3 groups based on their location relative to the five directional illuminations in the training set. Compared to the training, the illumination conditions in Group 1 are milder, while those in Group 3 are significantly more extreme.

Table 2 presents a detailed breakdown of the recognition performance of Nearest Subspace Patch Matching, in terms of pose and illumination condition. For moderate pose ($\leq 22.5^\circ$) and illumination (Groups 1 and 2), the algorithm performs perfectly. Even for more extreme illuminations (in Group 3), if the pose remains moderate, the algorithm still performs quite well. The algorithm breaks down only when both the lighting and pose are very extreme (i.e., Group 3 with 45° pose). Interestingly, some poses and illuminations combine to give a more difficult experimental

Pose	0°	-22.5°	$+22.5^\circ$	$\downarrow$	$\uparrow$	-45°	$+45^\circ$
	c27	c05	c29	c07	c09	c37	c11
Illum. Group 1							
6	100	100	100	100	100	98.5	98.5
7	100	100	100	100	100	97.1	100
8	100	100	100	100	100	98.5	98.5
9	100	100	100	100	100	97.1	98.5
11	100	100	100	100	100	97.1	100
12	100	100	100	100	100	95.6	95.6
Group 2							
5	100	100	100	100	100	88.2	95.6
10	100	100	100	100	100	100	94.1
13	100	100	100	100	100	95.6	100
14	100	100	100	100	100	94.1	89.7
19	100	100	100	100	100	97.1	100
20	100	100	100	100	100	95.6	100
21	100	100	100	100	100	95.6	92.7
Group 3							
2	100	100	100	100	100	61.8	55.9
3	100	100	100	100	100	86.8	76.5
4	100	98.5	100	100	100	64.7	86.8
15	100	100	92.7	100	100	94.1	60.3
16	100	100	95.6	100	100	85.3	73.5
17	100	100	91.2	100	100	67.7	48.5
18	100	98.5	100	100	100	77.9	92.7
22	100	100	100	100	100	92.7	75.0

Table 2. Breakdown of recognition rate of NSPM in terms of pose and illumination on the PIE database. $\uparrow$ and $\downarrow$ represent the upward and downward poses, approximately 20-25° from the frontal pose.

Pose	0°	-22.5°	$+22.5^\circ$	$\downarrow$	$\uparrow$	-45°	$+45^\circ$
	c27	c05	c29	c07	c09	c37	c11
Group 1							
NSPM	100	100	100	100	100	97.3	98.5
SIFT	100	100	100	100	100	96.3	97.3
[5]	99.8	98.2	N/A	N/A	N/A	N/A	N/A
Group 2							
NSPM	100	100	100	100	100	95.2	96.0
SIFT	100	100	100	100	100	89.7	89.1
[5]	96.7	97.6	N/A	N/A	N/A	N/A	N/A
Group 3							
NSPM	100	99.6	97.4	100	100	78.9	71.2
SIFT	99.5	90.1	89.5	96.3	96.3	48.5	49.6
[5]	89.9	81.2	N/A	N/A	N/A	N/A	N/A

Table 3. Average recognition rates for NSPM, SIFT, and [5] on PIE database. All algorithms perform well for mild conditions, but NSPM generalizes much more effectively to extreme illuminations.

condition. For example with f5 and -45° pose, the illumination comes from the far right, but the image is taken from the far left, yielding a more extreme effective illumination. This explains the relatively lower performance of e.g., f5 with -45°, f14 with $+45^\circ$, f2 with -45°, and others.

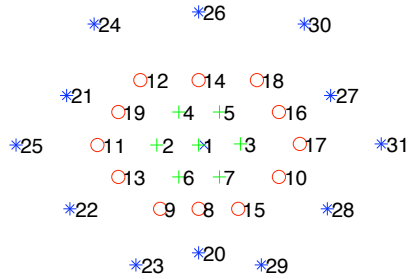


Figure 7. Grouping and relative locations of illuminations in the Extended Yale B database. Subset 1: +. Subset 2: o. Subset 3: * .

Table 3 gives the average recognition performance across the three groups. For comparison, it also gives the recognition rate for the SIFT matching scheme outlined in Section 3.[5] Notice that for mild illuminations, this scheme also performs reasonably well. However, for more extreme poses and illumination conditions, NSPM significantly outperforms SIFT. For example, in Group 3 with $\leq 22.5^\circ$ pose, NSPM remains nearly perfect, while SIFT's performance drops by as much as 10%.

Table 3 also compares our results to those reported in [5], which reports better results than comparable techniques such as [8].[6] That work uses less training data than ours (only a single image per class), but a more complicated algorithm (stereo matching to a cylindrical face model) to recognize faces across varying pose. From Table 3, NSPM consistently outperforms the results of [5], with the greatest gain for more extreme illuminations. For example, with -22.5° pose and illuminations in Group 3, NSPM achieves an average recognition rate of 99.6%, compared to 81.2% for [5]. This difference is most likely due to the lack of an explicit illumination model in [5], which instead relies on the use of normalized cross-correlation in the matching process. Finally, note that for test poses $\leq 45^\circ$, [10] reports perfect accuracy across a subset of 34 subjects from the PIE database. That work learns a model for patch mismatch at a (small) discrete set of poses. It does not cope with illumination variations; the training and test set contain only a single frontal illumination. From Table 3, NSPM achieves similar recognition accuracy across a larger test set, under more challenging illumination conditions.

Experiment 2: Extended Yale B database. We next test our algorithm on the Extended Yale B database [7]. This dataset contains images of 38 subjects under varying pose

[5] We use David Lowe's toolbox for SIFT matching, downloaded from `http://www.cs.ubc.ca/ lowe/keypoints/`. We have experimented with many different choices of training sets and parameters for the SIFT matching scheme. Table 3 reports the best results we have achieved with this technique. Interestingly, SIFT achieves its best performance with a smaller training set (f8, f10, f13, f20), suggesting that NSPM better utilizes the information contained in the additional training images.

[6] The result for [5] for Group 2, pose c27 does not include illumination f12, which was used as training.

Testing	12° pose		24° pose	
	NSPM	SIFT	NSPM	SIFT
Subset 1 and 2	98.7	98.7	91.2	93.3
Subset 3	96.8	95.1	80.8	85.4

Table 4. Average recognition rate on Extended Yale B database.

and illumination.[7] The illuminations are more challenging than those PIE, but the variation in pose is less severe (between 0 and 24 degrees). We use Subsets 1 (mild illumination), 2 (moderate illumination), and 3 (extreme illumination) for this experiment. These illuminations are visualized and indexed in Figure 7. Here, we use 80 patches, located on a 10×8 grid. We use frontal images (pose 1) take at 5 illuminations (1,8,14,11,17) from Subsets 1 and 2 for the training set, and test on the 12° and 24° poses. (poses 4 and 8, respectively)

Table 4 gives the result for NS, as well as for the SIFT matching scheme outlined in Section 3. At 12° pose, both algorithms perform quite well, with NSPM slightly outperforming SIFT on more extreme illuminations. For 24° pose and moderate illumination, the recognition rates remain above 90%. However, on this dataset, SIFT slightly outperforms NSPM for more extreme pose and illumination. This appears to be due to the presence of more severe shadowing in the Yale database: whereas the shadowed regions are rejected by the K-means step in the SIFT matching algorithm, these regions still contribute to the voting step in NSPM, potentially leading to misclassifications. Nevertheless, we will see in the next experiment that NSPM performs much more competitively under more realistic lighting conditions and with larger databases.

Experiment 3: FERET database. Finally, we test our algorithms on a portion of the FERET database [17]. This dataset contains far more subjects than Extended Yale B or PIE, and the images are taken under less controlled settings. It therefore provides a better test of the scalability of our approach, as well as its applicability to real face recognition scenarios. Here, we use 100 patches per image on a 10×10 grid.

We select a set of gray-level images of 200 subjects (labeled *ba*, *bc*, *bd*, *bk*, *be* and *bf*). Except for *bk*, all images are taken with identical (ambient) illumination. We experiment with two different training sets, one with two frontal images per subject (*ba* and *bk*) and one with only *ba*.[8] Table 5 summarizes the two algorithms' performance with test sets *bd* (25° pose) and *bc* (40° pose). Algorithm 1 performs quite well (95.5% recognition rate) in the more mild setting. Notice that Algorithm 1 outperforms SIFT by more than 10% in all cases, suggesting that the voted subspace patch classifications are a more discriminative statistic than the number of SIFT correspondences for classification across large databases.

[7] In this experiment, two subjects (B02 and B16) are discarded due to corrupted training images, leaving a total of 36 subjects.

[8] With one training image per person, NSPM reduces to performing nearest neighbor with cosine (angular) distance on each test patch, followed by voting.

Training	Testing	Pose	Rec. rate(%)	
			NSPM	SIFT
ba, bk	*bd*	$+25^\circ$	95.5	77
ba	*bd*	$+25^\circ$	92	80
ba, bk	*bc*	$+40^\circ$	59.5	38
ba	*bc*	$+40^\circ$	57.5	38.5

Table 5. Performance of NSPM (Algorithm 1) and SIFT matching on a set of 200 subjects of the FERET database.

Training	Testing	Pose	Rec. rate(%)	
			NSPM	[4]
ba	*be*	$+15^\circ$	98.5	99.5
ba	*bf*	-15°	99.0	97.4
ba	*bd*	$+25^\circ$	91.8	96.9
ba	*bc*	$+40^\circ$	57.7	95.4

Table 6. Comparison of NSPM (Algorithm 1) with results reported in [4] on a set of 194 subjects of the FERET database. Images *ba* are used as the training for all experiments.

We also compare our results on the FERET database to those reported in [4], which uses 3D morphable models to achieve very high recognition rates. The experiments reported in [4] use a subset of 194 subjects (01013 - 01206) from the FERET database, with only one training image per person (*ba*). We test Algorithm 1 with the same setting. The results are presented in Table 6, along with results reported in [4]. Again, Algorithm 1 performs very competitively for moderate (up to 25°) pose changes, suggesting that it offers a much faster, simpler alternative to 3D models if the pose is moderate. Notice, though, that the performance degrades as the pose becomes more extreme, suggesting that 3D information, or more expressive deformation models are necessary in this circumstance.

5. Conclusion and Discussion

In this paper, we have examined a simple, scalable approach to face recognition under moderate variations in pose and illumination. Unlike previous approaches that treat illumination as a nuisance factor to be removed by preprocessing, our algorithm directly tackles pose and illumination simultaneously, by matching to a linear combination of frontal training images rather than just a single training image. The experimental results of Algorithm 1 on all three face databases confirm that

> *Under moderate variation in both pose (upto 25° in all directions) and illumination (within the span of the training images), simply matching sufficiently many 2D patches as a linear combination of frontal training images is competitive with*

more advanced methods. [9]

Moderate variation in pose and illumination is arguably the most common situation encountered in real-world applications of face recognition. Furthermore, the experiments on the PIE database suggests that if there are sufficient illuminations in the training and the variation in the testing is moderate (Groups 1 and 2), the above method even works well with pose variation up to $\pm 45^\circ$. These findings resemble results on handwritten digit recognition reported in [11], where a very simple patch deformation model effectively handles large within-class deformations, yielding state-of-the-art performance.

Nevertheless, this simple scheme does start to break down if both pose varies beyond 25° and illumination lies outside the span of the training images. Only in those more extreme situations do more complex models for lighting, pose, deformation, or 3D face shape and more advanced learning methods appear to become necessary.

Our findings in this paper also justify the use of linear subspace models for face recognition under moderate pose and illumination variation. The speed and performance of the basic Algorithm 1 can be further improved if one uses the faster nearest subspace algorithm [1] or more robust classification method for linear regression models [23]. This remains an important direction for future investigation. We are also investigating whether this method can be extended to more extreme poses by including some images at different poses in the training.

References

[1] R. Basri, T. Hassner, and L. Zelnik-Manor. Approximate nearest subspace search with applications to pattern recognition. In *CVPR*, pages 1–8, 2007.

[2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[3] A. Batur and M. Hayes. Segmented linear subspaces for illumination-robust face recognition. *IJCV*, 57(1):49–66, 2004.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *PAMI*, 25(9), 2003.

[5] C. Castillo and D. Jacobs. Using stereo matching for 2-D face recognition across pose. In *CVPR*, pages 1–8, 2007.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001.

[7] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6), 2001.

[8] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *PAMI*, 26(4):449–465, 2004.

[9] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(11):1080–1093, 2006.

[10] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach towards pose-invariant face recognition. In *proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*, volume 2, pages 954–959, 2003.

[11] D. Keysers, T. Deselaers, C. Gollan, and H. Ney. Deformation models for image recognition. *PAMI*, 29(8):1422–1435, 2007.

[12] T. Lindeberg. Scale-space theory: a basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2), 1994.

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[14] S. Lucey and T. Chen. Learning patch dependencies for improved pose mismatched face verification. In *CVPR*, 2006.

[15] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. Lu. Person-specific SIFT features for face recognition. In *proc. ICASSP*, volume 2, pages 593–596, 2007.

[16] F. Perronin, J. Dugelay, and K. Rose. A Probabilistic Model of Face Mapping with Local Transformations and its Application to Person Recognition. *PAMI*, 27(7):1157–1171, 2005.

[17] P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face recognition algorithms. *PAMI*, 22(10):1090–1104, 2000.

[18] S. Romdhani and T. Vetter. 3D probabilistic feature point model for object detection and recognition. In *CVPR*, pages 1–8, 2007.

[19] A. Shashua. On photometric issues in 3D visual recognition from a single 2D image. *IJCV*, 21(1-2):99–122, 1997.

[20] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination and expression database. *PAMI*, 25(12):1615–1618, 2003.

[21] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *PAMI*, 19(7), 1997.

[22] L. Wiskott, J. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *PAMI*, 19(7), 1997.

[23] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. to appear in *PAMI*, 2008.

[24] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.

[25] S. Zhou and R. Chellappa. Image-based face recognition under illumination and pose variations. *Journal of the Optical Society*, 22:217–229, 2005.

[9] such as those using deformation-invariant image descriptors [13], illumination-invariant objective functions [5], pose statistics [10], or 3D face models [7, 4].