To Appear in the IEEE Transactions on Pattern Analysis and Machine Intelligence

Appearance-Based Face Recognition and Light-Fields

Ralph Gross, Iain Matthews, and Simon Baker

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Arguably the most important decision to be made when developing an object recognition algorithm is selecting the scene measurements or *features* on which to base the algorithm. In appearance-based object recognition the features are chosen to be the pixel intensity values in an image of the object. These pixel intensities correspond directly to the radiance of light emitted from the object along certain rays in space. The set of all such radiance values over all possible rays is known as the *plenoptic function* or *light-field*. In this paper we develop a theory of appearance-based object recognition from light-fields. This theory leads directly to an algorithm for face recognition across pose that uses as many images of the face as are available, from one upwards. All of the pixels, whichever image they come from, are treated equally and used to estimate the (eigen) light-field of the object. The *eigen light-field* is then used as the set of features on which to base recognition, analogously to how the pixel intensities are used in appearance-based face and object recognition.

Keywords: Appearance-based object recognition, face recognition, light-fields, eigen light-fields, face recognition across pose.

1 Introduction

Arguably the most important decision to be made when developing an object recognition algorithm is selecting the scene measurements or *features* on which to base the algorithm. One of the most successful and well-studied approaches to object recognition is the *appearance-based* approach. Although the expression "appearance-based" was introduced by Murase and Nayar [19], the approach itself dates back to Turk and Pentland's *Eigenfaces* [28] and perhaps before [27]. The defining characteristic of appearance-based algorithms is that they directly use the pixel intensity values in an image of the object as the features on which to base the recognition decision.

The pixel intensities that are used as features in appearance-based algorithms correspond directly to the radiance of light emitted from the object along certain rays in space. Although there may be various non-linearities caused by the optics (e.g. vignetting), the CCD sensor itself, or by gamma correction in the camera, the pixel intensities can be thought of as approximately equivalent to the radiance of light emitted from the object in the direction of the pixel.

The *plenoptic function* [1] or *light-field* [13, 17] specifies the radiance of light along all rays in the scene. Hence, the light-field of an object is the set of all possible features that could be used by an appearance-based object recognition algorithm. It is natural, therefore, to investigate using light-fields (as an intermediate representation) for appearance-based object recognition. In the first part of this paper we develop a theory of appearance-based object recognition from light-fields. In the second part we propose an algorithm for face recognition across pose based on an algorithm to estimate the (eigen) light-field of a face from a set of images.

1.1 Theoretical Properties of Light-Fields for Recognition

There are a number of important theoretical questions pertaining to object recognition from light-fields. Some examples are:

1. The fundamental question "what is the set of images of an object under all possible illumination conditions?" was recently posed and answered in [5]. Because an image simply consists of a subset of measurements from the light-field, it is natural to ask the same question about the set of all light-fields of an object. Answering this second question may also

help understand the variation in appearance of objects across both pose and illumination.

2. "When can two objects be distinguished from their images?" is perhaps the most important theoretical question in object recognition. Various attempts have been made to answer it in one form or another. For example, it was shown in [4] that, given a pair of images, there is always an object that could have generated those two images (under different illuminations.) Similarly one might ask "when can two objects be distinguished from their light-fields?"

In the first part of this paper we derive a number of fundamental properties of object light-fields. In particular, we first investigate the set of all possible light-fields of an object under varying illumination. Amongst other things we show that the set of all light-fields is a convex cone, analogously to the results in [5] for single images. Afterwards we investigate the degree to which objects are distinguishable from their light-fields. We show that, under arbitrary illumination conditions, if two objects have the same shape they cannot be distinguished, even given their light-field. The situation for objects with different shapes is different however. We show that two objects can almost always be distinguished from their light-fields if they have different shapes.

1.2 Face Recognition Using Light-Fields

One implication of this theory is that "appearance-based" object recognition from light-fields is theoretically more powerful than object recognition from single images. Capturing an entire light-field is normally not appropriate for object recognition however; it requires either a large number of cameras, a great deal of time, or both. This does not mean that it is impossible to use light-fields in practical object recognition algorithms. In the second part of this paper we develop an algorithm for face recognition across pose that is based on an algorithm to estimate the (eigen) light-field of an object from an arbitrary collection of images [14]. (This algorithm is closely related to an algorithm for dealing with occlusions in the eigen-space approach [7, 16].) The eigen light-field, once it has been estimated, is then used as an enlarged set of features on which to base the face recognition decision. Some of the advantageous properties of this algorithm are as follows:

1. Any number of images can be used, from one upwards, in both the training (gallery) and the test (probe) sets. Moreover, none of the training images need to have been captured from the

same pose as any of the test images. For example, there might be two test images for each person, a full frontal view and a full profile, and only one training image, a half profile. In this way, our algorithm can perform "face recognition across pose."

2. If only one test or training image is available, our algorithm behaves "reasonably" when estimating the light-field. In particular, we prove that the light-field estimated by our algorithm correctly re-renders images across pose (under suitable assumptions about the objects.) We also validate this re-rendering empirically.

3. If more than one test or training image is available, the extra information (including the implicit shape information) is incorporated into a better estimate of the light-field. The final face recognition algorithm therefore performs better with more input images.

4. It is straightforward to extend our algorithm to perform "face recognition across both pose and illumination," as we showed in [15].

1.3 Paper Overview

We begin in Section 2 by introducing object light-fields and deriving some of their fundamental properties. We continue in Section 3 by describing eigen light-fields and their use in our algorithm for face recognition across pose. In extensive evaluations we compare the performance of eigen light-fields to standard face recognition algorithms on both the FERET [22] and CMU PIE [26] databases. We conclude in Section 4 with a summary and a discussion.

2 Object Light-Fields and Their Properties for Recognition

2.1 Object Light-Fields

The *plenoptic function* [1] or *light-field* [17] is a function which specifies the radiance of light in free space. It is usually assumed to be a 5D function of position (3D) and orientation (2D). In addition, it is also sometimes modeled as a function of time, wavelength, and polarization, depending on the application in mind. Assuming that there is no absorption or scattering of light through the air [20], the light-field is actually only a 4D function, a 2D function of position defined

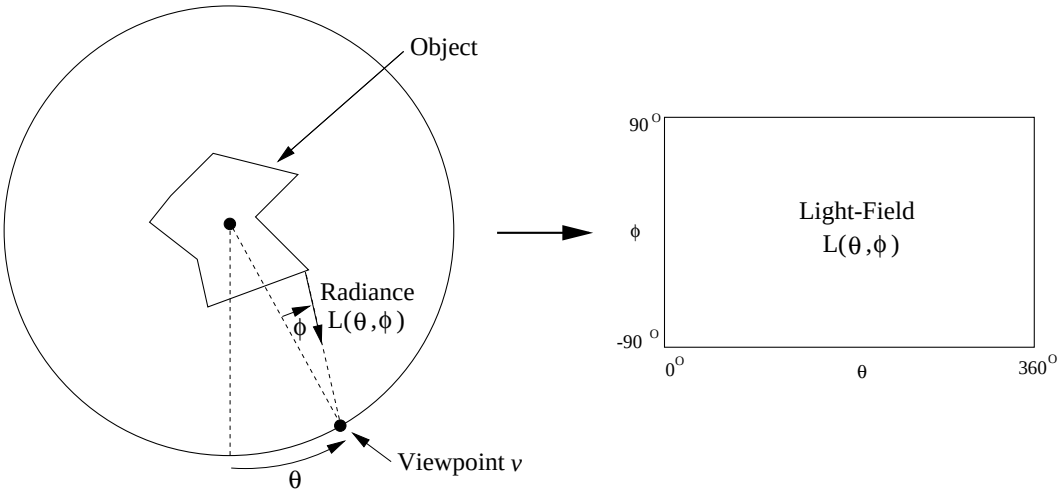


Figure 1: An illustration of the 2D light-field [17] of a 2D object. The object is conceptually placed within a circle. The angle to the viewpoint v around the circle is measured by the angle θ and the direction that the viewing ray makes with the radius of the circle by ϕ. For each pair of angles θ and ϕ, the radiance of light reaching the viewpoint is denoted $L(\theta, \phi)$, the *light-field* [17]. Although the light-field of a 3D object is actually 4D, we will continue to use the 2D notation of this figure for ease of explanation.

over a 2D surface, and a 2D function of direction [13, 17]. In 2D, the light-field of a 2D object is only 2D. See Figure 1 for an illustration of the 2D light-field of a 2D object.

2.2 The Set of All Light-Fields of an Object Under Varying Illumination

The fundamental question "what is the set of images of an object under all possible illumination conditions?" was recently posed and answered by Belhumeur and Kriegman [5]. We begin our analysis by asking the analogous question for light-fields. Since an image just consists of a subset of the rays in the light-field, it is not surprising that the same result also holds for light-fields:

Theorem 1 *The set of n-pixel* light-fields *of any object, seen under all possible lighting conditions, is a convex cone in $\mathbf{R}^n$.*

This result holds for any object, even if the object is non-convex and non-Lambertian. As pointed out in [5], the proof is essentially a trivial combination of the additive property of light and the fact that the set of *all* illumination conditions is itself a *convex cone.* For this reason, the same result holds for any subset of illumination conditions that is a convex cone. One example is an arbitrary number of point light sources at infinity. It is straightforward to show that this subset of illumination conditions is a convex cone and therefore that the following theorem also holds:

Table 1: A comparison of image illumination cones and light-field illumination cones. The main point to note is that in three of the four cases, the light-field illumination cone is a "smaller" subset of the set of all light-fields than the corresponding image illumination cone is a subset of the set of all images.

	Image Illumination Cone	Light-field Illumination Cone
Arbitrary Illumination Conditions Any Convex Object	*Always exactly equals* the set of all images	*Can sometimes be* the set of all light-fields
Arbitrary Illumination Conditions Convex Lambertian Object	*Always exactly equals* the set of all images	*Never is* the set of all light-fields
Point Light Sources at Infinity Any Convex Object	*Can sometimes be* full-dimensional	*Can sometimes be* full-dimensional
Point Light Sources at Infinity Convex Lambertian Object	*Can sometimes be* full-dimensional	*Never is* full-dimensional

Theorem 2 *The set of n-pixel* light-fields *of any object, illuminated by an arbitrary number of point light sources at infinity, is a convex cone in* $\mathbf{R}^n$.

These results are analogous to those in [5]. Moreover, since Theorems 1 and 2 clearly also hold for any subset of rays in the light-field, the analogous results in [5] are special cases of these theorems.

When we investigate the nature of the illumination cones in more detail, however, we find several differences between images and light-fields. Some of the differences are summarized in Table 1. If we consider arbitrary illumination conditions and any convex object, the image illumination cone *always exactly equals the set of all images* because every point on the object can be illuminated independently and set to radiate any intensity. This result holds for any reflectance function. The only minor requirement is that no point on the object has zero reflectance.

The situation is different for light-fields. It is possible to choose reflectance functions for which the light-field illumination cone is equal to the set of all light-fields. One simple example is to use a "mirrored" object. However, for most reflectance functions the light-field illumination cone is *not equal* to the set of all light-fields. One example is Lambertian reflectance. In this case, the light-field cone *never equals* the set of all light-fields because any two pixels in the light-field that image the same point on the object will always have the same intensity. For Lambertian objects the image illumination cone across arbitrary illumination conditions still *exactly equals* the set of all images because the pixels can still all be set independently by choosing the illumination appropriately.

For point light sources at infinity (rather than for arbitrary illumination conditions), the results are similar. The image illumination cone can sometimes be full-dimensional. For convex Lambertian objects the dimensionality equals the number of distinct surface normals. (See [5] Proposition 5.) If each surface normal is different, the image illumination cone is full-dimensional. For light-fields, however, the light-field illumination cone of a convex Lambertian object with point light sources at infinity is never full dimensional because any two pixels in the light-field that image the same point on the surface will always have the same intensity.

The trend in Table 1 is clear. Object recognition in the presence of illumination changes is "theoretically" easier using light-fields than with images. Using either model (arbitrary illumination or point light sources at infinity) the light-field illumination cone is a "smaller" subset of the set of all light-fields than the image illumination cone is a subset of the set of all images.

2.3 Distinguishability of Objects from Their Images and Light-Fields

As mentioned in [5] the convex cone property is potentially very important for object recognition because it implies that if the illumination cones of two objects are disjoint, they can be separated by a linear discriminant function. This property makes classification much easier because applying a linear classifier is in general far easier than determining which illumination cone an image or light-field lies closest to. However, to take advantage of this property, the two illumination cones must be disjoint. If they are not the two objects will not always be distinguishable anyway. These arguments, of course, apply equally to both image and light-field illumination cones. In this section we study the distinguishability (intersection) of illumination cones and show that the task is theoretically easier for light-fields than for images. We begin with image illumination cones.

2.3.1 Distinguishability of Objects from Their Images

An immediate corollary of the fact that the image illumination cones of convex objects under arbitrary lighting are exactly equal to the set of all images (see Table 1) is that no two convex objects (Lambertian or not) can *ever* be distinguished without some assumptions about the illumination:

Corollary 1 *The image illumination cones of any two convex objects seen under all possible lighting conditions are exactly equal. It is therefore never possible to say which convex object an image*

came from. It is not even possible to eliminate any convex objects as possibilities.

Perhaps one of the most important results of [5] is to show that, if the illumination consists of point sources at infinity the situation is more favorable; empirically the volume of the image illumination cone is much less than the space of all images. It is also possible to show that there are pairs of objects that are distinguishable under this smaller set of lighting conditions:

Theorem 3 *There exist pairs of objects for which the intersection of their illumination cones (over the set of illumination conditions consisting of arbitrary numbers of point light sources at infinity) only consists of the black (all zero) image; i.e. there are pairs of objects that are always distinguishable (over the set of illumination conditions which consist of point light sources at infinity.)*

Proof: (Sketch) One example is to consider two Lambertian spheres, one with an albedo function that has multiple step discontinuities (which appear in every image), one that varies smoothly everywhere. All of the images of the object with the step discontinuity in the albedo map will also have a step discontinuity in the image, whereas none of the images of the other object will. □

Although we have shown that there are pairs of objects for which the image illumination cones (for point light sources at infinity) only intersect at the all black image, there are pairs of objects for which their image illumination cones do intersect.

Theorem 4 *There exist pairs of objects for which the intersection of their illumination cones (over the set of point light sources at infinity) consists of more than just the black (all zero) image; i.e. there are pairs of objects that are sometime indistinguishable (over point light sources at infinity.)*

Proof: Consider two convex Lambertian objects in different illuminations. If each object has albedo variation proportional to the foreshortened incoming illumination of the other object, the two objects will generate the same image. (The constants of proportionality must be the same.) □

2.3.2 Distinguishability of Same-Shape Objects from Their Light-Fields

In the previous section we showed that distinguishing objects from their images under varying illumination is often very difficult, and in many cases "theoretically" impossible. If the objects are the same shape, convex, and Lambertian, intuitively the light-field should not contain any additional information. It is no surprise, then, that it is fairly straight-forward to prove an analogy of Corollary 1 for (convex Lambertian) objects of the same shape:

Theorem 5 *The light-field illumination cones over all possible lighting conditions of any two convex, Lambertian objects of the same shape are exactly equal.*

Proof: Given arbitrary lighting, it is possible to generate any incoming radiance distribution over the surface of the (convex) object using lasers. It is therefore possible to generate any light-field for any convex object (subject to the necessary and sufficient constraint that rays imaging the same point on the surface of the object have the same intensity.) □

Distinguishing (convex Lambertian) objects of the same shape from their light-fields is therefore impossible without any assumptions on the illumination. If assumptions are made about the illumination, the situation is different. As in Theorems 3 and 4 above, if the illumination consists of point light sources at infinity two objects of the same shape may or may not be distinguishable.

Theorem 6 *There exist pairs of same-shape convex, Lambertian objects for which the intersection of their light-field illumination cones (over the set of point light sources at infinity) only consists of the black (all zero) light-field; i.e. there are pairs of same-shape objects that are always distinguishable (over the set of point light sources at infinity.)*

Proof: Essentially the same as the proof of Theorem 3. □

Theorem 7 *There exist pairs of convex, Lambertian objects with the same shape for which the intersection of their light-field illumination cones (over the set of point light sources at infinity) consists of more than just the black (all zero) image; i.e. there are pairs of same-shape objects that are sometime indistinguishable even given their light-fields.*

Proof: Essentially the same as the proof of Theorem 4. □

2.3.3 Distinguishability of Differently-Shaped Objects from Their Light-Fields

Intuitively the situation for differently shaped objects is different. The light-field contains considerable information about the shape of the objects. In fact, we recently showed in [2] that, so long as the light-field does not contain any extended constant intensity regions, it uniquely defines the shape of a Lambertian object. This means that the intersection of the light-field cones of two differently shaped objects must only contain light-fields that have constant intensity regions.

Theorem 8 *The intersection of the light-field illumination cones over all possible lighting conditions of any two Lambertian objects that have different shapes only consists of light-fields that have constant intensity regions.*

Table 2: The distinguishability of objects from their images and light-fields. The main point to note is that if two objects have the same shape, the light-field adds nothing to the ease with which they can be distinguished, compared to just a single image. On the other hand, if the two objects have different shapes, it is theoretically far easier to distinguish them from their light-fields than it is from single images.

	Arbitrary Illumination Conditions	Point Light Sources at Infinity
Images of Two Convex Lambertian Objects	Never Distinguishable (Corollary 1)	Sometimes Distinguishable (Thm. 3) Sometime Indistinguishable (Thm. 4)
Light-fields of Two Same Shape Convex Lambertian Objects	Never Distinguishable (Theorem 5)	Sometimes Distinguishable (Thm. 6) Sometime Indistinguishable (Thm. 7)
Light-fields of Two Differently Shaped Lambertian Objects	Distinguishable if No Constant Intensity (Thm. 8)	Always Distinguishable if No Constant Intensity (Thm. 8)

This theorem implies that two differently shaped Lambertian objects can always be distinguished from any light-field that does not contain constant intensity regions.

2.3.4 Summary

We have described various conditions under which pairs of objects are distinguishable from their images or light-fields. See Table 2 for a summary. When nothing is assumed about the incoming illumination, it is impossible to distinguish between any pair of objects from their images. If the illumination consists of a collection of point light sources at infinity, the situation is a little better. Some pairs of objects can always be distinguished, but other pairs are sometimes indistinguishable.

If the objects have the same shape the situation is the same with light-fields. Light-field don't add to the discriminatory power of a single image. If the objects have different shapes the light-field adds a lot of discriminatory power. So long as the light-field has no constant intensity regions, any pair of differently shaped objects can be distinguished under any illumination conditions.

2.4 Implications

The implication of these theoretical results is as follows. The light-field provides considerable information about the shape of objects that can help distinguish between them in unknown, arbitrary illumination conditions under which they would be indistinguishable from single images.

Although it is practically impossible to capture the entire light-field for most object recognition tasks, sometimes it may be possible to capture 2-3 images. Ideally we would like an object recognition algorithm that can use any subset of the light-field; a single image, a pair of images, multiple images, or even the entire light-field. Such an algorithm should be able to take advantage of the implicit shape information in the light-field. In the remainder of this paper we describe exactly such an algorithm, the first step of which is to estimate the light-field from the input image(s).

3 Eigen Light-Fields for Face Recognition Across Pose

In many face recognition application scenarios the *pose* of the probe and gallery images are different. The gallery image might be a frontal "mug-shot" and the probe might be a 3/4 view captured from a surveillance camera in the corner of the room. The number of gallery and probe images may also vary. The gallery may consist of a pair of images of each subject, perhaps a frontal mug-shot and full profile view, like the images typically captured by police departments. The probe may be a similar pair of images, a single 3/4 view, or even a collection of views from random poses.

Until recently face recognition across pose (i.e. when the gallery and probe have different poses) has received very little attention in the literature with a few exceptions [6]. Algorithms have been proposed which can recognize faces [21] or more general objects [19] at a variety of poses. In a recent generalization of eigenfaces multi-linear algebra is used to compute a representation that separates the different modes underlying the formation of face images such as pose, illumination and expression [29]. In limited experiments an algorithm based on this representation performs better than eigenfaces. Most of these algorithms require gallery images at every pose, however. Algorithms have been proposed which do generalize across pose, for example [12], but this algorithm computes 3D head models using a gallery containing a large number of images per subject captured with controlled illumination variation. It cannot be used with arbitrary galleries and probes. Note, however, that concurrent with this paper there has been a growing interest in face recognition across pose. For example, Vetter *et al.* have developed an algorithm based on fitting a 3D morphable model [9, 24].

On a technical level (although not on an application level), the work most closely related to ours

is by Vetter and Poggio [30]. In their paper Vetter and Poggio introduced an algorithm for the re-rendering of faces across pose. Our algorithm (described in Section 3.2) is a strict generalization of this algorithm. Our algorithm is formulated in terms of light-fields, and so works with any collection of input images, or even input rays. When we reduce our algorithm to only two views it is (essentially) the same as the algorithm in [30]. Another major difference between this paper and [30] is that the algorithm in [30] is never actually used to perform face recognition.

In this section we propose an algorithm for face recognition across pose using light-fields. Our algorithm can use any number of gallery images captured at arbitrary poses, and any number of probe images also captured with arbitrary poses. A minimum of 1 gallery and 1 probe image are needed, but if more images are available the performance of our algorithm generally gets better.

Our algorithm operates by estimating (a representation of) the light-field of the subject's head. First, generic training data is used to compute an eigen-space of head light-fields, similar to the construction of eigen-faces [28]. Light-fields are simply used rather than images. Given a collection of gallery or probe images, the projection into the eigen-space is performed by setting up a least-squares problem and solving for the projection coefficients similarly to approaches used to deal with occlusions in the eigenspace approach [7, 16]. This simple linear algorithm can be applied to any number of images, captured from any poses. Finally, matching is performed by comparing the probe and gallery light-fields using a nearest neighbor algorithm.

The remainder of this section is organized as follows. We begin in Section 3.1 by introducing the concept of *eigen light-fields* before presenting the algorithm to estimate them from a collection of images in Section 3.2. After describing some of the properties of this algorithm in Section 3.3, we then describe how the algorithm can be used to perform face recognition across pose in Section 3.4. Finally, we present our experimental results in Section 3.5.

3.1 Eigen Light-Fields

Suppose we are given a collection of light-fields $L_i(\theta, \phi)$ where $i = 1, \ldots, N$. See Figure 1 for the definition of this notation. If we (vectorize the light-fields and then) perform an eigen-decomposition of the vectors using Principal Components Analysis (PCA), we obtain $d \leq N$ eigen light-fields $E_i(\theta, \phi)$ where $i = 1, \ldots, d$. Assuming the eigen-space of light-fields is a good

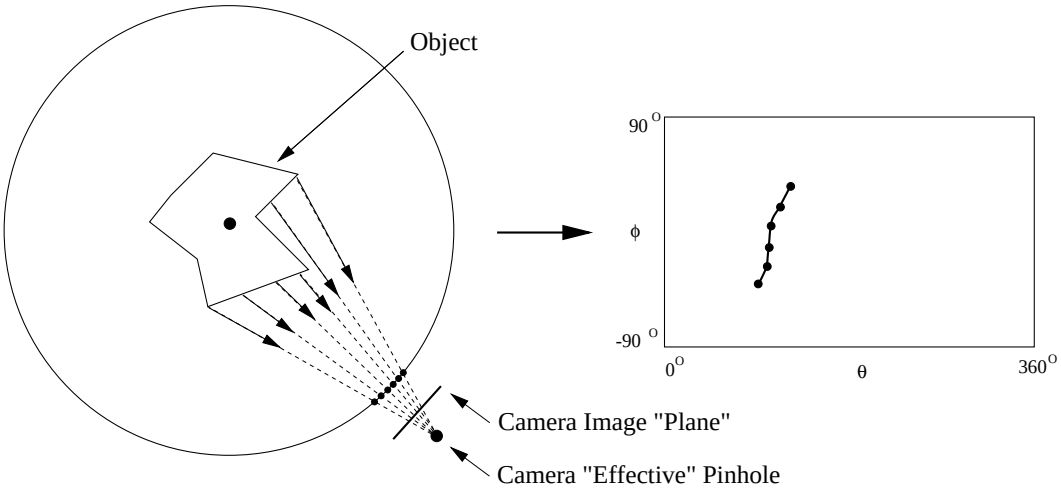


Figure 2: The 1D image of a 2D object corresponds to a curve (surface for a 2D image of a 3D object) in the light-field. Each pixel corresponds to a ray in space through the camera pinhole and the location of the pixel in the image. In general this ray intersects the light-field circle at a different point for each pixel. As the pixel considered "moves" in the image, the point on the light-field circle traces out a curve in θ-ϕ space. This curve is a straight vertical line iff the "effective pinhole" lies on the circle used to define the light-field.

representation of the set of light-fields under consideration, we can approximate any light-field:

$$L(\theta, \phi) \approx \sum_{i=1}^{d} a_i E_i(\theta, \phi) \tag{1}$$

where $a_i = \langle L(\theta, \phi), E_i(\theta, \phi) \rangle$ is the inner (or dot) product between $L(\theta, \phi)$ and $E_i(\theta, \phi)$. This decomposition is analogous to that used in face and object recognition [19, 28]; it is just performed on the entire light-field rather than on single images. (The mean light-field can be included as a constant additive term in Equation (1) and subtracted from the light-field in the definition of a_i if so preferred. There is very little difference in doing this however.)

3.2 Estimating Light-Fields from Images

Capturing the complete light-field of an object is a difficult task, primarily because it requires a huge number of images [13, 17]. In most object recognition scenarios it is unreasonable to expect more than a few images of the object; often just one. As shown in Figure 2, however, any image of the object corresponds to a curve (for 3D objects, a surface) in the light-field. One way to look at this curve is as a highly occluded light-field; only a very small part of the light-field is visible.

Can the coefficients a_i be estimated from this highly occluded view? Although this may seem hopeless, note that light-fields are highly redundant, especially for objects with simple reflectance properties such as Lambertian. An algorithm is presented in [16] to solve for the unknown a_i for eigen-*images*. A similar algorithm was used in [7]. Rather than using the inner product $a_i = \langle L(\theta, \phi), E_i(\theta, \phi) \rangle$, Leonardis and Bischof [16] solve for a_i as the least squares solution of:

$$L(\theta, \phi) - \sum_{i=1}^{d} a_i E_i(\theta, \phi) \; = \; 0 \tag{2}$$

where there is one such equation for each pair of θ and ϕ that are un-occluded in $L(\theta, \phi)$. Assuming that $L(\theta, \phi)$ lies *completely within the eigen-space* and that enough pixels are un-occluded, then it is well-known that the solution of Equation (2) will be exactly the same as that obtained using the inner product:

Theorem 9 *Assuming that $L(\theta, \phi)$ is in the linear span of $\{E_i(\theta, \phi) \mid i = 1, \ldots d\}$, then $a_i = \langle L(\theta, \phi), E_i(\theta, \phi) \rangle$ is always an exact minimum solution of Equation (2).*

Since there are d unknowns ($a_1 \ldots a_d$) in Equation (2), at least d un-occluded light-field pixels are needed to over-constrain the problem, but more may be required due to linear dependencies between the equations. In practice, $2 - 3$ times as many equations as unknowns are typically required to get a reasonable solution [16]. Given an image $I(m, n)$, the following is then an algorithm for estimating the eigen light-field coefficients a_i:

Eigen Light-Field Estimation Algorithm

1. For each pixel (m, n) in $I(m, n)$ compute the corresponding light-field angles $\theta_{m,n}$ and $\phi_{m,n}$. (This step assumes that the camera intrinsics are known, as well as the relative orientation between the camera and object. In Section 3.4.1 we will describe how to avoid this step and instead use a simple "normalization" to convert the input images into light-field vectors.)
2. Find the least-squares solution (for $a_1 \ldots a_d$) to the set of equations:

$$I(m, n) - \sum_{i=1}^{d} a_i E_i(\theta_{m,n}, \phi_{m,n}) \; = \; 0 \tag{3}$$

where m and n range over their allowed values. (In general, the eigen light-fields E_i need to be interpolated to estimate $E_i(\theta_{m,n}, \phi_{m,n})$. Also, all of the equations for which the pixel $I(m,n)$ does not image the object should be excluded from the computation.)

Although we have described this algorithm for a single image $I(m,n)$, any number of images can obviously be used. The extra pixels from the other images are simply added in as additional constraints on the unknown coefficients a_i in Equation (3).

3.3 Properties of the Eigen Light-Field Estimation Algorithm

The Eigen Light-Field Estimation Algorithm can be used to estimate a light-field from a collection of images. Once the light-field has been estimated, it can then, theoretically at least, be used to render new images of the same object under different poses. See [30] for a related algorithm. In this section we show that if the objects used to create the eigen-space of light-fields all have the same shape as the object imaged to create the input to the algorithm, then this re-rendering process is in some sense "correct," assuming that all the objects are Lambertian. As a first step, we show that the eigen light-fields $E_i(\theta, \phi)$ capture the shape of the objects in the following sense:

Theorem 10 *If $\{L_i(\theta,\phi) \mid i = 1, \ldots, N\}$ is a collection of light-fields of Lambertian objects with the same shape, then all of the eigen light-fields $E_i(\theta,\phi)$ have the property that if (θ_1, ϕ_1) and (θ_2, ϕ_2) define two rays which image the same point on the surface of any of the objects then:*

$$E_i(\theta_1, \phi_1) \;=\; E_i(\theta_2, \phi_2) \qquad \forall i = 1 \ldots d. \tag{4}$$

Proof: The property in Equation (4) holds for all of the light-fields $\{L_i(\theta,\phi) | i = 1, \ldots, N\}$ used in the PCA because they are Lambertian. Hence, it also holds for any linear combination of the L_i. Therefore it holds for the eigen-vectors because they are linear combinations of the L_i. □

The property in Equation (4) also holds for all linear combinations of the eigen light-fields. It therefore holds for the light-field recovered in Equation (3) in the Light-Field Estimation Algorithm, assuming that the light-field from which the input image is derived lies in the eigen-space so that Theorem 9 applies. This means that the Light-Field Estimation Algorithm estimates the light-field in a way that is consistent with the object being Lambertian and of the appropriate shape:

Corollary 2 *Suppose $\{E_i(\theta,\phi) \mid i = 1, \ldots, d\}$ are the eigen light-fields of a set of Lambertian objects with the same shape and $I(m,n)$ is an image of another Lambertian object with the same*

shape. If the light-field from which $I(m,n)$ is derived lies in the light-field eigen-space, then the light-field recovered by the Light-Field Estimation Algorithm has the property that if $\theta_{m,n}, \phi_{m,n}$ is any *pair of angles which image the same point in the scene as the pixel (m,n) then:*

$$I(m,n) \;=\; E(\theta_{m,n}, \phi_{m,n}). \tag{5}$$

where $E(\theta_{m,n}, \phi_{m,n})$ is the light-field estimated by the Light-Field Estimation Algorithm; i.e. the algorithm correctly re-renders the object under the Lambertian reflectance model.

Theorem 2 implies that the algorithm is acting reasonably in estimating the light-field, a task which is impossible from a single image without a prior model on the shape of the object. Unlike in [30], here the shape model is implicitly contained in the eigen light-fields. Theorem 2 assumes that all of the objects are approximately the same shape, but that is a common assumption for faces [23]. Even if there is some shape variation in faces, it is reasonable to assume that the eigen light-fields will capture this information. Theorem 2 also assumes that faces are Lambertian and that the light-field eigenspace accurately approximates any face light-field. The extent to which these assumptions are valid are illustrated in Figure 3 where we present results of using our algorithm to re-render faces across pose. In each case the algorithm received the left-most (frontal) image as input and created the rotated view in the middle. For comparison, the original rotated view is included as the right-most image. The re-rendered image for the first subject is very similar to the original. While the image created for the second subject still shows a face in the correct pose, the identity of the subject is not as accurately recreated. We conclude that overall our algorithm works fairly well, but that more training data is needed so that the eigen light-field of faces can more accurately represent any given face light-field.

3.4 Application to Face Recognition Across Pose

The Eigen Light-Field Estimation Algorithm described above is somewhat abstract. In order to be able to use it for face recognition across pose we need to do the following things:

Vectorization: The input to a face recognition algorithm consists of a collection of images (possibly just one) captured from a variety of poses. The Eigen Light-Field Estimation Algorithm operates on light-field vectors (light-fields represented as vectors). Vectorization consists of converting the input images into a light-field vector (with missing elements, as appropriate.)

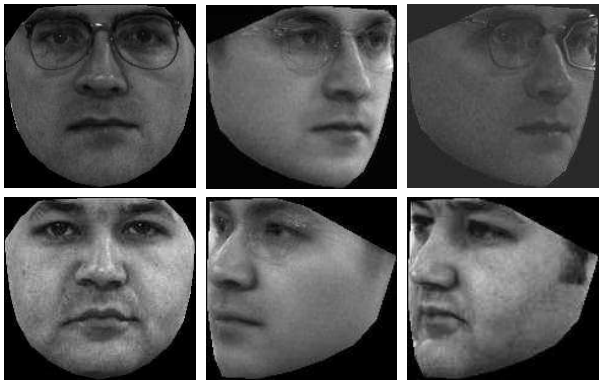


Figure 3: An illustration of using our eigen light-field estimation algorithm for re-rendering a face across pose. The algorithm is given the left-most (frontal) image as input from which it estimates the eigen light-field and then creates the rotated view shown in the middle. For comparison, the original rotated view is shown in the right-most column. In the figure we show one of the better results (top) and one of the worst (bottom.) Although in both cases the output looks like a face, the identity is corrupted in the second case.

Classification: Given the eigen coefficients $a_1 \ldots a_d$ for a collection of gallery (training) faces and for a probe (test) face, we need to classify which gallery face is the most likely match.

Selecting Training and Testing Sets: To evaluate our algorithm we have to divide the database(s) used into (disjoint) subsets for training and testing.

We now describe each of these tasks in turn.

3.4.1 Vectorization by Normalization

Vectorization is the process of converting a collection of images of a face into a light-field vector. Before we can do this we first have to decide how to discretize the light-field into pixels. Perhaps the most natural way to do this is to uniformly sample the light-field angles, θ and ϕ in the 2D case of Figure 2. This is not the only way to discretize the light-field. Any sampling, uniform or non-uniform, could be used. All that is needed is a way of specifying what is the allowed set of light-field pixels. For each such pixel, there is a corresponding index in the light-field vector; i.e. if the light-field is sampled at K pixels, the light-field vectors are K dimensional vectors.

We specify the set of light-field pixels in the following manner. We assume that there are only a finite set of poses $1, 2, \ldots P$ in which the face can occur. Each face image is first classified into the nearest pose. (Although this assumption is clearly an approximation, its validity is demonstrated

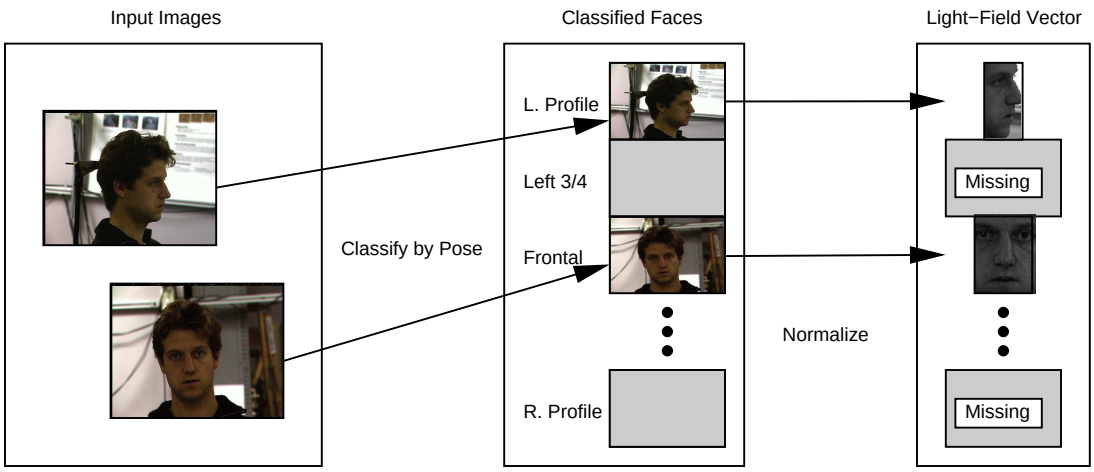


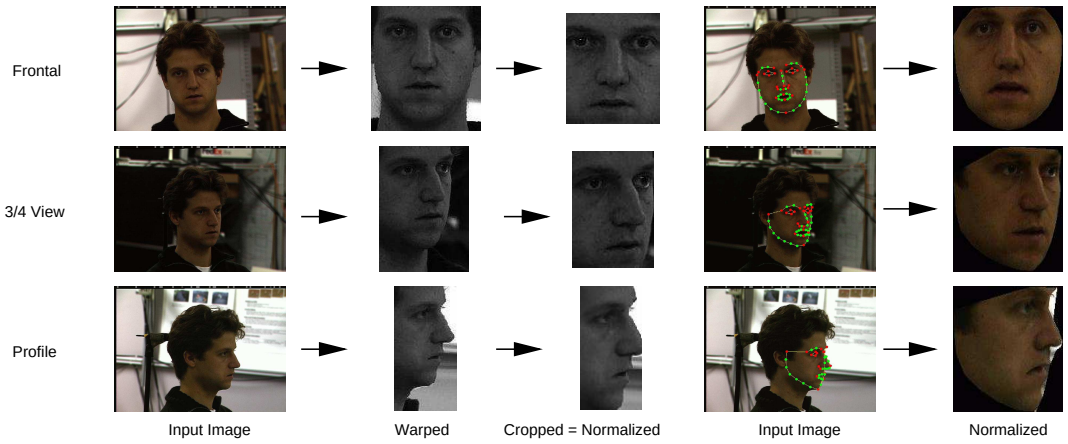
Figure 4: Vectorization by normalization. Vectorization is the process of converting a set of images of a face into a light-field vector. Vectorization is performed by first classifying each input image into one of a finite number of poses. For each pose, a normalization is then applied to convert the image into a sub-vector of the light-field vector. If poses are missing, the corresponding part of the light-field vector is missing.

by the empirical results in Section 3.5.3. In both the FERET [22] and PIE [26] databases, there is considerable variation in the pose of the faces. Although the subjects are asked to place their face in a fixed pose, they rarely do this perfectly. Both databases therefore contain considerable variation away from the finite set of poses. Since our algorithm performs well on both databases, the approximation of classifying faces into a finite set of poses is validated.)

Each pose $i = 1, \ldots, P$ is then allocated a fixed number of pixels K_i. The total number of pixels in a light-field vector is therefore $K = \sum_{i=1}^{P} K_i$. If we have images from pose 3 and 7, for example, we know $K_3 + K_7$ of the K pixels in the light-field vector. The remaining $K - K_3 - K_7$ are unknown, missing data. This vectorization process is illustrated in Figure 4.

We still need to specify how to sample the K_i pixels of a face in pose i. This process is analogous to that needed in appearance-based object recognition and usually performed by "normalization." In eigenfaces [28], the standard approach is to find the positions of several canonical points, typically the eyes and the nose, and to warp the input image onto a coordinate frame where these points are in fixed locations. The resulting image is then masked. To generalize eigenface normalization to eigen light-fields, we just need to define such a normalization for each pose.

In this paper we experimented with two different normalizations. The first one, illustrated in Figure 5(a) for three poses, is a simple one based on the location of the eyes and the nose. Just as in



(a) 3-Point Normalization (b) Multi-Point Normalization

Figure 5: (a) The first, simpler normalization for three poses in the set in Figure 4, one frontal, one a 3/4 view, the final a full profile. Just as in eigenfaces, we assume that the eye and nose locations are known, warp the face into a coordinate frame in which these canonical points are in a fixed location and finally crop the image with a (pose dependent) mask. (b) The second, more complex normalization. In this case, a large number (39–54 depending on the pose) of points on the face are used to perform the normalization.

eigenfaces, we assume that the eye and nose locations are known, warp the face into a coordinate frame in which these canonical points are in a fixed location and finally crop the image with a (pose dependent) mask to yield the K_i pixels. For this simple 3-point normalization, the resulting masked images vary in size between 7200 and 12600 pixels.

The second normalization is more complex and is motivated by the success of Active Appearance Models [10]. This normalization is based on the location of a large number (39–54 depending on the pose) of points on the face. These canonical points are triangulated and the image warped with a piecewise affine warp onto a coordinate frame in which the canonical points are in fixed locations. See Figure 5(b) for an illustration of this multi-point normalization. The resulting masked images for this multi-point normalization vary in size between 20800 and 36000 pixels. Although currently the multi-point normalization is performed using hand-marked points, it could be performed by fitting an Active Appearance Model [10] and then using the implied canonical point locations. Further discussion of this way of automating our algorithm is contained in Section 4.2.

3.4.2 Classification using Nearest Neighbor

The Eigen Light-Field Estimation Algorithm outputs a vector of eigen coefficients $(a_1, \ldots, a_d)$. Given a set of gallery (training) faces, we obtain a corresponding set of vectors $(a_1^{\mathrm{id}}, \ldots, a_d^{\mathrm{id}})$, where id is an index over the set of gallery faces. Similarly, given a probe (or test) face, we obtain a vector $(a_1, \ldots, a_d)$ of eigen coefficients for that face. To complete the face recognition algorithm we need an algorithm which classifies $(a_1, \ldots, a_d)$ with the index id which is the most likely match. Many different classification algorithms could be used for this task. For simplicity, we use the nearest neighbor algorithm which classifies the vector $(a_1, \ldots, a_d)$ with the index:

$$\arg\min_{\mathrm{id}} \mathrm{dist}\left((a_1, \ldots, a_d), (a_1^{\mathrm{id}}, \ldots, a_d^{\mathrm{id}})\right) \;=\; \arg\min_{\mathrm{id}} \sum_{i=1}^{d} \left(a_i - a_i^{\mathrm{id}}\right)^2. \tag{6}$$

All of the results reported in this paper use the Euclidean distance in Equation (6). Alternative distance functions, such as the Mahalanobis distance, could be used instead if so desired.

3.4.3 Selecting the Gallery, Probe, and Generic Training Data

In each of our experiments we divided the database(s) into three disjoint subsets:

Generic Training Data: Many face recognition algorithms such as eigenfaces, and including our algorithm, require "generic training data" to build a generic face model. In eigenfaces, for example, generic training data is needed to compute the eigenspace. Similarly, in our algorithm generic data is needed to construct the eigen light-field.

Gallery: The gallery is the set of "training" images of the people to be recognized; i.e. the images given to the algorithm as examples of each person that might need to be recognized.

Probe: The probe set contains the "test" images; i.e. the examples of images to be presented to the system that should be classified with the identity of the person in the image.

The division into these three subsets is performed as follows. First we randomly select half of the subjects as the generic training data. The images of the remaining subjects are used for the gallery and probe. There is therefore never any overlap between the generic training data and the gallery

Figure 6: An illustration of the pose variation in the CMU PIE database [26]. The pose varies from full right profile (c22) to full frontal (c27) and on to full left profile (c34). The 9 cameras in the horizontal sweep are each separated by about 22.5°. The 4 other cameras include 1 above (c09) and 1 below (c07) the central camera, and 2 in the corners of the room (c25 and c31), typical locations for surveillance cameras.

and probe. (Note that it is often a good idea for the gallery and generic training data to be the same. There is nothing in our algorithm that precludes this. We just require the gallery and generic training data to be disjoint for fairness and to avoid any bias in our results. In order to include gallery subjects during training, an algorithm performing PCA with missing data would have to be used to compute the eigen light-fields [11, 25])

After the generic training data has been removed, the remainder of the databases(s) are divided into probe and gallery sets based on the pose of the images. For example, we might set the gallery to be the frontal images and the probe set to be the left profiles. In this case, we evaluate how well our algorithm is able to recognize people from their profiles given that the algorithm has only seen them from the front. In the experiments described below we choose the gallery and probe poses in various different ways. The gallery and probe are always completely disjoint however.

3.5 Experimental Results

3.5.1 Databases

We used two databases in our face recognition across pose experiments, the CMU Pose, Illumination, and Expression (PIE) database [26] and the FERET database [22]. Each of these databases contains substantial pose variation. In the pose subset of the CMU PIE database (see Figure 6), the 68 subjects are imaged simultaneously under 13 different poses totaling 884 images. In the FERET database, the subjects are imaged non-simultaneously in 9 different poses. See Figure 7

for an example. We used 200 subjects from the FERET pose subset giving 1800 images in total. (In both cases, we used greyscale images even if the database actually contains color images.). If not stated otherwise we used half of the available subjects for training of the generic eigenspace (34 subjects for PIE, 100 subjects for FERET) and the remaining subjects for testing. In all experiments (if not stated otherwise) we retain a number of eigenvectors sufficient to explain 95% of the variance in the input data.

bb bc bd be ba bf bg bh bi

Figure 7: An illustration of the pose variation in the FERET database [22]. The poses of the 9 images vary from $+60^\circ$ (bb) to full frontal (ba) and on to -60° (bi). Overall, the variation in pose is somewhat less than in the CMU PIE database. See Figure 6 for an illustration of the pose variation in the PIE database.

3.5.2 Example Eigen Light-Field

Figure 8 illustrates an example eigen light-field for the PIE database. To best illustrate the appearance variation, we actually display the mean vector plus a multiple of each eigenvector. That way the eigenvectors actually look like real faces that are typical of the appearance variation modeled by the eigenspace. As can be seen, the eigenvectors mostly encode identity information.

3.5.3 Experiment 1: Comparison with Other Algorithms

We first conducted an experiment to compare our algorithm with two others. In particular we compared our algorithm with eigenfaces [28] and FaceIt, the commercial face recognition system from Identix (formerly Visionics). Eigenfaces is the defacto baseline standard by which face recognition algorithms are compared. FaceIt finished top overall in the Face Recognition Vendor Test 2000 [8].

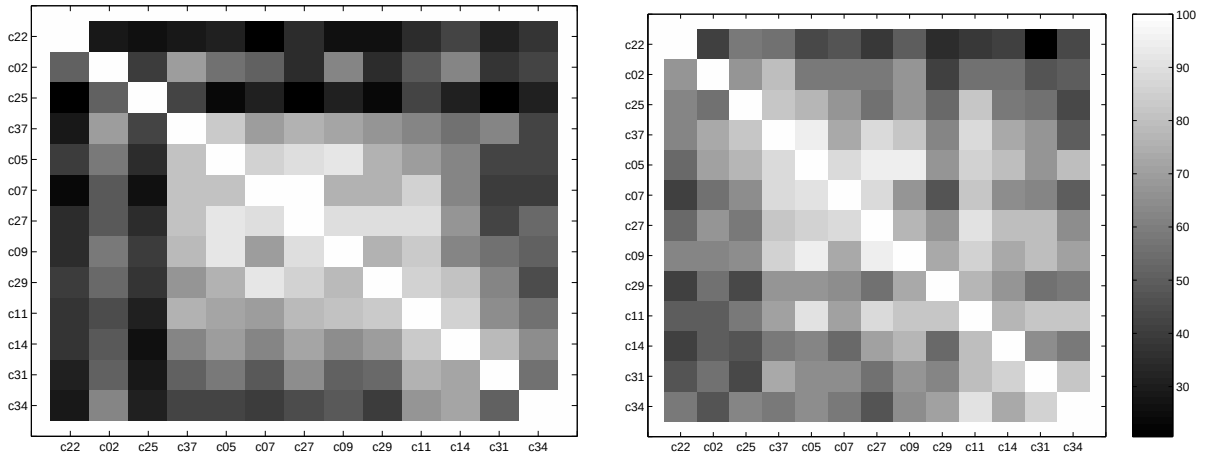
We first performed a comparison using the PIE database [26]. After randomly selecting the generic training data, we selected the gallery pose as one of the 13 PIE poses and the probe pose as any other of the remaining 12 PIE poses. For each disjoint pair of gallery and probe poses, we

Figure 8: The mean vector and the first three eigen light-field vectors for 6 poses of the PIE database. To best illustrate the eigenvectors, we display the mean vector plus a multiple of each eigenvector. That way the eigenvectors actually look like real faces, and are typical of the appearance variation in the eigenspace.

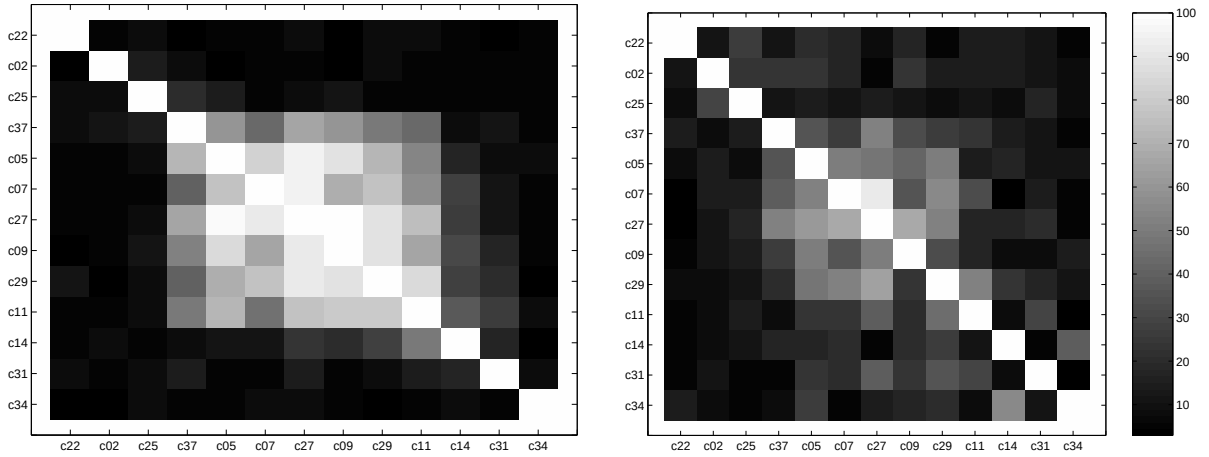
compute the average recognition rate over all subjects in the probe and gallery sets. The details of the results are included in Figures 9–10 and a summary is included in Table 3.

In Figure 9 we plot color-coded 13×13 "confusion matrices" of the results. The row denotes the pose of the gallery, the column the pose of the probe, and the displayed intensity the average recognition rate. A lighter color denotes a higher recognition rate. (On the diagonals the gallery and probe images are the same and so all three algorithms obtain a 100% recognition rate.) Eigen light-fields performs far better than the other algorithms, as is witnessed by the lighter color of Figures 9(a–b) compared to Figures 9(c–d). Note how eigen light-fields is far better able to generalize across wide variations in pose, and in particular to and from near profile views.

Several "cross-sections" through the confusion matrices in Figure 9 are shown in Figure 10. In each cross-section, we fix the pose of the gallery images and vary the pose of the probe image. In each graph we plot four curves, one for eigenfaces, one for FaceIt, one for eigen light-fields with the 3-point normalization, and one for eigen light-fields with the multi-point normalization. As can be seen, eigen light-fields outperforms the other two algorithms. In particular, it is better able to recognize the face when the gallery and probe poses are very different. This is witnessed by the eigen light-field curves in Figure 10 being higher at the extremities of the probe pose range.



(a) Eigen Light-Fields - 3-Point Normalization (b) Eigen Light-Fields - Multi-point Normalization

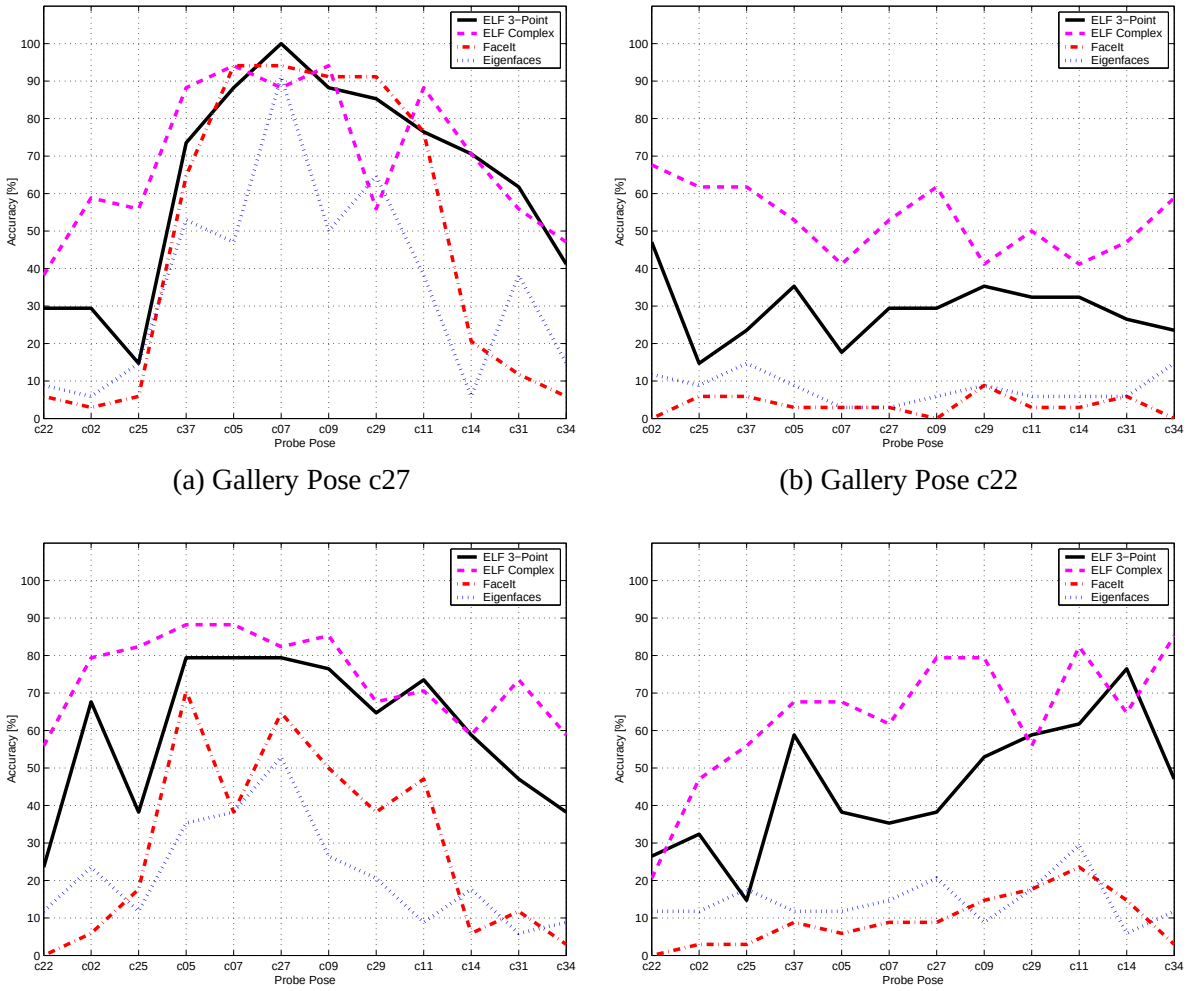


(c) FaceIt (d) Eigenfaces

Figure 9: A comparison with FaceIt and eigenfaces for face recognition across pose on the PIE database. For each pair of gallery and probe poses, we plot the color-coded average recognition rate. The fact that the images in (a) and (b) are lighter in color than those in (c) and (d) implies that our algorithm performs better.

The results in Figures 9 and 10 are summarized in Table 3. In this table we include the average recognition rate computed over all disjoint gallery-probe poses. As can be seen, eigen light-fields outperforms both the standard eigenfaces algorithm and the commercial FaceIt system. (In [24] better recognition results across pose on the PIE database are reported. See Section 4.3 for a detailed comparison of our algorithm with the model-based one in [24].)

We next performed a similar comparison using the FERET database [22]. Just as with the PIE database, we selected the gallery pose as one of the 9 FERET poses and the probe pose as any other of the remaining 8 FERET poses. For each disjoint pair of gallery and probe poses, we compute the average recognition rate over all subjects in the probe and gallery sets, and then average the



(a) Gallery Pose c27

(b) Gallery Pose c22

(c) Gallery Pose c37

(d) Gallery Pose c31

Figure 10: Several "cross-sections" through the confusion matrices in Figure 9. In each figure we fix the pose of the gallery and only vary the pose of the probe. We plot four curves, one each for eigen light-fields with the 3-point normalization, eigen light-fields with the multi-point normalization, eigenfaces, and FaceIt. The performance of eigen light-fields is superior to that for the other two algorithms, particularly when the pose of the gallery and probe are radically different. Eigen light-fields recognizes faces better across pose.

results. The results are very similar to those for the PIE database and are summarized in Table 4. Again, eigen light-fields performs significantly better than both FaceIt and eigenfaces.

Overall, the performance improvement of eigen light-fields over the other two algorithms is more significant on the PIE database than on the FERET database. This is because the PIE database contains more variation in pose than the FERET database. See Figures 6 and 7.

Table 3: A comparison of eigen light-fields with FaceIt and eigenfaces for face recognition across pose on the PIE database. The table contains the average recognition rate computed across all disjoint pairs of gallery and probe poses; i.e. this table summarizes the average performance in Figure 9.

	Eigenfaces	FaceIt	Eigen Light-Fields 3-Point Normalization	Eigen Light-Fields Multi-Point Normalization
Average Recognition Accuracy	16.6%	24.3%	52.5%	66.3%

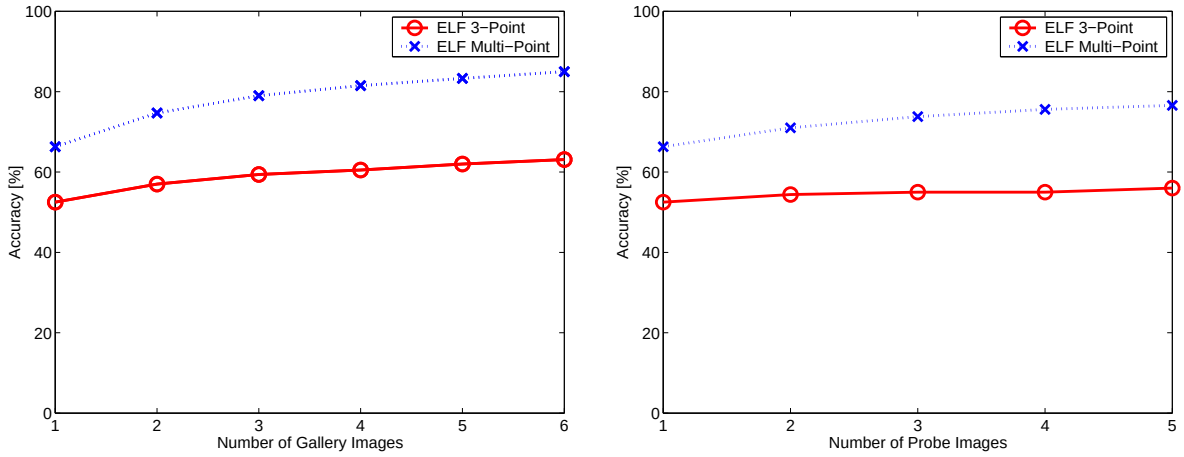
Table 4: A comparison of eigen light-fields with FaceIt and eigenfaces for face recognition across pose on the FERET database. The table contains the average recognition rate computed across all disjoint pairs of gallery and probe poses. Again, eigen light-fields outperforms both eigenfaces and FaceIt.

	Eigenfaces	FaceIt	Eigen Light-Fields 3-Point Normalization
Average Recognition Accuracy	39.4%	59.3%	75%

3.5.4 Experiment 2: Improvement with the Number of Input Images

So far we have assumed that just a single gallery and probe image are available to the algorithm. What happens if more gallery and/or probe images are available? In Experiment 2 we investigate the performance of eigen light-fields with different numbers of images using the PIE database. To compute the recognition rate with n gallery images, we select every possible set of n gallery poses and 1 probe pose. In total this amounts to $13 \times 12 \times \ldots (13-n)/n!$ different combinations of poses. We then compute the average recognition rate for each such combination and average the results. We plot the overall average recognition rate against the number of gallery images in Figure 11(a). As can be seen, eigen light-fields is able to estimate a more accurate light-field using more gallery images and thereby obtain a higher recognition rate.

Eigen light-fields can also take advantage of more than one probe image. We therefore repeated Experiment 2 but reversed the roles of the gallery and probe. The results are shown in Figure 11(b). Again the performance increases with the number of probe images, however the benefit of using multiple probe images is not as much as the benefit of using multiple gallery images. With multiple gallery images the accuracy of the light-field of every subject in the gallery is improved. With more probe images, the accuracy of the light-field of just the single probe subject is improved.



(a) Varying the Number of Gallery Images (b) Varying the Number of Probe Images

Figure 11: (a) The improvement in the performance of our algorithm with increasing numbers of gallery images. Using the additional images, eigen light-fi elds is able to estimate the light-fi elds more accurately and thereby obtains a higher recognition rate. (b) The performance of eigen light-fi elds also improves with the number of probe images. The performance increase is greater with increased numbers of gallery images because the accuracy of the light-fi eld of every gallery subject is improved. On the other hand, with more probe images, the accuracy of just the one probe subject is improved.

3.5.5 Experiment 3: Matching Sub-Images

We just illustrated how the performance of eigen light-fields improves if more gallery and/or probe images are available. Eigen light-fields can use any subset of the light-field. In particular, it does not even need a complete image. To validate this property, we ran the following experiment. We repeated Experiment 1, but for each pair of gallery and probe poses, we randomly selected a certain percentage of the pixels in the masked image. We then compute the average recognition rate just using this subset of the pixels. This process is repeated for 100 random samples of pixels and the results averaged. The results are shown in Figure 12 for a variety of pixel percentages ranging from 1% to 100% (the complete image). These results were obtained using the 3-point normalization and so the performance with 100% is 52.5%, as per Table 3. The figure demonstrates graceful degradation of the recognition performance when subsets of the images are used. The algorithm achieves remarkable recognition rates with just 1% of the image information. In this case the light-field contains an average of 206 pixels (3-point normalization) and 548 pixels (multi-point normalization), respectively.

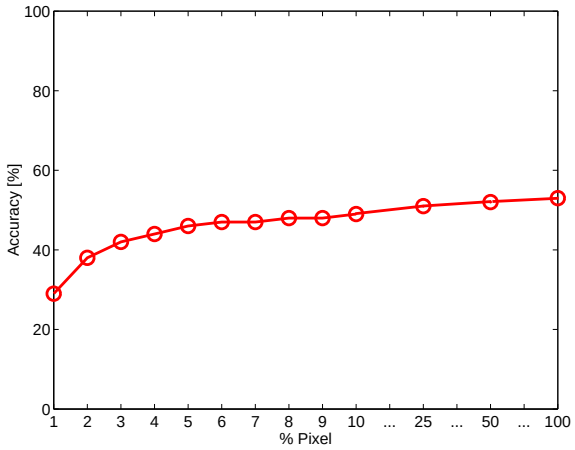


Figure 12: The performance of eigen light-fi elds with a subset of the images using the 3-point normalization and the PIE database. The average recognition rate is plotted against the percentage of pixels in the probe and gallery images. A subset of the images can be used without any signifi cant reduction in the recognition rate.

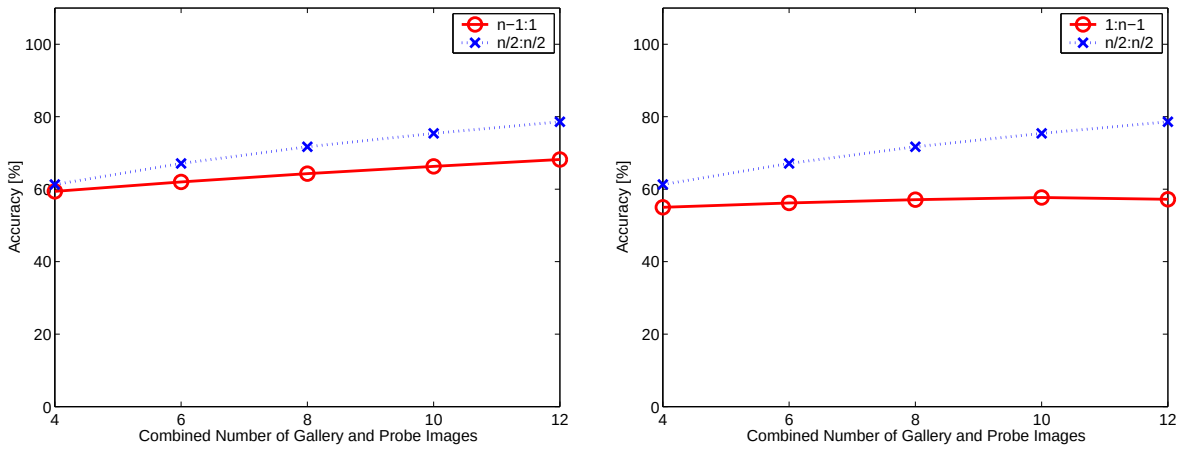
3.5.6 Experiment 4: Division of the Input Images between Gallery and Probe

In Experiment 2 we examined the benefits of using more than one gallery or probe image. Suppose that n gallery and probe images are available in total. Is it better to use $n-1$ gallery and 1 probe images or $n/2$ gallery and $n/2$ probe images? In order to answer this question, we conducted Experiment 4. Given n images, we generated every possible combination of $n-1$ gallery images and 1 probe image (as in Experiment 2) and every possible combination of $n/2$ gallery images and $n/2$ probe images. We then computed the average recognition rate for each case. Similarly we switched the roles of gallery and probe. The results are shown in Figure 13. The conclusion is clear. It is better to divide the images equally between gallery and probe rather than asymmetrically.

One possible conclusion from this result is that adding more that one image to each of the probe and gallery allows a better estimate of the light-field. Having two more accurate estimates results in better performance than having one very accurate estimate and one not so accurate estimate.

3.5.7 Experiment 5: Influence of Eigenspace Parameters

The computation of eigen light-fields is influenced primarily by the number of subjects used during training and the number of eigenvectors retained from Principal Components Analysis. In order to quantify the effect that these parameters have on recognition performance we repeated Exper-



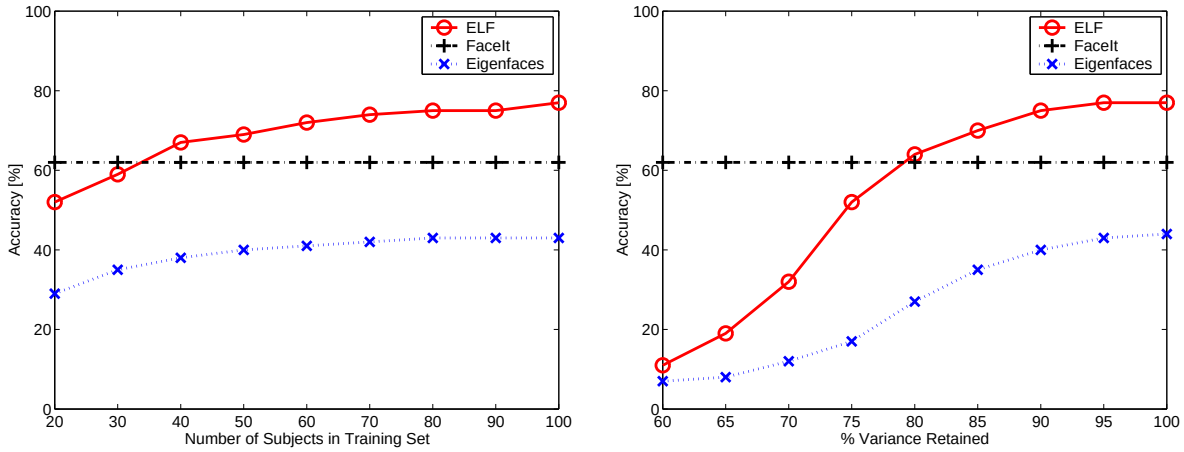
(a) $n-1$ gallery, 1 probe vs. $n/2$ of each (b) 1 gallery, $n-1$ probe vs. $n/2$ of each

Figure 13: (a) The performance of using $n-1$ gallery images and 1 probe image versus using $n/2$ of each. The empirical evidence suggests to split up the images evenly into gallery and probe. (b) The performance of using 1 gallery image and $n-1$ probe images versus using $n/2$ of each. Again splitting up the images evenly achieves higher recognition rates. Having two more accurate estimates of the light-fi elds results in better performance than having one very accurate estimate and one not so accurate estimate.

iment 1 on the FERET database and systematically changed their values. The recognition accuracies for FaceIt stay constant across the different parameter settings since we did not (and could not) retrain the system. Figure 14(a) shows recognition accuracies of eigen light-fields, eigenfaces and FaceIt for varying numbers of training subjects in the training set. For each experiment the same set of 100 gallery/probe subjects was used. It can be seen that eigen light-fields outperform FaceIt when more than 40 subjects are used in the training set. In a similar fashion Figure 14(b) shows recognition accuracies for the three algorithms for changing percentages of variance retained from Principal Component Analysis using 100 subjects in the training set. Here, eigen light-fields surpass the FaceIt performance when more than 80% of the variance (corresponding to eight eigenvectors) is retained. Again, the performance of eigen light-fields degrades gracefully.

3.5.8 Experiment 6: Recognition across databases

For all experiments shown so far the training and gallery/probe subjects were taken from the same database. In Figure 15 we show recognition accuracies for round-robin tests with seven poses of the PIE database. Training for both eigen light-fields and eigenfaces was done using all 200 subjects of the FERET database. Correspondence between FERET and PIE poses was determined



(a) (b)

Figure 14: Comparative performance of the three algorithms on the FERET database: (a) varying number of subjects in the generic training data and (b) varying the percentage of variance retained in the eigenspace (using 100 subjects). The recognition accuracies are averages of round-robin experiments using each pose in turn as gallery pose and all other poses as probe poses. The performance of the eigen light-fi elds surpasses FaceIt performance at comparatively low values of the parameter settings. The recognition accuracies for FaceIt stay constant across the different parameter settings since we did not (and could not) retrain the system.

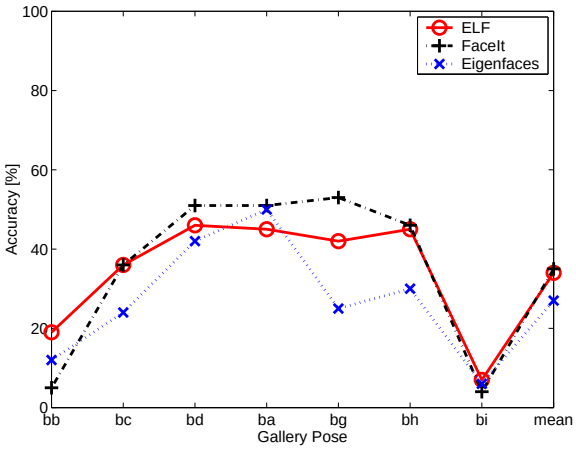


Figure 15: Recognition accuracies of eigen light-fi elds, eigenfaces and FaceIt for round-robin tests over seven PIE poses. Eigen light-fi elds and eigenfaces were trained using only FERET subjects and tested using only PIE subjects. The poses vary from $+60^\circ$ (bb) to full frontal (ba) and on to -60° (bi). See Figure 7 for example images.

manually. Note that there is a considerable mismatch between the FERET and PIE poses (see Figures 6 and 7). While low overall, the performance of eigen light-fields compares reasonably with the FaceIt performance on the same task.

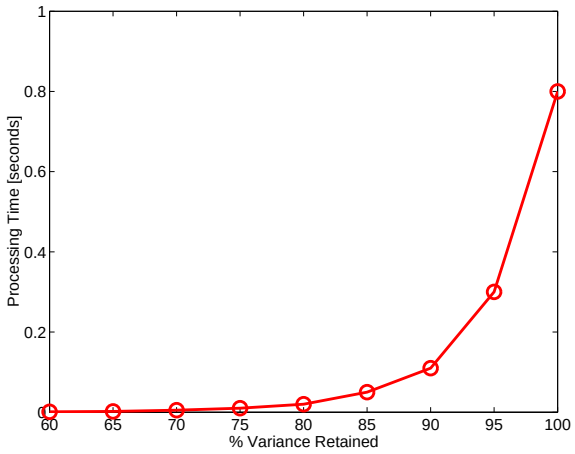


Figure 16: The processing time (on a 2.8 GHz Pentium 4) to solve the linear system to compute the eigenspace coeffi cients for a singe probe image (using 100 gallery subjects of the FERET database.) The results are averages for round-robin tests where each pose is used in turn as gallery pose with all other poses as probe poses. Typically 95% of the variance should be retained for maximum performance.

3.5.9 Experiment 7: Computational Complexity

Since the computation of the eigen light-fields and the eigenspace coefficient for the gallery images can be done offline, the online complexity is dominated by the cost of solving the overdetermined linear systems for the probe coefficients. Figure 16 shows the processing times for this step per probe subject on a 2.8 GHz Pentium 4 processor for different amounts of variance retained from Principal Component Analysis, corresponding to between 1 (for 60% variance) and 99 eigenvectors (for 100% variance). The experiment was conducted using the FERET database with 100 training subjects. The results are averages for round-robin tests where each pose is used in turn as the gallery pose with all other poses as probe poses. As shown in Figure 14(b) typically 95% of the variance should be retained for maximum performance.

4 Conclusion

4.1 Summary

Appearance-based object recognition uses pixels or measurements of light in the scene as its features. In the ultimate limit, the set of all such measurements is the plenoptic function or light-field. In this paper we have explored appearance-based object recognition from light-fields. We first analyzed the theoretical distinguishability of objects from their images and light-fields. We pre-

sented a number of results which show that theoretically objects can be distinguishable from their light-fields in cases that they are ambiguous from just a single image. This theoretical analysis motivates trying to build appearance-based object recognition algorithms that use as much of the light-field as is available, be it a single image, a pair of images, or multiple images.

In the second half of this paper we proposed an appearance-based algorithm for face recognition across pose based on an algorithm to estimate the eigen light-field from a collection of images. This algorithm can use any number of gallery images captured form arbitrary poses and any number of probe images also captured from arbitrary poses. The gallery and probe poses do not need to overlap. We showed that our algorithm can reliably recognize faces across pose and also take advantage of the additional information contained in widely separated views to improve recognition performance if more than one gallery or probe image is available. Note that the Eigen Light-Fields Algorithm can be extended to recognize faces across pose and illumination simultaneously by generalizing eigen light-fields [14] to Fisher light-fields [15], analogously to how eigen faces [28] can be generalized to Fisherfaces [3]. (Note, however, that care must be taken because Equation (3) assumes that the basis vectors (E_i) are able to reconstruct the image $I(m, n)$ well. This is not the case for Fisher light-fields and so a two-step algorithm is required that first reconstructs the eigen light-field and then constructs a Fisher light-field from that.)

4.2 Light-Fields, Vectorization, and Appearance-Based Face Recognition

One common reaction to the algorithm described in this paper is to question what it has to do with Light-Fields. On one level this is valid. The algorithm could be described without mention of the term. It could be described as face recognition across pose, the technical meat consisting simply of using a eigenspace computed on pairs (or n-tuples) of views and the (standard) approach to dealing with occlusions in the eigenspace approach [16]. Such a paper could have been published right after [21] as a simple extension to the view-based approach described there to avoid the need for gallery images of each subject from each pose.

There is another reason we used the term. Appearance-based face recognition uses pixels in images as features. The reason for using the term light-field was to highlight the relationship between these two quantities and the implications for appearance-based object recognition. In

our "appearance-based" algorithm we include a step (vectorization) to convert the light-field into a vector of pixels that is used as the input to our appearance-based algorithm (PCA followed by nearest neighbor.) Vectorization converts measurements of light into features for a pattern recognition algorithm. When described for light-fields this step seems somewhat ad-hoc, and it is questionable whether it could ever be reliably performed in a fully automatic face recognition system. There is an equivalent step in eigenfaces or any other appearance-based algorithm, but because this step is essentially a "null step" it is often overlooked. Any face recognition algorithm must, either explicitly or implicitly, convert measurements of light (the light field, or pixels in images) into measurements of the face that can then be passed to a classification algorithm. In frontal appearance-based face recognition, this conversion reduces to a null step. Fortunately (or perhaps unfortunately) good results are often obtained even with this choice. By presenting an appearance-based algorithm in terms of a light-field we hope to illustrate the importance of this step, rather than glossing over it as appearance-based approaches have encouraged people to do.

4.3 Comparison with Model-Based Algorithms

How should the conversion from measurements of light to measurements of the object (face) be performed? Perhaps the most promising approach at this time is the model-based approach, best exemplified by Active Appearance Models [10] and 3D Morphable Models [9, 24]. Fitting a face model can be regarded as converting measurements of light into measurements of the face (i.e. the model parameters). Empirically there is considerable evidence to back up this approach. One example is the superior performance we obtained using the multi-point (Active Appearance Model like) normalization over the simple 3-point (appearance-based like) normalization. Another example is the excellent results obtained by Vetter *et al.* [9, 24]. Concurrent with the research in this paper, we provided the PIE database to Vetter *et al.* [9, 24] to provide a comparison between the appearance-based approach (this paper) and the model based approach ([9, 24]). The results presented in [9, 24] are considerably better than those in this paper, in our opinion clearly demonstrating that the model-based approach is preferable to the appearance-based approach. It might be argued that this conclusion is only valid for the (largely 3D) task of face recognition across pose. We believe otherwise. Although the task of fitting the face model (a task we have been actively

working on [18]) currently makes the model-based approach less robust, we believe the approach is inherently superior for both (2D) frontal face recognition as well as face recognition across pose. The reason is that the model-based approach explicitly addresses the question of how to convert measurements of light (the light-field) into measurements of the face (a vector of parameters.)

Acknowledgments

Much of Section 3 first appeared in [14]. We would like to thank Terence Sim and Takeo Kanade for preliminary discussions on the light-field estimation algorithm and the reviewers of [14], [15], and this paper for their feedback. The research described in this paper was supported by U.S. Office of Naval Research contract N00014-00-1-0915 and in part by U.S. Department of Defense contract N41756-03-C4024. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

[1] E.H. Adelson and J. Bergen. The plenoptic function and elements of early vision. In Landy and Movshon, editors, *Computational Models of Visual Processing*. MIT Press, 1991.

[2] S. Baker, T. Sim, and T. Kanade. When is the shape of a scene unique given its light-field: A fundamental theorem of 3D vision? *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(1):100–109, 2003.

[3] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[4] P.N. Belhumeur and D.W. Jacobs. Comparing images under variable illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 610–617, 1998.

[5] P.N. Belhumeur and D.J. Kriegman. What is the set of images of an object under all possible lighting conditions? *International Journal of Computer Vision*, 28(3):1–16, 1998.

[6] D. Beymer and T. Poggio. Face recognition from one example view. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 500–507, June 1995.

[7] M. Black and A. Jepson. Eigen-tracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 36(2):101–130, 1998.

[8] D.M. Blackburn, M. Bone, and P.J. Phillips. Facial recognition vendor test 2000: Evaluation report, 2000.

[9] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illumination with a 3D morphable model. In *Proceedings of the Fifth International Conference on Face and Gesture Recognition*, pages 202–207, 2002.

[10] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[11] F. De La Torre and M Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1):117–142, 2003.

[12] A. Georghiades, P.N. Belhumeur, and D. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings of the Fourth International Conference on Face and Gesture Recognition*, pages 277–284, 2000.

[13] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH)*, pages 43–54, 1996.

[14] R. Gross, I. Matthews, and S. Baker. Eigen light-fields and face recognition across pose. In *Proceedings of the Fifth International Conference on Face and Gesture Recognition*, pages 1–7, 2002.

[15] R. Gross, I. Matthews, and S. Baker. Fisher light-fields for face recognition across pose and illumination. In *Proceedings of the German Symposium on Pattern Recognition (DAGM)*, pages 481–489, 2002.

[16] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.

[17] M. Levoy and M. Hanrahan. Light field rendering. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH)*, pages 31–41, 1996.

[18] I. Matthews and S. Baker. Active appearance models revisited. Technical Report CMU-RI-TR-03-02, Carnegie Mellon University Robotics Institute, April 2003.

[19] H. Murase and S.K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[20] S.G. Narasimhan and S.K. Nayar. Chromatic framework for vision in bad weather. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 598–605, 2000.

[21] A.P. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[22] P.J. Phillips, H. Moon, S. Rizvi, and P.J Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[23] T. Riklin-Raviv and A. Shashua. The Quotient image: Class based recognition and synthesis under varying illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 566–571, 1999.

[24] S. Romdhani, V. Blanz, and T. Vetter. Face identification by matching a 3D morphable model using linear shape and texture error functions. In *Proceedings of the European Conference on Computer Vision*, pages 3–19, 2002.

[25] H. Shum, K Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(9):855–867, 1995.

[26] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12), 2003.

[27] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519–524, 1987.

[28] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[29] M.A.O. Vasilescu and D. Terzopoulos. Multilinear image analysis for face recognition. In *International Conference on Pattern Recognition (ICPR)*, pages 511–514, 2002.

[30] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–741, 1997.