

Fisher Light-Fields for Face Recognition Across Pose and Illumination

Ralph Gross, Iain Matthews, and Simon Baker

The Robotics Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
{rgross, iainm, simonb}@cs.cmu.edu

Abstract. In many face recognition tasks the pose and illumination conditions of the probe and gallery images are different. In other cases multiple gallery or probe images may be available, each captured from a different pose and under a different illumination. We propose a face recognition algorithm which can use any number of gallery images per subject captured at arbitrary poses and under arbitrary illumination, and any number of probe images, again captured at arbitrary poses and under arbitrary illumination. The algorithm operates by estimating the *Fisher light-field* of the subject’s head from the input gallery or probe images. Matching between the probe and gallery is then performed using the Fisher light-fields.

1 Introduction

In many face recognition scenarios the pose of the probe and gallery images are different. The gallery contains the images used during training of the algorithm. The algorithms are tested with the images in the probe sets. For example, the gallery image might be a frontal “mug-shot” and the probe image might be a 3/4 view captured from a camera in the corner of the room. The number of gallery and probe images can also vary. For example, the gallery may consist of a pair of images for each subject, 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.

Face recognition across pose, i.e. face recognition where the gallery and probe images do not have the same poses, has received very little attention. Algorithms have been proposed which can recognize faces [1] (or more general objects [2]) at a variety of poses. However, most of these algorithms require gallery images at every pose. Algorithms have been proposed which do generalize across pose, for example [3], but this algorithm computes 3D head models using a gallery containing a large number of images per subject captured using controlled illumination variation. It cannot be used with arbitrary gallery and probe sets.

After pose variation, the next most significant factor affecting the appearance of faces is illumination. A number of algorithms have been developed for face recognition across illumination, but they typically only deal with frontal faces [4, 5]. Only a few approaches have been proposed to handle both pose and illumination variation at the same time. For example, [3] computes a 3D head

model requiring a large number of gallery images, and [6] fits a previously constructed morphable 3D model to single images. This last algorithm works well across pose and illumination, however, the computational cost is very high.

We propose an algorithm for face recognition across pose and illumination. Our algorithm can use any number of gallery images captured at arbitrary poses and under arbitrary illuminations, and any number of probe images also captured with arbitrary poses and illuminations. 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 [7] of the subject’s head. First, generic training data is used to compute a linear subspace of head light-fields, similar to the construction of Fisher-faces [4]. Light-fields are simply used rather than images. Given a collection of gallery or probe images, the projection into the subspace 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 [8, 9]. This simple linear algorithm can be applied to any number of images, captured from any poses under any illumination. Finally, matching is performed by comparing the probe and gallery Fisher light-fields using a nearest-neighbor algorithm.

2 Light-Fields Theory

The *plenoptic function* [10] or *light-field* [7] is a function which specifies the radiance of light in free space. It is 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. In 2D, the light-field of a 2D object is actually 2D rather, than the 3D that might be expected. See Figure 1 for an illustration.

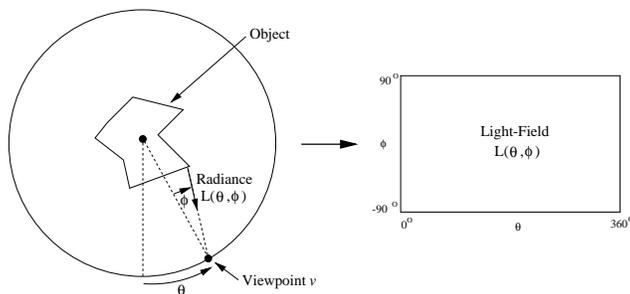


Fig. 1. The object is conceptually placed within a circle. The angle to the viewpoint v around the circle is measured by the angle θ , and the direction that the viewing ray makes with the radius of the circle is denoted ϕ . For each pair of angles θ and ϕ , the radiance of light reaching the viewpoint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

2.1 Eigen Light-Fields

Suppose we are given a collection of light-fields $L_i(\theta, \phi)$ of objects O_i (here faces of different subjects) where $i = 1, \dots, N$. See Figure 1 for the definition of this notation. If we perform an eigen-decomposition of these vectors using Principal Component Analysis (PCA), we obtain $d \leq N$ eigen light-fields $E_i(\theta, \phi)$ where $i = 1, \dots, d$. Then, assuming that the eigen-space of light-fields is a good representation of the set of light-fields under consideration, we can approximate any light-field $L(\theta, \phi)$ as:

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

where $\lambda_i = \langle L(\theta, \phi), E_i(\theta, \phi) \rangle$ 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 [2, 11]; The mean light-field could also be estimated and subtracted from all of the light-fields.

Capturing the complete light-field of an object is a difficult task, primarily because it requires a huge number of images [7, 12]. In most object recognition scenarios it is unreasonable to expect more than a few images of the object; often just one. 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 eigen coefficients λ_i be estimated from this highly occluded view? Although this may seem hopeless, consider that light-fields are highly redundant, especially for objects with simple reflectance properties such as Lambertian. An algorithm is presented in [8] to solve for the unknown λ_i for eigen-*images*. A similar algorithm was implicitly used in [9]. Rather than using the inner product $\lambda_i = \langle L(\theta, \phi), E_i(\theta, \phi) \rangle$, Leonardis and Bischof [8] solve for λ_i as the least squares solution of:

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

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 the solution of Equation (2) will be exactly the same as that obtained using the inner product [13]. Since there are d unknowns ($\lambda_1 \dots \lambda_d$) 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 [8]. Given an image $I(m, n)$, the following is then an algorithm for estimating the eigen light-field coefficients λ_i :

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 of the camera to the object.)

2. Find the least-squares solution (for $\lambda_1 \dots \lambda_d$) to the set of equations:

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

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 (so long as the camera intrinsics and relative orientation to the object are known for each image). The extra pixels from the other images are simply added in as additional constraints on the unknown coefficients λ_i in Equation (3). The algorithm can be used to estimate a light-field from a collection of images. Once the light-field has been estimated, it can then be used to render new images of the same object under different poses. (See [14] for a related algorithm.) In [13] we show that the algorithm correctly re-renders a given object assuming a Lambertian reflectance model.

2.2 Fisher Light-Fields

Suppose now we are given a set of light-fields $L_{i,j}(\theta, \phi)$, $i = 1, \dots, N, j = 1, \dots, M$ where each of N objects O_i is imaged under M different illumination conditions. We could proceed as described above and perform Principal Component Analysis on the whole set of $N \times M$ light-fields. An alternative approach is Fisher's Linear Discriminant (FLD) [15], also known as Linear Discriminant Analysis (LDA) [16], which uses the available class information to compute a projection better suited for discrimination tasks. Analogous to the algorithm above, we now find the least squares solution to the set of equations:

$$L(\theta, \phi) - \sum_{i=1}^m \lambda_i W_i(\theta, \phi) = 0 \quad (4)$$

where $W_i, i = 1, \dots, m$ are the generalized eigenvectors computed by the LDA. The extension of eigen light-fields to Fisher light-fields mirrors the step from eigenfaces to Fisher-faces in face recognition as proposed in [4].

3 Face Recognition Across Pose and Illumination

We will evaluate our algorithm on a subset of the CMU PIE database [17]. In the PIE database 68 subjects are imaged under 13 different poses and 21 different illumination conditions (see Figure 2). Many of the illumination directions introduce fairly subtle variations in appearance so we selected 12 of the 21 illumination conditions which span the set of variation widely. In total we use 10,608 images in the experiments.

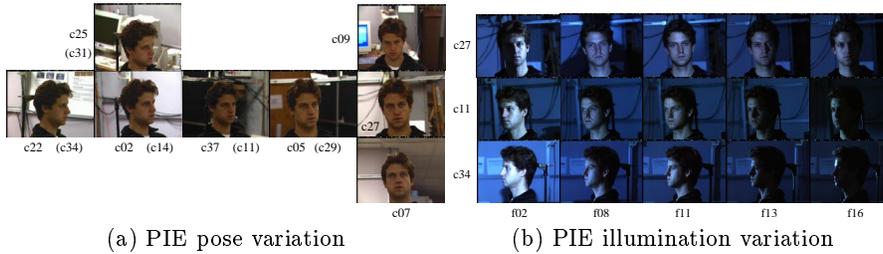


Fig. 2. The pose and illumination variation in the PIE database [17].(a) The pose varies from full right profile (c22) to full left profile (c34). (b) The illumination part shows 5 of the 12 illumination conditions used in the experiments here.

3.1 Gallery, Probe, & Generic Training Data

There are 68 subjects in the PIE database. We randomly select $N = 34$ of these subjects and use the images spanning all pose and illumination conditions as generic training data to construct the Fisher light-fields. Of the remaining 34 subjects, the images are divided into *completely disjoint* gallery and probe sets based on the pose and illumination condition they are captured in.

We determined the x-y positions of both eyes (pupils) and the tip of the nose in all 10,608 images we use in the experiments. Within each pose separately the face images are normalized for rotation, translation, and scale. The face region is then tightly cropped using the normalized feature point distances. See [13] for more details of this step.

3.2 Constructing the Light-Field Subspace

Suppose $Training \subset \{1, 2, \dots, 68\}$ denotes the set of generic training subjects, $Gallery_P$ and $Probe_P$ are the gallery and probe poses and $Gallery_I$ and $Probe_I$ are the gallery and probe illumination conditions. Note that $Gallery_P \cap Probe_P = \emptyset$ and $Gallery_I \cap Probe_I = \emptyset$ holds. We then assemble the set of images $I_{Train} =$

$$\{Im_{s,p,i} \mid s \in Training, p \in Gallery_P \cup Probe_P, i \in Gallery_I \cup Probe_I\}$$

The images are raster-scanned and concatenated. Between 7,000 and 14,000 pixels are extracted from each image depending on the pose. PCA and LDA is performed on these 408 vectors (34 subjects under 12 illumination conditions) to form the vectors W_i .

3.3 Processing the Gallery and Probe Images

Although the derivation in Section 2 is in terms of the entire light-field, the results also clearly hold for any subset of rays (θ, ϕ) in the light-field. We therefore do not need to densely sample the entire light-field to be able to use the algorithm. For each W_i ($i = 1, \dots, m$, the dimension of the light-field subspace) we extract the elements corresponding to the gallery and probe images and form shorter vectors W_i^G and W_i^P . For each non-training subject id $\notin Training$ we raster-scan the set of images:

$$\{Im_{id,p,i} \mid p \in Gallery_P, i \in Gallery_I\}, \{Im_{id,p,i} \mid p \in Probe_P, i \in Probe_I\}$$

to form a set of vectors of the same length. We solve Equation (4) for these shortened vectors resulting in λ_i^{id} for the gallery images and μ_i^{id} for the probe images.

3.4 The Classification Algorithm

In order to determine the closest gallery vector for each probe vector we perform nearest neighbor classification using the L_2 and Mahalanobis distance metrics on the PCA and the FLD subspaces. For each probe subject id we determine id_{\min} as

$$\text{id}_{\min} = \arg \min_{id^*} d(\mu^{id}, \lambda^{id^*}), d \in \{d_{L_2}, d_{Mahal}\} \quad (5)$$

If $\text{id}_{\min} = \text{id}$ the algorithm has correctly recognized the subject.

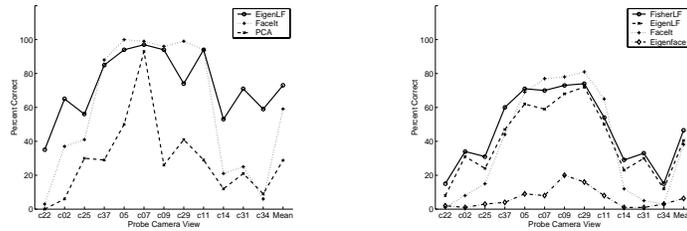
4 Experimental Results

We previously showed [13] that our algorithm outperforms eigenfaces [11] and FaceIt, the commercial face recognition system from Visionics. In Figure 3(a) we compare the recognition rate of the three algorithms for the gallery $E^G = \{27\}$ (frontal view). On average our algorithm achieves a recognition accuracy of 73% vs. 59% for FaceIt and 29% for eigenfaces. All images involved in the test were recorded with the same constant illumination. We also showed in [13] that the performance of our algorithm improves with the number of gallery images and that the role of the gallery and probe sets are approximately interchangeable. The average recognition accuracies are summarized in Table 1.

Table 1. Comparison of FaceIt, eigenfaces, eigen light-fields and Fisher light-fields over all three conditions. Due to the time constraints given for the preparation of this paper, complete FaceIt results are not presented. They will be made available at [18].

| Condition | FaceIt | Eigenface | Eigen LF | Fisher LF |
|------------------------------------|--------|-----------|----------|-----------|
| Varying pose, same illumination | - | 0.24 | 0.73 | - |
| Varying pose, varying illumination | 0.16 | 0.08 | 0.22 | 0.36 |
| Same pose, varying illumination | - | 0.60 | 0.60 | 0.81 |

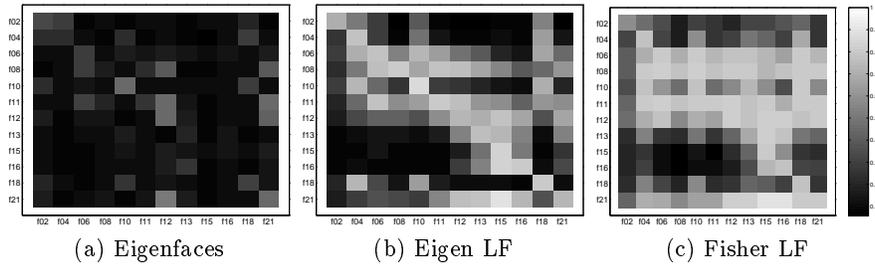
In Figure 3(b) we show a comparison between two light-field variants, FaceIt and eigenfaces for the gallery $E^G = \{27\}$ with frontal illumination. Here the recognition accuracies for the probe camera poses are obtained by averaging the results of testing the gallery illumination condition against a set of probe illumination conditions. Overall the Fisher light-field performs better (47% accuracy) than the eigen light-field (41% accuracy). The average accuracy for FaceIt is 38%. Eigenfaces perform poorly across most probe poses with an average accuracy of 6%. Figure 4 visualizes the differences in performance for eigenfaces and two light-field variants across all possible gallery and probe illumination conditions for the gallery $E^G = \{27\}$ (frontal view) and probe $E^P = \{37\}$ (3/4 view). The eigen light-field performs well close to the diagonal of the confusion matrix,



(a) Across pose (same illumination) (b) Across pose and illumination

Fig. 3. (a) A comparison of our algorithm with eigenfaces [11] and Facelt on the gallery $E^G = \{27\}$. The recognition rate of the eigen light-fields and Facelt is similar for the cameras $\{05, 07, 09\}$ closest to the gallery. For the profile views $\{02, 22, 25, 31, 34\}$ our algorithm outperforms Facelt by far. (b) A comparison of two light-field variants with Facelt and eigenfaces across pose and illumination with gallery $E^G = \{27\}$ and frontal gallery illumination. For each probe pose the accuracy is averaged over a set of probe illumination conditions.

whereas the Fisher light-field performs well across a broader range of conditions. Eigenfaces perform poorly in all tests. The average recognition accuracies are summarized in Table 1.



(a) Eigenfaces (b) Eigen LF (c) Fisher LF

Fig. 4. Comparison of eigenfaces with two light-field variants across all illumination conditions with gallery pose $E^G = \{27\}$ (frontal view) and probe pose $E^P = \{37\}$ (3/4 view). The gallery illumination conditions are shown along the x-axis, the probe illumination conditions along the y-axis.

In the case of identical gallery and probe poses, Fisher light-fields are identical to Fisherfaces [4]. As a baseline experiment we compare the recognition accuracies for identical gallery and probe poses across all illumination conditions. The results in Table 1 show that Fisher light-fields/Fisherfaces outperform eigen light-fields/eigenfaces by a large margin.

5 Discussion

In this paper we proposed an algorithm to recognize faces across pose and illumination. We have simplified this task in several ways: (1) the poses of the cameras are known and fixed, (2) the locations of the eyes and the nose used to extract the face region are marked by hand, and (3) the generic training data is captured with the same cameras that are used to capture the gallery and probe images. All of these factors make face recognition easier and are limitations on

the current algorithm. We are continuing to develop our algorithm to remove these limitations, while retaining the desirable properties of the algorithm.

To address part (3) we recently conducted preliminary experiments using PIE images as generic training data and FERET images as gallery and probe images. Our algorithm achieves a recognition accuracy of 81.3%, which compares very well to the performance of FaceIt over the same dataset (84.4%).

Acknowledgements

The research described in this paper was supported by U.S. Office of Naval Research contract N00014-00-1-0915. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

1. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: Proc. of CVPR. (1994)
2. Murase, H., Nayar, S.: Visual learning and recognition of 3-D objects from appearance. *IJCV* **14** (1995) 5–24
3. Georgiades, A., Belhumeur, P., Kriegman, D.: From few to many: Generative models for recognition under variable pose and illumination. In: Proc. of the 4th Conf. on Face and Gesture Recognition. (2000)
4. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE PAMI* **19** (1997) 711–720
5. Batur, A., Hayes, M.: Linear subspaces for illumination robust face recognition. In: Proc. of the 2001 CVPR. (2001)
6. Blanz, V., Romdhani, S., Vetter, T.: Face identification across different poses and illumination with a 3d morphable model. In: Proc. of the 5th Conf. on Face and Gesture Recognition. (2002)
7. Levoy, M., Hanrahan, M.: Light field rendering. In: Proc. of SIGGRAPH. (1996)
8. Leonardis, A., Bischof, H.: Dealing with occlusions in the eigenspace approach. In: Proc. of CVPR. (1996)
9. Black, M., Jepson, A.: Eigen-tracking: Robust matching and tracking of articulated objects using a view-based representation. *IJCV* **36** (1998) 101–130
10. Adelson, E., Bergen, J.: The plenoptic function and elements of early vision. In Landy, Movshon, eds.: *Computational Models of Visual Processing*. MIT Press (1991)
11. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: CVPR. (1991)
12. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The lumigraph. In: SIGGRAPH. (1996)
13. Gross, R., Baker, S., Matthews, I.: Eigen light-fields and face recognition across pose. In: Proc. of the 5th Conf. on Face and Gesture Recognition. (2002)
14. Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image. *IEEE Trans. on PAMI* **19** (1997) 733–741
15. Fukunaga, K.: *Introduction to statistical pattern recognition*. Academic Press (1990)
16. Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D., Weng, J.: Discriminant Analysis of Principal Components for Face Recognition. In: *Face Recognition: From Theory to Applications*. Springer Verlag (1998)
17. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: Proc. of the 5th Conf. on Face and Gesture Recognition. (2002)
18. <http://www.hid.ri.cmu.edu/FisherLF>.