ROBUST APPEARANCE MATCHING WITH FILTERED COMPONENT ANALYSIS

Fernando De la Torre, Alvaro Collet, Jeffrey F. Cohn and Takeo Kanade
Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

Keywords: Appearance Models, principal component analysis, Multi-band representation, learning filters.

Abstract: Appearance Models (AM) are commonly used to model appearance and shape variation of objects in images. In particular, they have proven useful to detection, tracking, and synthesis of people's faces from video. While AM have numerous advantages relative to alternative approaches, they have at least two important drawbacks. First, they are especially prone to local minima in fitting; this problem becomes increasingly problematic as the number of parameters to estimate grows. Second, often few if any of the local minima correspond to the correct location of the model error. To address these problems, we propose Filtered Component Analysis (FCA), an extension of traditional Principal Component Analysis (PCA). FCA learns an optimal set of filters with which to build a multi-band representation of the object. FCA representations were found to be more robust than either grayscale or Gabor filters to problems of local minima. The effectiveness and robustness of the proposed algorithm is demonstrated in both synthetic and real data.

1 INTRODUCTION

Component Analysis (CA) methods such as Principal Component Analysis (PCA) have been widely applied in visual, graphics, and signal processing tasks over the last two decades. PCA is a key learning component of Appearance Models (AM). AM have proven especially powerful for face tracking and synthesis relative to alternative approaches (e.g. optical flow) (Blanz and Vetter, 1999; Matthews and Baker, 2004; Cootes and Taylor, 2001b; de la Torre and Black, 2003; Black and Jepson, 1998).

In applications such as face detection and tracking, the goal is to search for a minimum residual between the image and the model across rigid (e.g. rotation and translation) and non-rigid parameters. For instance, consider fig. 1, in which a face has been placed in an arbitrary image. In fig. 1.a, we plot the normalized correlation surface error between the ideal template (face) and the image in a 40×40 patch centered in the middle of the face. This surface error has nice local properties: it has just one well defined global minimum that corresponds to the expected location of the face. However, if we learn a generic PCA model of the facial appearance variation from training data and try to locate the face again, two undesirable effects may occur. First, the location of the optimal

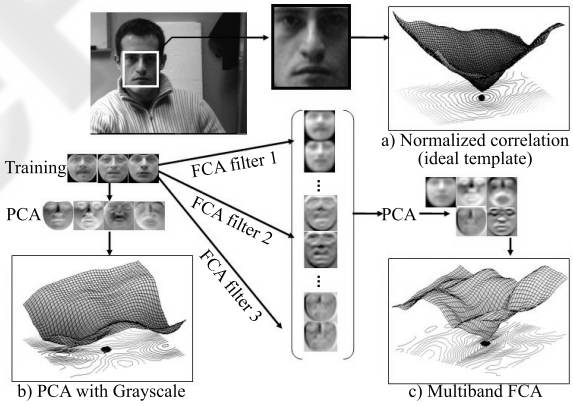

Figure 1: a). Normalized correlation error surface of the image with the face in a 40×40 patch. b) Error function with a generic graylevel appearance model. The black dot denotes the optimal position of the face. c) Error function of a multiband learned representation. The location of the face corresponds to the minimum of the function.

De la Torre F., Collet A., F. Cohn J. and Kanade T. (2007).
ROBUST APPEARANCE MATCHING WITH FILTERED COMPONENT ANALYSIS.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IFP/IA*, pages 207-212

parameter (translation) fails to correspond to the location of the face (delineated by the the black dot in the figure), see fig. (1.b). Second, many local minima may be found. Even if a gradient descent algorithm begins close to the correct solution, the occurrence of local minima is likely to divert convergence from the desired solution.

The aim of this paper is to explore the use of a new technique, Filtered Component Analysis (FCA). FCA learns a multiband representation of the image that reduces the number of local minima and improves generalization relative to PCA. Fig. (1.c) shows the main goal of the paper. By building a multiband representation with FCA, we are able to locate the minimum in the right location (black dot) and eliminate most local minima close to the optimal one.

2 PREVIOUS WORK

This section reviews work on subspace tracking and the role of representation in subspace analysis.

2.1 Subspace Detection and Tracking

Subspace trackers build the object's appearance/shape representation from the PCA of a set of training samples. Let $\mathbf{d}_i \in \Re^{d\times 1}$ (see notation [1]) be the i sample of a training set $\mathbf{D} \in \Re^{d\times n}$ and $\mathbf{B} \in \Re^{d\times k}$ the first k principal components. $\mathbf{B}$ contains the directions of maximum variation of the data. The principal components maximize $\max_{\mathbf{B}} \sum_{i=1}^{n} ||\mathbf{B}^T\mathbf{d}_i||_2^2 = ||\mathbf{B}^T \blacksquare \mathbf{B}||_F$, with the constraint $\mathbf{B}^T\mathbf{B} = \mathbf{I}$. The columns of $\mathbf{B}$ form an orthonormal basis that spans the principal subspace. If the effective rank of $\mathbf{D}$ is much less than d, we can approximate the column space of $\mathbf{D}$ with $k << d$ principal components. The data $\mathbf{d}_i$ can be approximated as a linear combination of the principal components as $\mathbf{d}_i \approx \mathbf{B}\mathbf{c}_i$ where $\mathbf{c}_i = \mathbf{B}^T\mathbf{d}_i$ are the coefficients obtained by projecting the training data onto the principal subspace.

Once the model has been learned (i.e. $\mathbf{B}$ is known), tracking is achieved by finding the parameters $\mathbf{a}$ of the geometric transformation $\mathbf{f}(\mathbf{x},\mathbf{a})$ that aligns the data w.r.t. the subspace. Given an image $\mathbf{d}_i$, subspace trackers or detectors find $\mathbf{a}$ and $\mathbf{c}_i$ that minimize: $min_{\mathbf{c}_i,\mathbf{a}} ||\mathbf{d}_i(\mathbf{f}(\mathbf{x},\mathbf{a})) - \mathbf{B}\mathbf{c}_i||_2^2$ (or some normalized error). In the case of an affine transformation, $\mathbf{f}(\mathbf{x},\mathbf{a}) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} a_3 & a_4 \\ a_5 & a_6 \end{pmatrix} \begin{pmatrix} x-x_c \\ y-y_c \end{pmatrix}$ where $\mathbf{a} = (a_1,a_2,a_3,a_4,a_5,a_6)$ are the affine parameters and $\mathbf{x} = (x_1,y_1,\cdots,x_n,y_n)$ is a vector containing the coordinates of the pixels to track. If $\mathbf{a} = (a_1,a_2)$ is just translation, the search can be done efficiently over the whole image using the Fast Fourier Transform (FFT). For $\mathbf{a} = (a_3 = a_6, a_5 = a_4)$, that is, for similarity transformation, the search also can be done efficiently in the log-polar representation of the image with the FFT.

2.2 Representation in Subspace Analysis

Most work on AM uses some sort of *normalized* graylevel to build the representation. However, regions of graylevel values can suffer from large ambiguities, camera noise, and changes in illumination. More robust representation can be achieved by local combination of pixels through filtering. Filtering of the visual array is a key element of the primate visual system (Rao and Ballard, 1995).

Representations for subspace recognition were explored by Bischof et al. (Bischof et al., 2004). In the training stage, they built a subspace by filtering the PCA-graylevel basis with steerable filters. In the recognition phase, they filtered the test images and performed robust matching, obtaining improved recognition performance over graylevel. Cootes et. al (Cootes and Taylor, 2001a) found that a non-linear representation of edge structure could improve the performance of model subspace matching and recognition. De la Torre et al. (de la Torre et al., 2000) found that subspace tracking was improved by using a multiband representation created by filtering the images with a set of Gaussian filters and its derivatives.

Our work differs in several aspects from previous work. First, we explicitly learn an optimal set of spatial filters adapted to the object of interest rather than using hand-picked ones. Once the filters are learned, we build a multiband representation of the image that has improved error surfaces with which to fit AM. We evaluate quantitatively the properties of the error surfaces and show how FCA outperforms current methods in appearance based detection.

[1] Bold capital letters denote a matrix $\mathbf{D}$, bold lower-case letters a column vector $\mathbf{d}$. $\mathbf{d}_j$ represents the j column of the matrix $\mathbf{D}$. d_{ij} denotes the scalar in the row i and column j of the matrix $\mathbf{D}$ and the scalar i-th element of a column vector $\mathbf{d}_j$. All non-bold letters will represent variables of scalar nature. $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$ designates Euclidean norm of $\mathbf{x}$. The $vec(\mathbf{D})$ operator transforms $\mathbf{D} \in \Re^{d\times n}$ into an dn-dimensional vector by stacking the columns. $\circ$ denotes the Hadamard or point-wise product. $\otimes$ denotes convolution. $\mathbf{1}_k \in \Re^{k\times 1}$ is a vector of ones. $\mathbf{I}_k \in \Re^{k\times k}$ is the identity.

3 FILTERED COMPONENT ANALYSIS

Many component analysis methods (PCA, LDA, etc) build data models based on the second order statistics (covariance matrices) of the signal. In particular, PCA finds a linear transformation that decorrelates the data by exploiting the correlation across samples. PCA models the correlation across pixels of different images, but not the spatial statistics within each of the images. In this section, we propose Filtered Component Analysis (FCA) that learns a bank of orthogonal filters that decorrelate the spatial statistics of a set of images. Once the FCA filters are learned, we build a multi-band representation that generalizes better and is more robust to different types of noise.

3.1 Learning Spatial Correlation

Previous research (de la Torre et al., 2000; Bischof et al., 2004; Cootes and Taylor, 2001a) has shown the importance of representation in AM. However, researchers have used hand-picked filters to represent the signal. Instead, FCA will learn a set of orthogonal spatial filters optimal for variance preservation. Variance preservation of image spatial statistics is a realistic assumption to build a generative model for detection or tracking appearance.

Given a set of training images, $\mathbf{D}^{d\times n}$, our aim is to model the spatial statistics of the signal by learning the filter $\mathbf{F}$ that minimizes:

$$E_1(\mathbf{F},\boldsymbol{\mu}) = \min_{\mathbf{F},\boldsymbol{\mu}} \sum_{i=1}^{n} ||\mathbf{d}_i \otimes \mathbf{F} - \boldsymbol{\mu}||_2^2 \qquad (1)$$

Recall that $\otimes$ denotes convolution, and $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i \otimes \mathbf{F}$ is the mean of the filtered signal. If $\boldsymbol{\mu}$ is known, the optimal $\mathbf{F}$ can be achieved by solving:

$$\mathbf{A}vec(\mathbf{F}) = \mathbf{b} \qquad \mathbf{A} = \sum_{i=1}^{n}\sum_{(x,y)} \mathbf{d}_i^{(x,y)} \mathbf{d}_i^{(x,y)^T}$$
$$\mathbf{b} = \sum_{i=1}^{n}\sum_{(x,y)} \boldsymbol{\mu}^{(x,y)} \circ \mathbf{d}_i^{(x,y)} \qquad (2)$$

where (x,y) is the domain where the convolution is valid and $\mathbf{d}_i^{(x,y)}$ is a patch of the filter size (f_x, f_y) centered at the coordinates (x,y). The matrix $\mathbf{A}$ can be computed efficiently in space or frequency from the autocorrelation function of $\mathbf{d}_i$. Analogously, $\mathbf{b}$ is estimated from the cross-correlation between $\mathbf{d}_i$ and $\boldsymbol{\mu}$. Alternatively, one could use the integral image (Lewis, 1995) to efficiently compute eq. 2.

Without imposing any constraints on the filter coefficients, the optimal solution of eq. 1 is given by $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{F} = \mathbf{0}$ (although an iterative algorithm will rarely converge to this solution). To avoid this trivial solution, we impose that the sum of squared coefficients is 1, i.e. $vec(\mathbf{F})^T vec(\mathbf{F}) = 1$. The latter constraint can be elegantly solved by noticing that the convolution is a linear operator, and , eq. 2 can be rewritten as:

$$E_2(\mathbf{F}) = \min_{\mathbf{F}} \sum_{i=1}^{n} ||(\mathbf{d}_i - \boldsymbol{\mu}') \otimes \mathbf{F}||_2^2 \qquad (3)$$

where $\boldsymbol{\mu}' = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i$ is the sample mean. Now Eq. 3 can be solved by finding the eigenvector with smallest eigenvalue of $\mathbf{A} = \sum_{i=1}^{n}\sum_{(x,y)} (\mathbf{d}_i - \boldsymbol{\mu}')^{(x,y)} (\mathbf{d}_i - \boldsymbol{\mu}')^{(x,y)^T}$ (see eq. 2).

3.2 Learning a Multiband Representation

In this section, we will build a multiband representation of the signal that preserves most spatial correlation among a given training set. In particular, we will find a set of filters $\mathbf{F}^{1,\cdots,F}$ that decorrelate the spatial statistics of the image and are orthogonal to each other. Observe that FCA is analogous to PCA but now rather than decorrelating the signal with the covariance of the data, we decorrelate the spatial statistics.

In our particular tracking application, we are interested in finding a set of filters that preserve the spatial statistics of the object of interest and has minimal response to background. This filter set can be obtained by maximizing $E_{FCA}(\mathbf{F}^{1,\cdots,F})$:

$$E_{FCA} = \sum_{f=1}^{F}\sum_{i=1}^{n} ||\mathbf{d}_i \otimes \mathbf{F}^f||_2^2 - \lambda \sum_{j=1}^{n_2} ||\mathbf{d}_j^b \otimes \mathbf{F}^f||_2^2 \qquad (4)$$

where $\mathbf{d}_j^b$ denotes the j^{th} sample of the background. Let $\mathbf{T} = [vec(\mathbf{F}^1)\ vec(\mathbf{F}^2)\ \cdots\ vec(\mathbf{F}^F)]$ be a matrix of all the vectorized filters, the filters should satisfy $\mathbf{T}^T\mathbf{T} = \mathbf{I}_{F\times F}$. After making the derivatives with respect to $\mathbf{F}^f$, it can be shown that the optimal solutions satisfies the following eigenvalue problem:

$$max_{\mathbf{F}^{1,\cdots,F}} \sum_{i=1}^{f} ||(\mathbf{A} - \lambda\mathbf{B}\alpha)vec(\mathbf{F}^i)||_2^2 \qquad (5)$$
$$\mathbf{A} = \sum_{i=1}^{n}\sum_{(x,y)} \mathbf{d}_i^{(x,y)} \mathbf{d}_i^{(x,y)^T} \quad \alpha = \frac{max(\mathbf{A})}{max(\mathbf{B})}$$
$$\mathbf{B} = \sum_{j=1}^{n_2}\sum_{(x,y)} \mathbf{d}_j^{b(x,y)} \mathbf{d}_j^{b(x,y)^T}$$
$$s.t.\ vec(\mathbf{F}^i)^T vec(\mathbf{F}^j) = 0\ \forall i \neq j \quad and$$
$$vec(\mathbf{F}^i)^T vec(\mathbf{F}^i) = 1\ \forall i$$

If λ is large, the set of filters will predominantly cancel the background. If λ is small the filters will be adapted to the object.With λ close to one the filters will achieve trade-off between modeling the signal (i.e object) and removing the background. Typically

$0 \leq \lambda \leq 2$. α is an artificially introduced parameter that normalizes the energies of $\mathbf{A}$ and $\mathbf{B}$.

The solution of eq. 5 is given by the leading eigenvectors of $(\mathbf{A} - \lambda\alpha\mathbf{B})$. At this point, it is interesting to consider again the analogy with PCA. PCA will find the leading eigenvectors of $\sum_{i=1}^{n} \mathbf{d}_i \mathbf{d}_i^T$ whereas FCA will find the leading eigenvectors (assuming $\lambda = 0$) $\mathbf{A} = \sum_{i=1}^{n} \sum_{(x,y)} \mathbf{d}_i^{(x,y)} \mathbf{d}_i^{(x,y)^T}$. While PCA finds the directions of maximum variation of the covariance matrix, FCA finds the directions of maximum variation of the sum of all overlapping patches.

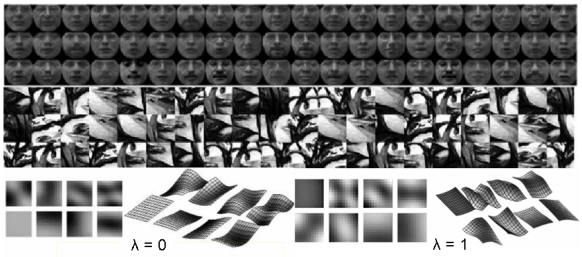

Figure 2: a) Training images of faces and background. b) FCA filters for $\lambda = 0$, $\lambda = 1$ and size 11×11.

Fig. (2.a) shows many examples of faces and background patches. Fig.(2.b) shows the set of FCA filters for $\lambda = 0$ and $\lambda = 1$ for size 11×11. Observe that the first FCA filter is an average filter, and the other filters are differential filters at different orientations and scales.

3.3 Multiband Subspace Detection

In subspace detection, PCA is computed from a set of training images. After the training stage, the goal is to detect the object of interest over different orientation, scales and translations. If the scale and orientation is known, dectection can be achieved finding the translational parameters $\mathbf{a} = (a_1, a_2)$ that minimize:

$$E_3 = min_{\mathbf{c}_i,\mathbf{a}} \frac{||\mathbf{d}_i(\mathbf{x}+\mathbf{a}) - \mathbf{B}\mathbf{c}_i||_2^2}{||\mathbf{d}_i(\mathbf{x}+\mathbf{a})||_2^2} \quad (6)$$

Evaluating eq. 6 at each location (x,y) can be computationally expensive. For a particular position (x,y) computing the coefficients (i.e. $\mathbf{c}_i$) is equivalent to correlating the image with each basis of subspace $\mathbf{B}$, and stacking all values for each pixel. For large regions, this correlation is performed efficiently in the frequency domain using the Fast Fourier Transform (FFT) (i.e. $\mathbf{C}_1 = \mathbf{b}_1^T \mathbf{I} = IFFT(FFT(\mathbf{b}_1) \circ FFT(\mathbf{I}))$). Similarly, the local energy term, $||\mathbf{d}_i(\mathbf{x}+\mathbf{a})||_2^2$, can be computed efficiently using the convolution in the space or frequency domain. Alternatively, these expressions can be computed efficiently using the integral image (Lewis, 1995).

In multiband tracking, we represent an image as a concatenation of filtered images. For a particular image $\mathbf{d}_i$ and a set of filters $(\mathbf{F}^1, \cdots, \mathbf{F}^f)$, there are several ways to modify eq. 6:

$$E_4 = \sum_{f=1}^{F} \beta_f \frac{||\mathbf{d}_i \otimes \mathbf{F}^f - \mathbf{B}^f \mathbf{c}_i||_2^2}{||\mathbf{d}_i \otimes \mathbf{F}^f||_2^2} \quad (7)$$

$$E_5 = \sum_{f=1}^{F} \beta_f \frac{||\mathbf{d}_i \otimes \mathbf{F}^f - \mathbf{B}^f \mathbf{c}_i^f||_2^2}{||\mathbf{d}_i \otimes \mathbf{F}^f||_2^2} \quad (8)$$

Parameters β_f are the eigenvalues of $(\mathbf{A} - \lambda\alpha\mathbf{B})$, obtained by FCA. E_4 filters the training images and builds PCA based on the set of stacked filtered images. E_5 computes an independent PCA for each representation such that the coefficients for each image are uncoupled (i.e. $\mathbf{c}_i^f$ differs for each filter).

4 EXPERIMENTS

To test the validity of our approach, we have performed several sets of experiments in face detection and facial feature tracking. The first set of experiments consists on detecting a face embedded in an arbitrary image (see fig. 1) using a generic model. In the second set, we test the ability of FCA to improve tracking in Active Appearance Models (Cootes and Taylor, 2001b; Blanz and Vetter, 1999; Matthews and Baker, 2004; de la Torre et al., 2000).

In all experiments a generic face model was built from 150 subjects from the IBM ViaVoice AV database (Neti et al., 2000), after aligning the data with Procrustes Analysis(Cootes and Taylor, 2001b). Once the FCA filters are learned, a multi-band representation is built for each of the 150 images, and PCA is computed retaining 80% of the total energy. For comparison purposes, multi-band PCA is also done for other representations (e.g. Gabor, graylevel and derivatives. In the experiments, we consider Gabor Filters because of the good results reported by other researchers in the area. In addition, these filters have been shown to possess optimal localization properties in both spatial and frequency domain and thus are well suited for tracking problems.

4.1 Understanding FCA

In order to compute FCA 150 subjects are selected randomly from the IBM database. We also extract 2000 random patches from several images of the IBM that do not contain faces. Using these training samples, we learn FCA filters at 5 different scales (3×3, 5×5, 7×7, 9×9 and 11×11 pixels), using eq. 5 for different λ values.

Given a new face image not present in the training set, we embedded it in a bigger background image (see fig. 3). Then, we efficiently search over all translations looking for a minimum of the subspace model. Fig. 3 shows an example of the error surface for each of the FCA bands, in comparison with the error surfaces from normalized graylevel. As it can be observed, Graylevel representation has several local minima and the global minimum is misplaced. On the other hand, the sum of the three FCA bands produces an error surface with a correctly-placed global minimum.

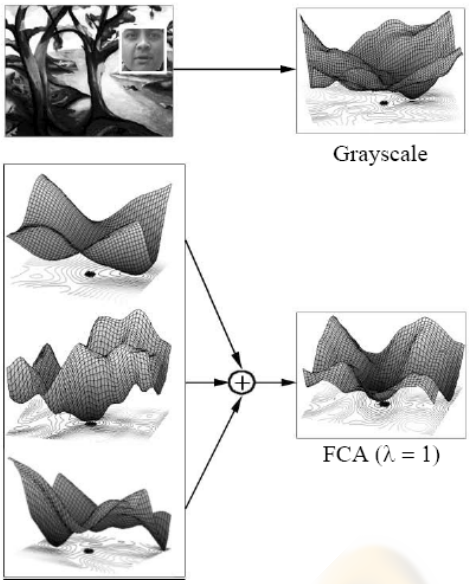

Figure 3: Error surfaces for graylevel and each of the bands for FCA.

4.2 Robustness to Noise/Illumination

This first experiment is designed to test the robustness of FCA to noise and varying illumination conditions. A subset of 100 subjects from the IBM database (not in the training set) are randomly chosen and embedded in background images. Then, random impulsional noise is added (see fig. 4.a) and the error in each location is efficiently computed (the orientation and scale is known). To quantitatively compare each filterbank, 3 different surface error statistics have been calculated. Given a patch of 100×100 pixels around the optimal location of the face (which is known beforehand), we compute the following statistics: 1) distance between the global minimum and the face center, 2) distance between the correct minimum and closest local minimum, 3) Amount of local minima. The amount of local minima in an error surface is calculated by counting those pixels with sign change in x and y derivatives and positive values in the second derivatives.

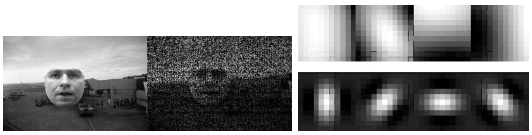

Figure 4: a) Original image and test image with added impulsional noise. b) FCA(11,4) and Gabor (8,4).

Table 1 shows the average results for the described error statistics for three representations: a set of four 11×11 pixels FCA filters (see fig. 4.a), the best-performing 11×11 pixels Gabor filter (see fig. 4.b) and the normalized graylevel. In all our experiments, we report the results of the set of Gabor filters that performs the best over several scales. A global minimum is said to be correct if it falls within a region of 3×3 pixels around the theoretical minimum. All the representations have similar accuracy; however, the amount of local minima is very high in the grayscale, and both grayscale and Gabor fail to provide a sufficiently high global-closest minimum margin in comparison with FCA filters. These results are quite stable across spatial domains of the FCA filter sets and have therefore been omitted in the interest of space.

Table 1: Experiments on noisy data. Statistics: (1) Percentage of correct global minimum. (2) distance between correct and closest local minimum. (3) Average number of local minima.

	gray	$FCA_{\lambda=0}$	$FCA_{\lambda=0.5}$	Gabor(8,4)
(1)	98	**99**	**99**	99
(2)	9.73	**24.36**	**24.03**	19.01
(3)	30.06	**1.45**	**1.49**	2.46

In the next experiment we test the robustness of FCA to illumination changes. We take 4 images under varying illumination conditions (see fig. 5) for 30 subjects from the PIE database(Sim et al., 2002) (total 120 images). We embedded this face into an image and compute the error surfaces. Results from this experiment can be seen in table 2. In this case, FCA clearly outperforms any other technique in all three statistics of the error function. The accuracy is higher than grayscale and Gabor by 33% and 12% resp., while keeping the closest minimum 25.37% pixels further away and the density of local minima is the lowest one. It is worth noting that the best-performing filters have been $FCA_{\lambda=0}$ (no background). Fig. 6 shows the error surface for a particular subject; as we can observe, the properties of FCA are more desirable

than graylevel or Gabor filters in terms of location and density of local minima.

Figure 5: Changes in illumination on the PIE database.

Table 2: Experiments on illumination.(1),(2),(3) see table1.

	gray	$FCA_{\lambda=0}$	$FCA_{\lambda=0.5}$	Gabor(8,4)
(1)	41	**74**	**73**	62
(2)	14.59	**26.37**	**26.04**	19.68
(3)	3.28	**1.4**	**1.41**	1.92

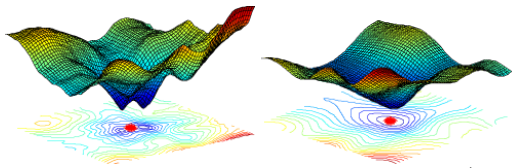

Figure 6: Error surface for **graylevel** and $FCA_{\lambda=0}(11,4)$.

The last experiment of this section explores FCA performance on images taken in the lab. 10 images have been collected in the lab (see Fig. 7) with an inexpensive webcam, and roughly selecting the same scale manually. Table 3 shows the detection results of this experiment. As we can see FCA consistently outperforms other representations that included Gabor and graylevel in all metrics.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed FCA to build a multi-band representation of the image to achieve more robust fitting and detection with appearance models. FCA outperforms Gabor, oriented pair filters and graylevel representations. Additionally, we have introduced quantitative metrics for evaluating the error surface. FCA has shown promising results, however future work should consider the use of different constraints for the filters (e.g. $vec(\mathbf{F})^T \mathbf{1}_{f_x \times f_y} = 1$). Also, it will be worth to explore the use of some recent non-linear filters.

Acknowledgements The work was partially supported by National Institute of Justice award 2005-IJ-CX-K067 and NIMH grant. Thanks to Iain Matthews and Simon Lucey for helpful discussions and comments.

Table 3: Experiments on images taken in the lab.(1), (2), (3) see table 1.

	gray	$FCA_{\lambda=0}$	$FCA_{\lambda=0.5}$	Gabor(8,4)
(1)	20	**80**	**80**	70
(2)	15.71	**18.05**	**25.52**	13.53
(3)	2	**2**	**1.2**	2.4

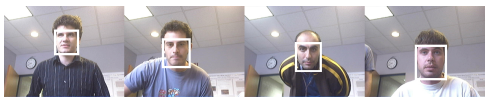

Figure 7: Some test images.

REFERENCES

Bischof, H., Wildenauer, H., and Leonardis, A. (2004). Illumination insensitive recognition using eigenspaces. *Computer Vision and Image Understanding*, 1(95):86 – 104.

Black, M. J. and Jepson, A. D. (1998). Eigentracking: Robust matching and tracking of objects using view-based representation. *International Journal of Computer Vision*, 26(1):63–84.

Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3d faces. In *SIGGRAPH*.

Cootes, T. and Taylor, C. (2001a). On representing edge structure for model matching. In *CVPR*.

Cootes, T. F. and Taylor, C. J. (2001b). Statistical models of appearance for computer vision. In *http://www.isbe.man.ac.uk/bim/refs.html*.

de la Torre, F. and Black, M. J. (2003). Robust parameterized component analysis: theory and applications to 2d facial appearance models. *Computer Vision and Image Understanding*, 91:53 – 71.

de la Torre, F., Vitrià, J., Radeva, P., and Melenchón, J. (2000). Eigenfiltering for flexible eigentracking. In *International Conference on Pattern Recognition*, pages 1118–1121.

Lewis, J. P. (1995). Fast normalized cross-correlation. In *Vision Interface*.

Matthews, I. and Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164.

Neti, C., Potamianos, G., Luettin, J., Matthews, I., Glotin, H., Vergyri, D., Sison, J., Mashari, A., and Zhou, J. (2000). Audio-visual speech recognition. Technical Report WS00AVSR, Johns Hopkins University, CLSP.

Rao, R. and Ballard, D. (1995). An active vision architecture based on iconic representations. *Artificial Intelligence*, 12:441–444.

Sim, T., Baker, S., and Bsat, M. (2002). The cmu pose, illumination, and expression (pie) database. In *IEEE Conference on Automatic Face and Gesture Recognition*.