

Robust Contrast-Invariant EigenDetection

Chakra Chennubhotla, Allan Jepson & John Midgley
Department of Computer Science
University of Toronto, 6 King's College Road
Toronto, ON M5S 3H5, Canada
Email: {chakra,jepson,jmidgley}@cs.toronto.edu

Abstract

We achieve two goals in this paper: (1) to build a novel appearance-based object representation that takes into account variations in contrast often found in training images; (2) to develop a robust appearance-based detection scheme that can handle outliers such as occlusion and structured noise. To build the representation, we decompose the input ensemble into two subspaces: a principal subspace (within-subspace) and its orthogonal complement (out-of-subspace). Before computing the principal subspace, we remove any dependency on contrast that the training set might exhibit. To account for pixel outliers in test images, we model the residual signal in the out-of-subspace by a probabilistic mixture model of an inlier distribution and a uniform outlier distribution. The mixture model, in turn, facilitates the robust estimation of the within-subspace coefficients. We show our methodology leads to an effective classifier for separating images of eyes from non-eyes extracted from the FERET dataset.

1 Introduction

Eigenspace representations enable an approximate but compact encoding of object-specific ensembles, such as a database of face images or images of a gesturing hand, using a small set of orthonormal basis images. The basis set is obtained by finding the principal components (PCA) of the image ensemble. The basis images span a subspace, also called the eigenspace, and a linear combination of these basis images can be used to approximately reconstruct images in the input ensemble. Eigenspace representations have proven to be useful in various contexts such as coding [9], appearance-based detection and recognition [13, 4, 6], and tracking [1].

Our work builds upon three key observations regarding eigenspace representation. First, eigenspace methods involve least squares approximations which are notoriously sensitive to large outliers. Hence, it is important to account for outliers both for training and testing eigenspace methods [1, 11]. A related issue is one of variation in the image contrast that is often found in the training data. Contrast is measured as the deviation from the average brightness. In general, images with a range of different contrasts will have the property that the images with higher contrasts will have a larger variance. If some images have much larger variance than others, then these images will dominate over others in the computation of the eigenspace. Often a contrast normalization step such as histogram equalization is performed. Such a method is unlikely to highlight object-specific structure that we wish

eigenspaces to capture.

Second, principal component analysis (PCA) amounts to a rotation of input coordinate axes and as such, PCA does not define a probability density model for the input data. However, eigenspaces can also be derived from the perspective of density estimation [4, 7, 12]. The advantage in estimating the input density is that it allows for the design of *probabilistic* methods to detect, recognize and/or classify test images as appearances of known objects. In particular, one strategy for object representation is to divide the signal space into a principal subspace and its orthogonal complement and then build probability models separately for the two subspaces. The detection strategy then is to apply a threshold on the likelihood assigned by the combined density model to a test image [4].

The third key observation, as noted in [5], is that the variance estimate *per-pixel* given by the eigenspace density models for the residual signal in the out-of-subspace is overly conservative. We defer the actual details of variance estimation used in our appearance-based detector to a later section.

In this paper we build on and extend the work done addressing the three issues mentioned. While we follow the strategy of splitting the signal space into a principal subspace and its orthogonal complement, we depart from the usual in several ways. First, we explicitly model any variations that the training set might exhibit with contrast. This allows us to build an eigenspace that is invariant to contrast changes. Also, we assume that only a small fraction of the training images have outliers, while we apply no such restriction on the test images.

Our second contribution is to invoke a novel mixture model formulation for the out-of-subspace residual signal. The mixture model consists of an inlier component and a uniform outlier distribution. We use a Gaussian distribution for the inlier but other densities such as Laplace can be used as well, as long as they provide a good fit to the out-of-subspace residuals. The mixture coefficients are updated using just one iteration of expectation maximization (EM) algorithm. The resulting pixel ownership probabilities are then used for a weighted least-square computation of the within-subspace coefficients. We show that on a database of eye images cropped from the FERET database of facial images, our contrast-invariant, robust eigenspace representation leads to an effective classifier.

2 Database

We generated an eye database by cropping eye regions from the FERET face database [2]. The original face images were scaled and rotated so that, in the warped image, the left and right eyes have a horizontal separation of 28 pixels. These warped images were then cropped to 20×25 regions centered on each eye. The images vary in contrast along with changes of the person, the lighting, position, pose (eg. open/closed eyes), and the occasional reflection from glasses. The resulting eye database contained 2392 images (Fig. 1). For non-eyes, we ran an interest point detector on many natural images and collected those image patches passing a suitable threshold [8]. The database of non-eyes consisted of 3839 images. Next, we use this database to explore the representation issues.

3 Contrast Model

Contrast is the variation in image intensities around a mean brightness value. We need to estimate the brightness variation that is due to the underlying signal, and not due to noise. We accomplish this by using two components, namely a constant ‘‘DC’’ component and a component in the direction of mean of the training set. We define the training set as $\{T_k(\vec{x})\}_{k=1}^K$ where K is the total number of training images and $T_k(\vec{x})$ is the gray value at pixel position \vec{x} in the k^{th} training image. The DC component $D(\vec{x})$ is given by $D(\vec{x}) = 1/\sqrt{N}$, where N is the total number of pixels in the image, \vec{x} corresponds to a pixel position and $D(\vec{x})$ is chosen to be unit length. The mean image $M(\vec{x})$ is taken to be the mean image of the training set minus their DC components, that is $M(\vec{x}) = \frac{s_m}{K} \sum_{k=1}^K [T_k(\vec{x}) - D(\vec{x})\langle T_k(\vec{x}), D(\vec{x}) \rangle]$. Here the inner product is defined by: $\langle T_k(\vec{x}), D(\vec{x}) \rangle \equiv \sum_{\vec{x}} T_k(\vec{x})D(\vec{x})$, where the sum is over all the pixel positions and s_m is chosen to make $M(\vec{x})$ a unit vector.

Define the components of $T_k(\vec{x})$ in the directions of DC and the mean vector as $d_k = \langle T_k(\vec{x}), D(\vec{x}) \rangle$ and $m_k = \langle T_k(\vec{x}), M(\vec{x}) \rangle$ respectively. To study contrast dependency, we project out the mean and DC vectors from the training set and observe the relationship between the *left-over variance*, given by $L_k = \frac{1}{N} \sum_{\vec{x}} [T_k(\vec{x}) - d_k D(\vec{x}) - m_k M(\vec{x})]^2$, and the mean component m_k .

It is possible that the spatial structure of eye images, which we wish to capture with the PCA basis $B_j(\vec{x})$, changes as the mean component m_k increases. Alternatively, these basis images may stay roughly the same and only the variation of their components given by $c_{k,j} = \langle T_k, B_j \rangle$ increases with the mean component m_k . Indeed, if the variation of the mean component is primarily due to lighting and imaging effects, then we might assume that the underlying signal is invariant to contrast. In this case we would expect the PCA basis images to be independent of m_k and only the variance of components $c_{k,j}$ to scale as a function of m_k .

In Fig. 2(Left) we plot $\sqrt{L_k}$ vs the mean coefficient for

each eye image in the FERET dataset. It is clear from the figure that $\sqrt{L_k}$ is linearly increasing with the mean coefficient. This is confirmed by the non-zero slope of the straight line fit to the data going through the origin (black line in Fig. 2(Left)). Note, the straight line fit may be biased in a small way by the outliers in the eye images.

To balance the variances across different image contrasts, define rescaled images as $A_k(\vec{x}) = [T_k(\vec{x}) - d_k D(\vec{x}) - m_k M(\vec{x})]/s(m_k)$, where $s(m_k)$ is a scaling factor. It is convenient to assume a minimum value for the left-over variance, say due to independent pixel noise, and use this in the estimate for scaling: $s(m_k) = \sqrt{\sigma_{\min}^2 + f(m_k)}$. The red curve in Fig. 2a is obtained with $f(m_k) = p \times m_k^2$, where p is the slope of the straight line fit (black line in Fig. 2(Left)) and all points on the red curve will have a variance of one (Fig. 2(Middle)).

Fig. 2(Right) reveals the property of non-eyes show in red, while the eyes are drawn in green. The vertical axis is indicative of the variance that could not be explained by the DC and the mean images alone. Observe the non-eye images exhibit a strong dependency with the mean coefficient. This behavior is consistent with what we know about the statistics of natural images. In particular, assuming that the mean eye image is dominated by low-frequency information, the spatial structure in natural images that could not be explained by the DC and the mean eye image is very likely to possess information at higher frequencies. As shown in [10], there are strong correlations between wavelet amplitudes for the same image position across scale and orientation in a space-frequency decomposition. Thus a high value for the mean coefficient must give rise to a high value for the high-frequency components in the left-over signal.

4 Principal Subspace and its Complement

Once the images are rescaled into $A_k(\vec{x})$ for $k = 1, \dots, K$, we first trim the database by removing extreme images that either have a low or negative mean image coefficient and/or very large left-over variances. Most of these extreme images have identifiable problems, such as the eye being closed, not centered, having a different scale, or having reflections from glasses. For the database we consider here they constitute only a small fraction of the input data. We then use singular value decomposition (SVD) to perform principal component analysis. Let $B_j(\vec{x})$ and σ_j denote the basis images and singular values obtained from the SVD over the trimmed dataset (Fig. 3).

Suppose we approximate the normalized images $A_k(\vec{x})$ with just the first n PCA basis images. We compute the residual signal $e_k(\vec{x})$ in the complement space as, $e_k(\vec{x}) = [A_k(\vec{x}) - \sum_{j=1}^n c_{k,j} B_j(\vec{x})]$, where the expansion coefficient $c_{k,j} = \langle A_k(\vec{x}), B_j(\vec{x}) \rangle$ and n is taken to be much less than the total number of training images K . We can now define the residual variance to be $V_n(\vec{x}) \equiv \frac{1}{K} \sum_{k=1}^K e_k^2(\vec{x})$. It



Figure 1: Database of eye (top) and non-eye (bottom) images.

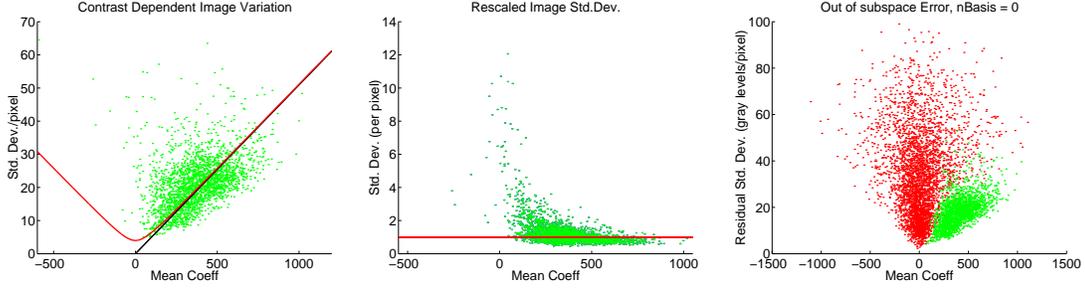


Figure 2: Characterizing the eye (green) and non-eye (red) spaces.

is easy to show that $V_n(\vec{x})$ satisfies the relation: $V_n(\vec{x}) = \sum_{j=n+1}^N (B_j(\vec{x})\sigma_j)^2$, where N is the total number of pixels. The residual variance is the estimate of the unexplained signal at each pixel because of the approximation of input with only n basis [5]. Although the residual variances at nearby pixels are likely to be correlated, we ignore this correlation.

5 Detection Strategy

In order to detect a test image as an eye, we expand it in terms of the DC and mean images, along with the first n PCA basis images. There are three statistics to consider for detection, namely, m_k , S_k^{wis} and S_k^{oos} . Here m_k is the coefficient of the mean image, S_k^{wis} is the within-subspace statistic given by $S_k^{\text{wis}} = \sum_{j=1}^n (c_{k,j}/\sigma_j)^2$ and $S_k^{\text{oos}} = \frac{1}{N} \sum_{\vec{x}} e_k^2(\vec{x})/V_n(\vec{x})$. Notice that S_k^{wis} measures the squared expansion coefficients $c_{k,j}$ compared to the squares of the singular values σ_j^2 , where σ_j^2 is also just the variance of $c_{k,j}$ across the training set. In addition S_k^{oos} measures the variance of what we call the out-of-subspace signal, which is the residual signal $e_k(\vec{x})$ normalized to have unit variance by division with $\sqrt{V_n(\vec{x})}$.

The variance plot drawn in Fig. 2(Right) shows $\sqrt{S_k^{\text{oos}}}$ computed with just the DC and the mean images and it is clear from the figure that the feature spaces of S_k^{oos} and m_k are amenable to a simple classification strategy. In particular, we use the constraint that $m_k > m_{\text{min}}$ where m_{min} is a small positive value. Negative values of m_k correspond to contrast reversals and small positive values for m_k generate a mean image component which varies only by a few gray levels. Additionally, we apply a contrast invariant threshold of the form $\arctan(m_k, \sqrt{S_k^{\text{oos}}}) \leq \tau_{\text{oos}}$, which requires that eyes be below a line drawn through the origin. The true detection and rejection rates are shown in Fig. 4 for three different choices for the total number of eigenbasis used, $n = 0, 5$, and 20. Using just the DC and the mean images ($n = 0$), this simple

strategy gives a detection rate of 94% and a false target rate of 6%. Also, observe the increase in separation of the eye-space from the non-eye space with the addition of eigenbases.

6 Mixture Model for Out-of-Subspace Signal

S_k^{oos} is a least-square measure and hence its computation is suspect when test images have pixel outliers. What we need is a robust method to compute the out-of-subspace variances. We achieve this by invoking a mixture model for the out-of-subspace signal, i.e. $\text{oos}_k(\vec{x}) = e_k(\vec{x})/\sqrt{V_n(\vec{x})}$. If there were no outliers, normalizing $e_k(\vec{x})$ by dividing it with $\sqrt{V_n(\vec{x})}$ will cause the out-of-subspace signal $\text{oos}_k(\vec{x})$ to have unit variance. Because pixel outliers destroy this property, we account for them with an explicit outlier model.

The mixture model contains an inlier component and a uniform outlier distribution. We choose a zero-mean, unit-variance Gaussian distribution $\mathcal{N}(\text{oos}_k(\vec{x}); 0, 1)$ as an inlier to explain the object-specific signal in the out-of-subspace. In particular, the probability for an out-of-subspace signal value of $\text{oos}_k(\vec{x})$ is taken to be independent of pixel position and is expressed as $p(\text{oos}_k(\vec{x})) = (1 - \pi_0) \times \mathcal{N}(\text{oos}_k(\vec{x}); 0, 1) + \pi_0 \times \mathcal{U}$, where π_0 and $1 - \pi_0$ are the mixing coefficients for the outlier distribution \mathcal{U} and the inlier distribution \mathcal{N} respectively. Now, the only unknown parameters of the mixture model are the mixing coefficients and the outlier probability.

Assuming a uniform distribution over the possible range of gray levels $[0, 255]$, the best case of an outlier that nearly looks like an eye image with a mean m_k can be shown to have a value $p_0(\text{oos}_k(\vec{x})) = m_k \sqrt{V_n(\vec{x})}/256$ for all $\text{oos}_k(\vec{x}) \in [-128/m_k \sqrt{V_n(\vec{x})}, 128/m_k \sqrt{V_n(\vec{x})}]$ and 0 otherwise. We treat out-of-space signal as independent of image position \vec{x} and hence, we set $\mathcal{U} = m_k [\sqrt{\sum_{\vec{x}} V_n(\vec{x})/N}]/256$.

We update the mixing coefficients in two steps. First, we begin with an initial guess for π_0 . Then for each pixel in



Figure 3: Eigenbasis for the eye space, arranged from left to right in a decreasing order of the input variance captured. The first sub-image on the left is the mean image.

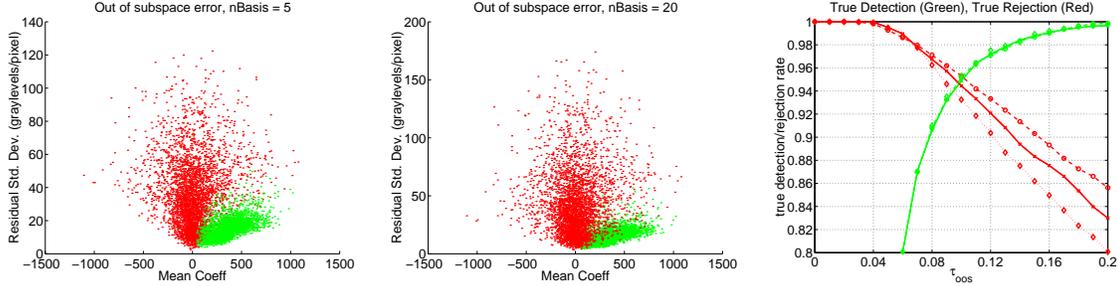


Figure 4: Separation of the eye and the non-eye clouds in the feature space of S_k^{oos} vs m_k with the addition of 5 (Left) and 20 (Middle) eigenbasis. The true detection (green) and rejection (red) rates are drawn on the right using lines annotated with diamonds ($n = 0$), crosses ($n = 5$) and circles ($n = 20$).

the out-of-subspace signal of a test image we compute an ownership probability given by: $\tau_k(oos_k(\vec{x})) = (1 - \pi_0) \times \mathcal{N}(oos_k(\vec{x}); 0, 1) / ((1 - \pi_0) \times \mathcal{N}(oos_k(\vec{x}); 0, 1) + \pi_0 \times \mathcal{U})$. This leads to a new value for π_0 given by: $\pi_0 = 1 - \frac{1}{K} \sum_{k=1}^K \tau_k(oos_k(\vec{x}))$. A high value for $\tau_k(oos_k(\vec{x}))$ implies that the out-of-subspace signal at pixel position \vec{x} is being accounted mostly by the inlier component of the mixture model.

In the second step, we use the ownership responsibilities to rederive the within-space coefficients. Imagine stacking the ownerships in a diagonal matrix W_k and performing a weighted least-squares computation such as: $\vec{d}_k = (U' * W_k * U)^{-1} U' * W_k' * T_k$, where matrix U has in its columns the DC vector, the mean image and zero or more eigenbases B_j , T_k is the test image and \vec{d}_k contains the within-space coefficients along with the components in the DC and the mean directions. The weighted least-squares is taking advantage of the ownership probability so that outlier pixels are discounted in estimating within-space coefficients.

Using the re-estimated within-space coefficients, we compute the out-of-subspace signal and then update the ownership probabilities one more time. These iterations are the usual steps of the expectation-maximization algorithm [3]. As shown in Fig. 5 the bright portions of the ownership maps clearly indicate which pixels in the out-of-subspace signal belong to the inlier component. The ownership maps can now be used to robustly estimate the out-of-subspace statistic. Any distribution other than a Gaussian is acceptable as long as it fits the out-of-subspace signal well. Finally, a simple detection strategy is to apply a threshold on the fraction of the total number of pixels in a test image the mixture model assigns to the inlier component.

References

[1] M. Black and A. Jepson EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation. *IJCV*, 26(1), pp. 63-84, 1998.

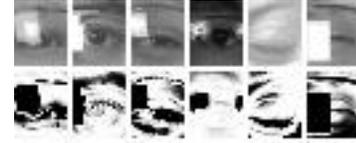


Figure 5: Ownership maps (Bottom) assigned by the mixture model to the Gaussian inlier component for images with outliers (Top).

[2] Phillips P. J, H. Wechsler, J. Huang and P. Rauss The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5), 295-306, 1998.

[3] McLachlan G.J. and K. E. Basford Mixture Models: Inference and Applications to Clustering. Marcel Dekker Inc. N.Y., 1988.

[4] Moghaddam, B. and A. Pentland Probabilistic Visual Learning for Object Representation. *PAMI*, 19(7), 696-710, July 1997.

[5] John Midgley. Probabilistic eigenspace object recognition in the presence of occlusions. Masters Thesis in Computer Science. University of Toronto 2001.

[6] Murase, H. and S. Nayar Visual Learning and Recognition of 3D Objects from Appearance. *Intl Journal of Computer Vision*, 14(1):5-24, 1995.

[7] S. Roweis EM Algorithms for PCA and SPCA. *Neural Information Processing Systems 10 (NIPS'97)* pp.626-632.

[8] Schmid, C. and R. Mohr. Local gray value invariants for image retrieval. *PAMI*, 19(5), 530-535, 1997.

[9] Sirovich, L and M. Kirby. Low-dimensional procedure for the characterization of human faces. *JOSA*, 4, 519-524, 1987.

[10] R W Buccigrossi and E P Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans Image Processing*, 8(12):1688-1701, Dec 1999.

[11] Fernando de la Torre and M. Black Robust PCA *ICCV*, Vol. I, pp. 362-369, 2001.

[12] M. Tipping and C. Bishop Probabilistic principal component analyzers. *Neural Computation*, 11(2):443-482, 1999.

[13] Turk, M. A. and A. P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.