Modelfest: Principal Component Analysis Reveals Underlying Channel Structure

Chien-Chung Chen
Christopher W. Tyler
Smith-Kettlewell Eye Research Institute
2318 Fillmore Street, San Francisco, CA 94115

ABSTRACT

In Year One of the Modelfest project, several laboratories collaborated to collect threshold data of human observers on 45 pattern stimuli. In this preliminary study, we used a principal component analysis (PCA) and a confirmatory factor analysis on the variations among observers to explore the underlying visual mechanisms for detecting Modelfest Stimuli. This analysis is based on the assumption that there are channels in common among observers that are represented with variations in sensitivity level only. We found three principal components. Assuming that each principal component represents a single mechanism, we compute the sensitivity profile of each mechanism as the sum of test stimuli weighted by the factor loadings on each component. The first mechanism is a spot detector. The second mechanism is dominated by a horizontal periodic pattern around 4 c/deg and the third may be characterized as a narrow bar detector.

Keywords: Modelfest, principal components, human visual system, spatial vision, receptive fields

1. Introduction

ModelFest[3,4] is a joint effort of several laboratories to construct a generally acceptable model for human vision. In the first year of ModelFest, efforts have been focused on collecting human observer data on the detection of luminance-defined spatial patterns[3]. One benefit of the Modelfest data set is that it included data for an extensive stimulus set on the same observers. Therefore, it is possible to make a meaningful comparison among visual system responses to a large number of images.

Here, we use principal component analysis (PCA) and confirmatory factor analysis to explore the mechanisms that might underlie the Modelfest data. Under the assumption of a set of linear visual channels with variations in sensitivity across observers, PCA should be able to reveal underlying structures of the visual system.

1.1 Principal components analysis

Suppose that the visual system consists of an array of linear mechanisms and that these linear mechanisms have the same form in all observers, but are expressed at different sensitivity levels (e.g., variations in receptive field density across observers). Let r_{ik} be the response of the k-th linear mechanism to the i-th pattern. For the purpose of our analysis, we assume that the responses of linear mechanisms are independent from each other. Thus, the response vector of k-th mechanism $R_k = [r_{1k}, r_{2k}, \ldots, r_{pk}]^T$ is orthogonal to the response vector of the i-th mechanism R_l where $k \neq i$ and p is the number of patterns (43 in this paper) in the data set. The measured detection threshold of the j-th observer to the i-th pattern y_{ij} is determined by a linear combination of the mechanism responses pluses an error term. That is, for the j-th observer,

$$Y_j - Y_m = e_{1j}R_1 + e_{2j}R_2 + \ldots + e_{qj}R_q + \underline{\varepsilon}_j \quad (1)$$

In *Human Vision and Electronic Imaging V*, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors, Proceedings of SPIE Vol. 3959 (2000) • 0277-786X/00/$15.00

where q is the number of independent mechanisms, Y_m is the mean threshold vector, $Y_j = [y_{1j}\ y_{2j}, \ldots, y_{pj}]^T$ is the threshold vector for observer j, e_{kj}, k=1,2,,...q are weights for mechanisms 1,2, ..., q for observer j respectively and $\underline{\varepsilon}_j = [\varepsilon_{1j}, \varepsilon_{2j}, \ldots \varepsilon_{pj}]^T$ is the error term. If we plot all the data in a nn dimensional space, the q linear mechanisms span a q-dimensional subspace. An error term is the deviation from each observer's data to this subspace. A good linear model should minimize the error terms and maximize the variation of the data explained by the q linear mechanisms. An orthogonal solution to this problem given by principal component analysis[7].

According to the method of singular value decomposition, for any real matrix A, there exist orthogonal matrices $\underline{Q}$ and $\underline{R}$ such that

$$A = \underline{Q}\ \Lambda\ \underline{R}^T \qquad (2)$$

where Λ is a p x p diagonal matrix with its diagonal been called the singular values of matrix A. The principal components can be computed from the covariance matrix with this method. Let $\underline{\Sigma}$ be the p x p covariance matrix of thresholds. For $\underline{\Sigma}$, $\underline{Q}$ and $\underline{R}$ are equal and singular values are the same as the Eigenvalues of Σ. Each column of $\underline{R}$, R_k, k = 1, 2, ...,p is a principal component of Σ. The proportion of variance explained by k principal components is equivalent to its corresponding singular values divided by the sum of all singular values[8].

Substituting A by S in Eq (2), we can get

$$\Sigma = \underline{R}\ \Lambda\ \underline{R}^T$$

which can be rearranged as

$$\begin{aligned}\Sigma &= \underline{R}\ \Lambda^{1/2}\ (\Lambda^{/1/2})^T\ \underline{R}^T \\ &= \underline{W}\ \underline{W}^T\end{aligned} \qquad (3)$$

where $\underline{W} = \underline{R}\ \Lambda^{1/2}$ and each element of $\Lambda^{1/2}$ is a square root of the corresponding element in Λ. The k-th column of matrix $\underline{W}$, $e_k R_k$, where e_k is the squared-root of the k-th singular value, defines the loading of all the images on the k-th principal component.

1.2 Confirmatory factor analysis

Once we extract the principal components from the covariance matrix, we can select the principal components (called common factors in the statistics literature) that account for the greatest amount of variation in the data and thus reduce the number of the parameters in the linear model. Suppose that we choose q factors (q < p) from the p principal components, on the assumption that q represents the number of significant mechanisms contributing to the observers' responses. We construct a p by q matrix $R_f = [R_1, R_2, \ldots, R_q]$, and q by q diagonal matrix with diagonal elements being $e_1, e_2, \ldots, e_f$. We can reconstruct the covariance matrix with the expression

$$\begin{aligned}\Sigma_f &= R_f\ \Lambda_f^{1/2}\ (\Lambda_f^{1/2})^T\ R_f^T \\ &= W_f\ W_f^T\end{aligned} \qquad (4)$$

where Σ_f is the estimated covariance matrix given the q factors we choose and the p x q matrix $W_f = R_f\ \Lambda_f^{1/2}$. The difference between Σ and Σ_f is the error of the model.

We may then perform an orthonormal linear transform represented by a q x q matrix V. Since the product of any orthonormal matrix and its transpose is an identity matrix, we have $(W_f\ V)\ (W_f\ V)^T = W_f\ V\ V^T\ W_f^T = W_f\ W_f^T$. That is, any orthonormal transform of W will give the same estimation of covariance matrix. One implication of this result is that we can rotate the common factors orthogonally by any angle and still have a linear model that retains the same estimation of the covariance matrix as the one estimated directly from the original principal components. Sometimes, it is informative to perform such a rotation with a particular criterion to have a better

understand of mechanisms contributing to the data. In this paper; we use a varimax criterion to maximize the variance on each chosen common factor.

2. Methods

The general experimental methods for ModelFest are presented elsewhere in the Proceedings[4]. The ModelFest Year One Project contains 45 pattern stimuli. By the time of press, five laboratories had submitted data on nine observers. We hope the number of the observers in the data set to increase over time as more researchers become involved in this project.

There were 45 stimuli in Modelfest data set. However, since not every observer submitted data on stimulus number 44 (dynamic random noise) and 45 (alternating phase), we based our analysis only on the data for the first 43 stimuli. A list of image can be found in the accompanying paper[4] or in the webpage http://www.neurometrics.com/projects/Modelfest/yearOneStimuli.htm

The principal components were computed with the singular value decomposition method on the covariance matrix. The confirmatory factor analysis was based on the principal components that each explained more than 10% of variation in the data. The factor rotation used a varimax criterion, which maximizes the variance of the data on each factor. The image of the sensitivity profile of a factor is computed as the sum of stimuli images weighted by the factor loading of that stimulus on the factor.

3. Results

3.1 General sensitivity

Figure 1 (a) plots the proportion of variance explained by each principal component. The data set is dominated by a single factor that accounts for 70% of the variance. The only other significant principal component accounts for 15% of the variance. Figure 1 (b) plots the loading of each image on the second principal component against that on the first. The first principal component has high positive loadings from all images. Apparently, it represents the difference in general sensitivities between observers. The only images with a lower loading on the first principal component are image 25 (the largest Gaussian spot), images 16 and 17 (two elongated Gabor patches with greatest length-to-width ratio), and images 9 and 10 (two Gabor patches with highest spatial frequencies in the stimuli set).

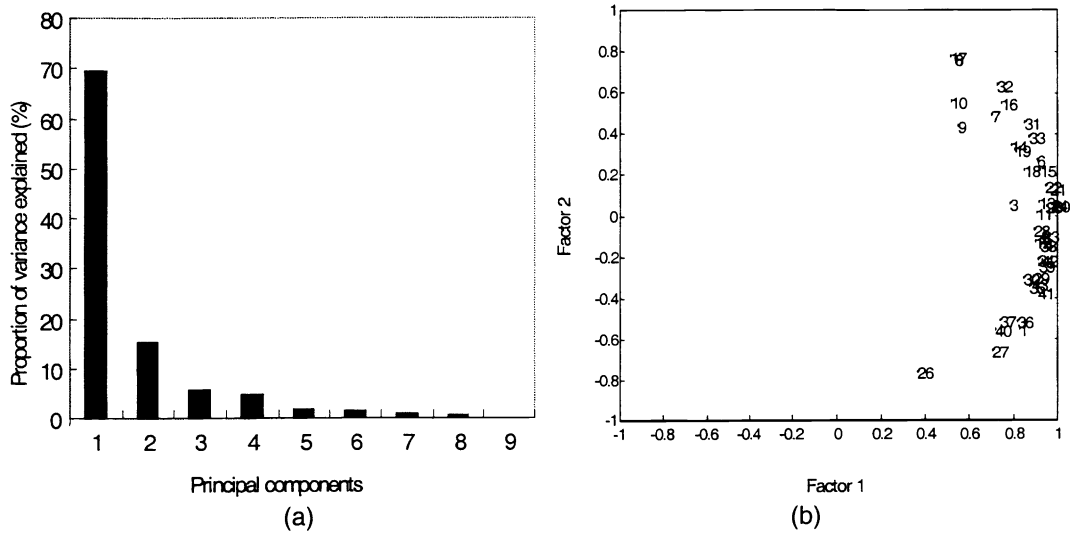


Figure 1. (a) Proportion of data variation explained by all principal components. (b) Scatter plot of factor loadings on common factor one and two.

3.2 Channel structures

Although the general sensitivity factor may be informative in understanding the visual system as a whole, it does not elucidate the underlying channel structures. In addition, a single dominating principal component may distort or obscure other properties in the data that might revealed by the principal component analysis[8]. Thus, we choose to scale each threshold by dividing threshold by the average of the corresponding observer's thresholds for all images. This scaling approach follows the hypothesis that all observers have the same channel structure but differ in the relative gain of those channels. That is, the threshold vector for j-th observer is

$$Y'_j - Y_m' = s_j (e_{1j}R_1 + e_{2j}R_2 + \ldots + e_{qj}R_q) + \underline{\varepsilon}_j' . \quad (5)$$

Dividing both sides of the equation by the general sensitivity sj, we get

$$Y_j - Y_m = (e_{1j}R_1 + e_{2j}R_2 + \ldots + e_{qj}R_q) + \underline{\varepsilon}_j$$

where $Y_j = Y'_j/s_j$, $\varepsilon_j = \varepsilon'_j/s_j$ and Y_m is the recalculated mean vector after normalization. Thus, reduce Eq (5) to the form of Eq (1).

After this transformation, we found that there were three principal components that each accounts for more than 10% of variation in the data, which were therefore chosen as the common factors in the model. Figure 2 plots the proportion of variation explained by each principal component. The three common factors account for more than 80% of variation in the data.

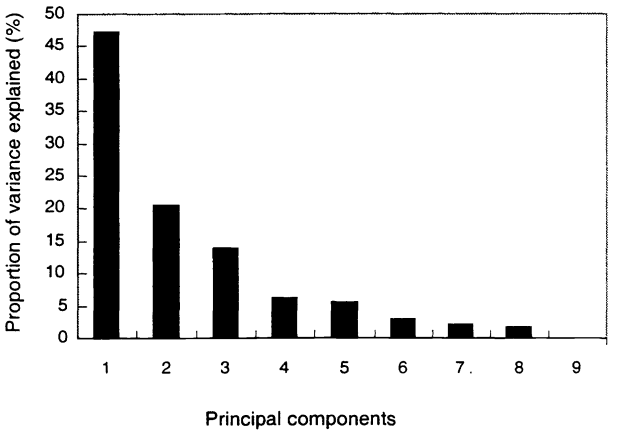
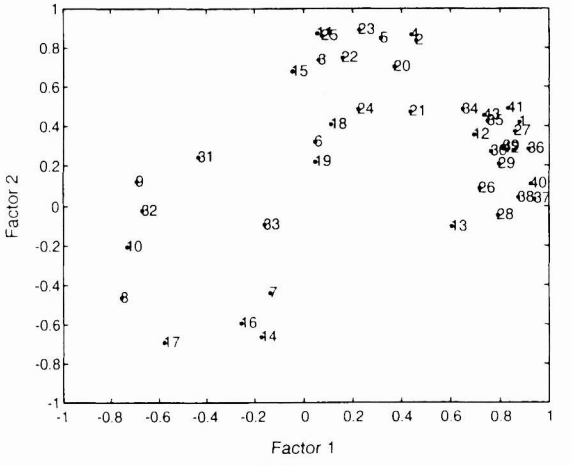


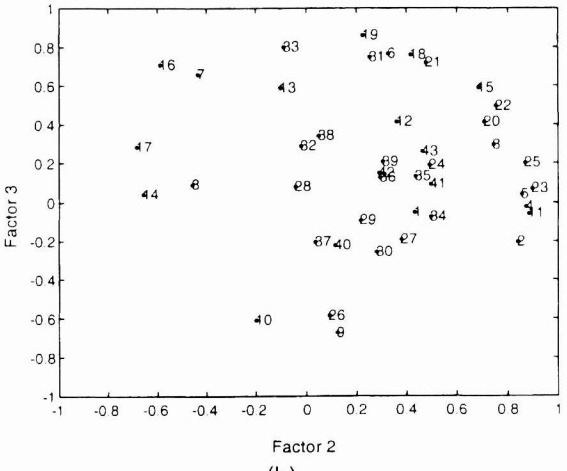
Figure 2. The proportion of variation of each principal component computed after data normalization.

Figure 3 contains scatter plots of factor loadings on all three common factors while Figure 4 contains images of the sensitivity profiles derived from factor loadings. Factor 1 has high positive loadings with the white noise pattern (image 35), the small Gaussian blobs (images 27, 28 & 29), the checker-board (image 42), the natural image (image 43), the oblique Gabor patches (images 36 & 37), the concentric pattern (image 41), the small disk (image 40), the plaids (images 38 & 39), and the alternating phased Gabor string (image 34); That is, this factor weights almost all the images in the stimulus set that are not horizontal periodic patterns. In addition, factor 1 has medium to high loadings for the horizontal Gabor patches with medium spatial frequencies (4 ~ 5.6 c/deg) and the one octave bandwidth (images 12 & 13). The images with medium to high negative loadings on factor 1 are the high spatial frequency Gabor patches (images 8, 9, 10, & 17), the dipole (image 32), and the line (image 31). The common features of those images are high spatial frequency (> 16 c/deg) components in the Fourier power spectrums. Those images with little or no loading on factor 1 tend to have an elongated shape or have a medium high spatial frequency. Therefore, we conclude that factor 1: (1) does not show a orientation tuning; (2) is excited by low-medium spatial frequencies and inhibited by high spatial frequencies; and (3) has a circular receptive field (because elongated stimuli do not produce a consistent result across observers and yield low factor loadings).

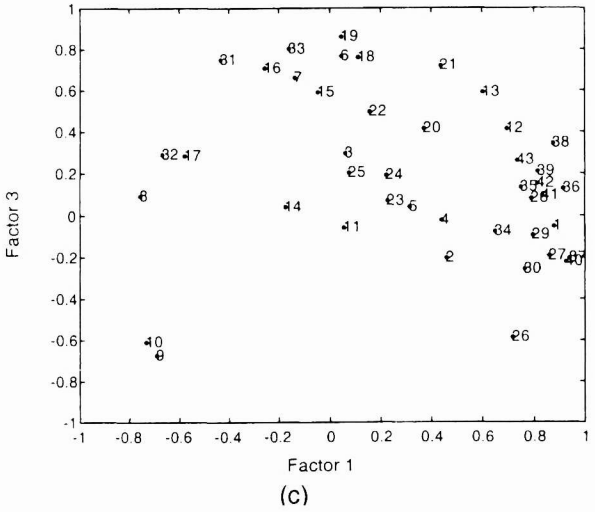
Factor 2 is apparently tuned to the horizontal orientation. In contrast to factor 1, which has high loadings for almost all images with non-horizontal components, factor 2 has virtually no loading on these images. The images with high positive loadings are all horizontal periodic patterns in low spatial frequencies (2 ~ 5.66 c/deg; images 2, 3, 4, 5, 11, 20,) or contain a low



(a)

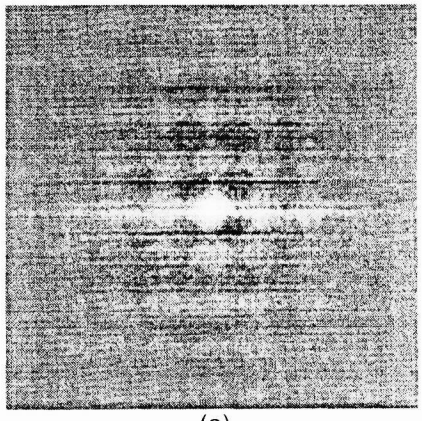


(b)

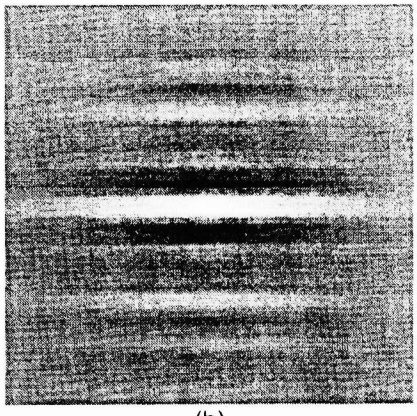


(c)

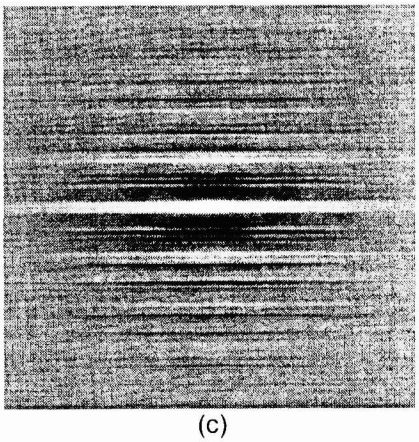
Figure 3. The scatter plots of factor loadings. (a) factor 2 vs factor 1; (b) factor 3 vs factor 2; (c) factor 3 vs factor 1.



(a)



(b)



(c)

Figure 4. The sensitivity profiles of three common factors. Panel (a) to (c) corresponding to factor 1 to 3 respectively.

spatial frequency components (images 22, 23 & 25). Those images that have high negative loadings on factor 2 (images 7, 8, 14, 16, & 17) are horizontal Gabor patterns of high spatial frequency (greater than 11.3 c/deg, except for image 16, which has an 8 c/deg spatial frequency). A Fourier transform of the sensitivity profile of factor 2 (figure 4 (b)) shows that the power spectrum peaks at 4 c/deg.

Factor 3 is sensitive to the shape of the image. Images with a high loading on this factor are Gabor patches elongated either along (images 15, 16, 18, & 19) or orthogonal to (images 20 & 21) their own orientation, a string of Gabor patches (image 33), a line (image 31), and other medium-high spatial frequency (8 ~ 11 c/deg) Gabor patches (images 6, 7, & 13). The images with negative loadings either have a very high spatial frequency (greater than 22 c/deg; images 9, 10) or a very low spatial frequency (image 26). Thus, the two properties of the factors are (1) tuning to a medium spatial frequency and (2) sensitivity to the elongation of the image.

4 Discussion

We show that there are three factors that account for most of the Modelfest data, other than the overall sensitivity variation among observers. The first factor has negligible orientation tuning, is sensitive to spots and disks, and is tuned low spatial frequencies. This factor has the properties of a "spot" detector. The second factor is sensitive to horizontal periodic patterns with spatial frequencies near 4 c/deg. It resembles a band-pass filter but seems to reflect squared-off bars near the center. The third factor is sensitive to the fine elongated patterns. It seems to be a narrow "bar" detector.

4.1 Spatial frequency tuning

The spatial frequency tuning of these three factors is rather interesting. In psychophysics, spatial frequency tuning has been measured by the threshold elevation of a target pattern induced by adapting[2] to or masking[9,10,11] by a series of patterns systematically varied in spatial frequency. The sensitivity of a target mechanism to a spatial frequency is defined as the inverse of the threshold elevation measured at that spatial frequency. A typical spatial frequency tuning function has an inverse-V shape; that is, the sensitivity peaks at a certain spatial frequency and decreases as spatial frequency moves away from the peak point. There is a controversy about the effects of spatial frequencies far away from the peak spatial frequency. Much literature has shown that spatial frequencies sufficiently distant from the peak should not produce a response in the target mechanism[11]. However, De Valois[6] argued that a distant spatial frequencies inhibit the target mechanism and thus produce a "Mexican hat" shape of tuning function (the sensitivity first decreases to a negative level and then recovers to zero sensitivity level as spatial frequency moves away from the peak frequency).

Suppose that distant frequencies could not produce a response in the target mechanism and that one of the common factors extracted from the Modelfest data reflects such mechanism. We would expect this factor to have high loadings for images near the peak frequency and low loadings for images far away from the peak frequency. However, all our three factors all have high negative loadings on spatial frequencies distant from those having high positive loadings. Thus, the results of our analysis support a mutual inhibition between, say, low and high spatial frequencies.

4.2 Independent component analysis is unsuitable for a linear model

Independent component analysis[5] (ICA) offers an alternative solution to the factor loading matrix $\underline{W}$ (eq. 3). It has been shown recently that ICA has several advantages over the principal component analysis in the investigation of natural images[1]. ICA can only be solved numerically by varying a guessed value for $\underline{W}$ with the learning rule

$$\Delta \underline{W} = c \, (\underline{I} + Z \, Y^T) \, \underline{W} \qquad (6)$$

where c is a constant, and Y is the threshold vector. The k-th element of Z,

$$z_k = \frac{\partial}{\partial y_k} \ln \frac{\partial g_k}{\partial y_k}$$

where g_k is a function of y_k. In most ICA literature, g_k is a nonlinear function. However, if g_k is a linear function, it follows that z_k is zero. Thus, in a linear model where g_k is linear for all k, the values of all elements of vector Z are zero. Eq (6) then reduces to

$$\Delta \underline{W} = c\, \underline{W}.$$

Thus, the learning rule can converge only if the initial guess is proportional to the final result; there is no way in ICA that guarantees a correct initial guess. It would have to rely on other methods such as PCA to supply such a guess and the final result would only be proportional to the accuracy of the guess. Therefore, it seems that no additional information can be gained from ICA in a linear model.

4.3 Limitations of our analysis

This paper is only a preliminary report. Our analysis is subject to limitations. The first one is the number of subjects in the data set. To obtain a stable solution from PCA, the number of observers should be equal to or greater than the number of variables. The Modelfest data set, though still growing, has not achieved this goal. We would not be surprised to find that the principal components vary as the number of observers in the data set increases. However, since our confirmatory factor analysis shows that three common factors can account for 80% of the data, we would not expect the change to be dramatic.

Our analysis is based the assumptions that the thresholds are determined by a linear system and that the common factors reflect the channels of the visual system. These two assumptions are subject to challenge. Our analysis would fail if these assumptions were not met. Even so, the analysis provides information on how the thresholds for the images in the stimulus set are related to each other. This information may provide hints for constructing a more elaborated model of the human visual system.

References

1. Bell, A. J. & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, **37**, 3327-3338.
2. Blakemore, C. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of the retinal images. *Journal of Physiology*, **203**, 237-260.
3. Carney, T., Klein, S. A., Tyler, C. W., Silverstein, A. D. Beutter, B., Levi, D. Watson, A. B., Reeves, A. J. Norcia, A. M., Chen, C. C., Makous, W. & Eckstein, M. P. (1999). Development of an image/threshold database for designing and testing human vision model. *SPIE Proceedings*, **3644**, 542-551.
4. Carney, T., Tyler, C. W., Watson, A. B., Makous, W., Beutter, B., Chen, C. C., & Norcia, A. M. (2000). Modelfest: year one results and plans for year two. **SPIE Proceedings**, **3959**-14.
5. Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, **36**, 287-314.
6. De Valois, K. K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research*, **17**, 1057-1065.
7. Jackson, J. E. (1991). *A Users's Guide to Principal Components*. NY: Wiley.
8. Johnson, R. A. & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall.
9. Legge, G. E. & Foley, J. M. (1980). Contrast masking of human vision. *Journal of the Optical Society of America*, **70**, 1458-1471.

10. Sachs, M. B., Nachmias, J. & Robson, J. G. (1971). Spatial frequency channels in human vision. *Journal of the Optical Society of America*, **61**, 1176-1178.
11. Wilson, H. R., McFarlane, D. K. & Philips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, **23**, 873-882.