# Perceptual image distortion*

Patrick C. Teo and David J. Heeger

Department of Computer Science and Department of Psychology
Stanford University, Stanford, CA 94309

## ABSTRACT

In this paper, we present a perceptual distortion measure that predicts image integrity far better than mean-squared error. This perceptual distortion measure is based on a model of human visual processing that fits empirical measurements of: (1) the response properties of neurons in the primary visual cortex, and (2) the psychophysics of spatial pattern detection. We also illustrate the usefulness of the model in measuring perceptual distortion in real images.

## 1 INTRODUCTION

A variety of imaging and image processing methods are based on measures of image integrity (distortion measures). Examples include: image data compression, dithering algorithms, flat-panel display and printer design. In each of these cases, the goal is to reproduce an image that looks as much as possible like the original.

It has long been accepted that distortion measures like mean-squared error (MSE) are inaccurate in predicting perceptual distortion. For example, in the context of image data compression, a number of methods have exploited the human visual system's insensitivity to higher spatial frequencies to achieve higher compression rates than schemes that simply used MSE as their distortion measure. Recently, some researchers have found that they were able to tolerate coarser quantization in areas of higher "texture energy" and still achieve the same perceptual results.[20,19] However, these techniques have often been rather ad hoc. A notable exception is recent work by Watson[25] that is based on psychophysical masking data.

In this paper, we present a perceptual distortion measure that predicts image integrity far better than MSE. Our distortion measure is a generalization of the model used by Watson,[25] and it encompasses the "texture energy" masking phenomenon mentioned above.[19,20]

Our perceptual distortion measure is based on fitting empirical measurements of: (1) the response properties of neurons in the primary visual cortex (also called visual area V1), and (2) the psychophysics of spatial pattern detection, that is, peoples' ability to detect a low contrast visual stimulus.

It is important to recognize the relevance of these empirical results for developing measures of image integrity. First, we discuss the relevance of the psychophysical data. In a typical spatial pattern detection experiment, the contrast of a visual stimulus (called the target) is adjusted until it is just barely detectable. In some experiments (called masking experiments), the target is also superimposed on a background (called the masker). Again, the contrast of the target is adjusted (while the masker

*Patrick C. Teo and David J. Heeger. Perceptual image distortion. *Human Vision, Visual Processing and Digital Display V, IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology,* 1994 (in press).

contrast is held fixed) until the target is just barely detectable. Typically, a target is harder to detect (i.e., a higher contrast is required) in the presence of a high contrast masker. A model that predicts spatial pattern detection is obviously useful in image processing applications. In the context of image compression, for example, the target takes the place of quantization error and the masker takes the place of the original image.

Next we discuss the relevance of the neurophysiological data. The primary visual cortex is a bottleneck in the primate visual pathway. Neural signals pass from the eyes to V1; from there, information fans out to a number of other cortical visual areas of the brain. Since V1 is a bottleneck, everything we see is mediated by neurons in that area of the brain. Two images will appear identical if they yield identical responses for all of the neurons in V1. Likewise, two images will appear very similar if they yield nearly identical responses in V1 neurons. A model that predicts the responses of V1 cells is, therefore, useful in developing a perceptual distortion measure.

An enormous amount of effort has been devoted over the past thirty years to research on spatial pattern psychophysics and on V1 physiology. A longstanding view is that the early stages of the visual system perform a linear transform: a frequency and orientation subband decomposition. This view has been supported by a variety of physiological[6,18] and psychophysical[11] results.

However, this linear transform model falls short of a complete account of early vision. One major fault with the linear model is the fact that V1 cell responses saturate at high contrasts (e.g., see Albrecht and Hamilton[3]). The responses of ideal linear operators, on the other hand, increase with stimulus contrast over the entire range of contrasts. A second fault with the linear model is revealed by testing superposition. A typical V1 neuron responds vigorously to its preferred orientation but not at all to the perpendicular orientation. According to the linear model, the response to the superimposed pair of stimuli (preferred plus perpendicular) should equal the response to the preferred stimulus presented alone. In fact, the response to the superimposed pair is about half that predicted (e.g., see Bonds[4]), a phenomenon known as cross-orientation inhibition. A third failure of the linear model is revealed by spatial masking psychophysics experiments.[9]

In recent years, we and others have developed a nonlinear model of early vision (actually, an extension of the linear transform model), hereafter referred to as the *normalization model*, to explain a significantly larger body of data[2,9,13–15] The normalization model has four stages. Here we introduce these four stages; each of the stages is described in detail later in the paper. The first stage of the model is a subband transform. In the second stage of the model, each coefficient of the subband transform is squared. The third stage of the model is a divisive normalization mechanism in which each squared coefficient is divided by the sum of a large number of squared coefficients. Each squared and normalized coefficient represents the response of a hypothetical V1 neuron. In physiological terms, normalization means that each V1 neuron is suppressed by the pooled activity of a large number of V1 neurons. The fourth and final stage of the normalization model is a detection mechanism.

The importance of normalization is that it preserves the essential features of linearity in a system (the brain) that has limited dynamic range. It is commonly believed that information about a visual stimulus, other than its contrast, is represented in terms of the relative responses of a collection of neurons. For example, the orientation of a stimulus might be represented as the ratio of the responses of two neurons, each with a different orientation preference. Indeed the ratio of a neuron's responses to two stimuli is largely independent of stimulus contrast.[3] Cortical neurons have a limited dynamic range

and their responses saturate at high contrasts. Normalization makes it possible for response ratios to be independent of stimulus contrast, even in the face of response saturation.

The normalization model explains response saturation of V1 neurons because the pooled activity (the sum of the squared coefficients) increases with stimulus contrast. The model explains cross-orientation inhibition because a given neuron is suppressed by many other neurons including those with perpendicular orientation tunings (see Heeger[13] for details).

The normalization model also explains spatial masking psychophysics. Foley and Boynton[9] recently performed an extensive series of spatial masking experiments. The normalization model provides a good fit to nearly all of his data (see Figures 3 and 6). By contrast, previous models of spatial masking (including a model that Foley himself proposed in the early 1980's) fail to explain his new data. According to the normalization model, spatial masking is a simple consequence of normalization (divisive suppression). The response of a neuron increases with target contrast, but it is suppressed by superimposing a masker. Hence, in the presence of a masker, the target must have a higher contrast to evoke a criterion response.

In particular, the model explains three general classes of spatial pattern detection results. First, it explains baseline contrast sensitivity (detection of a target when there is no masker). Second, the model explains the usual phenomena of contrast masking (when the target and masker have the same orientation). Third, and unlike previous models of spatial masking (like that used by Watson[25]), the normalization model explains the masking effect that occurs when the target and masker have very different orientations.

In Section 2, we derive the normalization model in detail and describe a perceptual distortion measure based on the model. In Section 3, we present the psychophysical methods used to collect spatial masking data. In Section 4, we demonstrate that the normalization model fits this psychophysical data. For examples of using the normalization model to fit neurophysiological data, see Heeger[13–15] and Albrecht and Geisler.[2] In Section 4, we also demonstrate that our perceptual distortion measure predicts image integrity far better than MSE.

## 2 THE MODEL

The model consists of four steps: (1) a front-end linear transform consisting of a suite of linear sensors tuned to different spatial orientations and frequencies, (2) squaring, (3) a divisive contrast normalization mechanism, and finally (4) a detection mechanism. Hence, the input images undergo both a linear and a non-linear transformation. The perceptual distortion of an image with respect to a reference is computed by passing both images through the model. The final detection step determines the extent to which the distortion is visible.

**Linear Transform.** In many signal processing applications, a signal is decomposed into a set of subbands, and the information within each subband is processed more or less independently of that in the other subbands. Recently it has become popular to use discrete subband decompositions that are critically sampled, i.e., in which the number of transform coefficients is equal to the number of samples in the original signal. Quadrature mirror filters (QMF's) and wavelets, which are closely related, are examples of critically sampled transforms that have proven to be useful in various signal processing

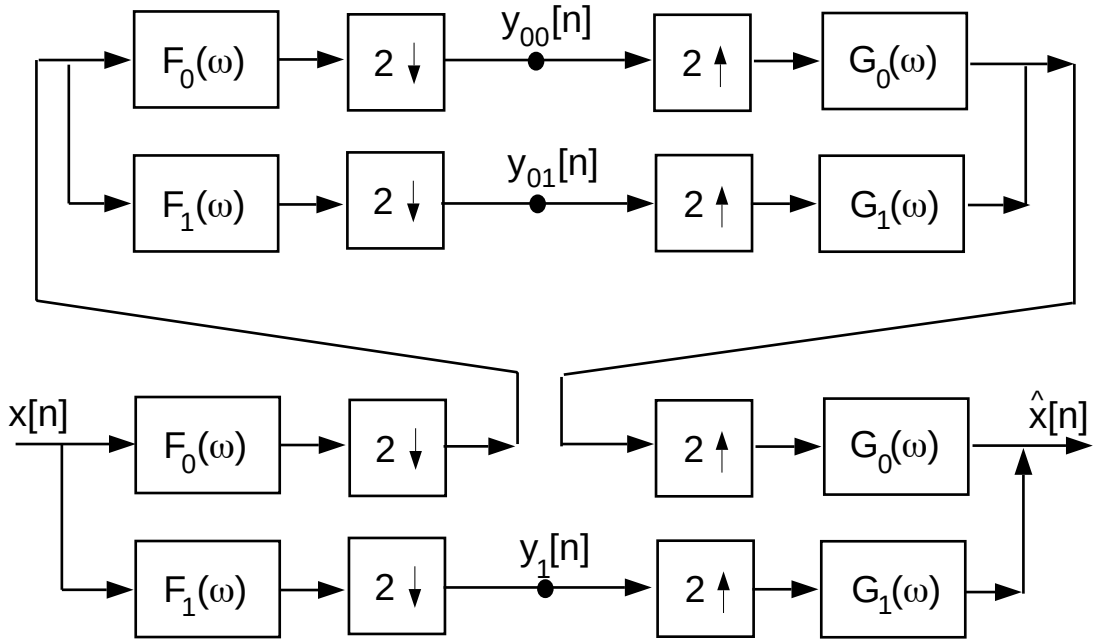


Figure 1: Example of a QMF pyramid analysis/synthesis cascade (reproduced from Simoncelli and Adelson[21]). An input signal $x[n]$ is convolved with two shift-invariant linear filters. The frequency responses of these filters are denoted by $F_1(\omega)$ and $F_2(\omega)$. The filtered signals are each subsampled by a factor of two. Then one of these filtered and subsampled signals is passed through the same set of operations (filtering and subsampling) again. The coefficients of the transform are denoted by $y_1[n]$, $y_{01}[n]$, and $y_{11}[n]$. The original signal may be reconstructed (if so desired) from the transform coefficients by inverting the operations.

applications.

For the linear transform front-end of our model, we use the hexagonally sampled quadrature mirror filter transform (hex-QMF) designed by Simoncelli and Adelson.[21] Quadrature mirror filters, originally introduced by Crosier et al.,[7,8] are used in an analysis/synthesis system that decomposes a signal into high-pass and low-pass frequency subbands. Vetterli was the first to propose the use of QMF's for image decomposition.[24] Mallat[16] related QMF transforms to wavelet theory and proposed their use in machine vision.

The hex-QMF has a number of desirable properties. First, the hex-QMF transform is a pyramid transform,[5] and hence it can be computed efficiently as a cascade of filtering and subsampling operations. An example of this recursive filtering and subsampling algorithm is diagramed in Figure 1.

A second desirable property of the hex-QMF is that the basis functions of the transform are jointly localized in space, spatial frequency, and orientation. The concept of joint localization in space and spatial frequency was introduced by Gabor.[10] Most applications of wavelets/QMF's to two or more dimensions have used separable filters; some of the basis functions of these separable transforms are not localized in orientation (they are non-oriented). For our perceptual distortion measure, it is important that all of the transform's basis functions be oriented. Figure 2 shows a diagram of the partition of the

spatial frequency plane created by the hex-QMF pyramid.

A third desirable property of the hex-QMF transform is self-similarity; the basis functions of the transform are translations, dilations, and rotations of a common kernel. This property of self-similarity is the defining property of wavelets (see Strang[23] for an introduction). A consequence of self-similarity is that the bandwidths of the basis functions are equal on a logarithmic frequency scale (in particular, the hex-QMF basis functions have octave bandwidths). Many authors in signal processing, computer vision, and biological vision have argued for the importance of such equal logarithmic-width (constant Q) transforms.[1,12,16,17,26]

A fourth desirable property of the hex-QMF transform is that the set of basis functions collectively tiles all possible orientations and frequencies.

Unfortunately, the hex-QMF also has some undesirable properties. One major drawback of using any transform with spatially localized, orthogonal basis functions (wavelet, QMF, blocked DCT) is the lack of translation invariance.[22] A second problem with the hex-QMF, for our purpose, is that the orientation bandwidth of the filters is too broad. The frequency and orientation bandwidths are particularly important as we shall see when we fit the model to psychophysical data (see *Section 4* below). For these two reasons, we plan to develop a better perceptual distortion measure using an overcomplete, steerable pyramid.[22] The steerable pyramid transform retains the desirable properties of the hex-QMF (for our application), but alleviates the undesirable properties.

**Squaring and Normalization.** The front-end linear transform yields a set of coefficient values. Each squared and normalized coefficient represents the response of a visual sensor (a hypothetical neuron in primary visual cortex).

Since the front-end transform is linear, a coefficient's magnitude increases linearly with the contrast of the input image. Furthermore, these linear coefficients are equally sensitive (or insensitive) to perturbations of the input regardless of image contrast. Contrast normalization results in a nonlinear relationship between input contrast and sensor output. It also assures a moderate level of sensitivity to inputs of smaller contrast.

The normalization scheme is divisive and is determined by two parameters: an overall scaling constant, $k$, and a saturation constant, $\sigma^2$. Let $A^\theta$ be a coefficient of the front-end linear transform. The squared and normalized output, $R^\theta$, is computed as follows:

$$R^\theta = k \, \frac{(A^\theta)^2}{\Sigma_\phi \, (A^\phi)^2 + \sigma^2} \tag{1}$$

In words, the normalized output of a sensor tuned to orientation $\theta$ is computed by dividing its original squared response, $(A^\theta)^2$, by the sum of the squared responses of a pool of sensors over all orientations. This summation, $\Sigma_\phi \, (A^\phi)^2$, *includes* the term, $A^\theta$, that appears in the numerator (i.e., each sensor suppresses itself). As long as $\sigma$ is nonzero, the normalized sensor response will always be a value between 0 and $k$, saturating for high contrasts.

Contrast normalization has a simple geometrical interpretation. It can be viewed as an embedding of the vector of transform coefficients into a space one dimension higher; the value of the extra di-

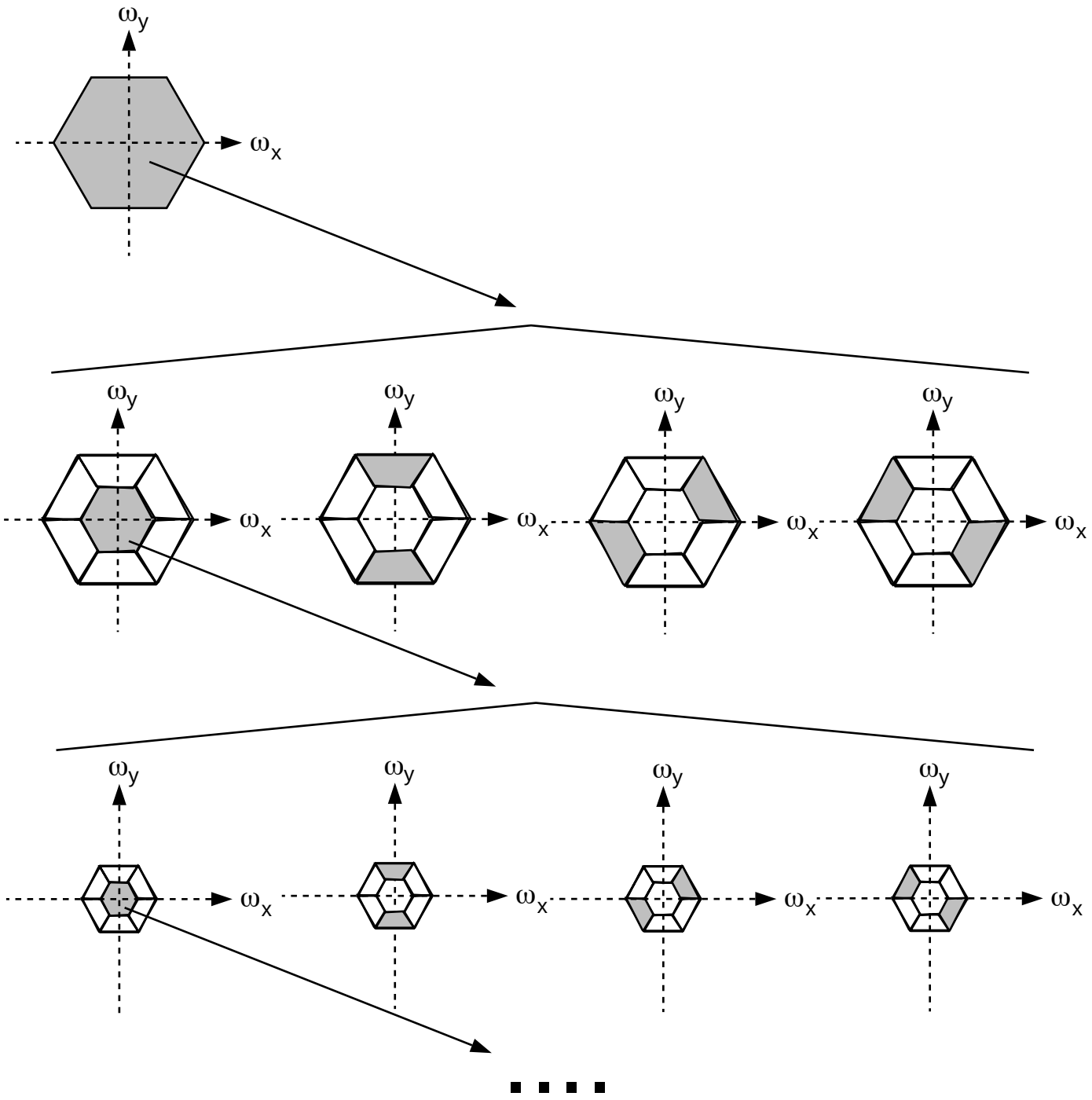


Figure 2: Diagram (reproduced from Simoncelli and Adelson[21]) of the partition of the spatial frequency plane created by the hex-QMF pyramid. The recursive filtering and subsampling algorithm gives rise to equal bandwidths on a logarithmic frequency scale. At each level, the bands are tuned to orientations of 0, 60, and 120 degrees with respect to the vertical.

mension's coordinate being set to $\sigma$. For example, since the hex-QMF transform has three orientation bands, a vector of linear transform coefficients is $(A^0, A^{60}, A^{120})$ and the higher dimensional vector is $(A^0, A^{60}, A^{120}, \sigma)$. This vector is then normalized (in the usual sense of vector normalization) to obtain a unit vector (in the higher dimensional space) which is then projected back onto the original lower dimensional space. Lastly, each coordinate of the resultant vector is squared.

The normalized sensors, like real neurons, each has a limited dynamic range, as shown in Figure 4(a). Each sensor is able to discriminate contrast differences only over a narrow range of contrasts. This range is determined by the scaling and saturation constants, $k$ and $\sigma^2$, respectively. Hence, several contrast normalization mechanisms, each having different $k_i$'s and $\sigma_i^2$'s, are required to discriminate contrast changes over the full range of contrasts. In the current implementation of the model, we have four different contrast discrimination bands (that is, four different choices of $\sigma_i^2$ and $k_i$).

The full set of normalized sensors is tuned for different spatial positions, spatial frequencies, orientations, and contrast discrimination bands. The outputs of these normalized sensors are then used by the detection mechanism to determine the level of perceptual distortion present between a pair of images.

**Detection Mechanism.** Let $\mathcal{R}_0$ be a vector of normalized sensor responses for an input image $I_0$. Let $\mathcal{R}_1$ be the corresponding vector of normalized responses for input $I_1$. The detection mechanism adopted by the model is the simple squared-error norm (i.e., the vector distance between $\mathcal{R}_0$ and $\mathcal{R}_1$):

$$\triangle\mathcal{R} = ||\mathcal{R}_0 - \mathcal{R}_1||_2^2 \tag{2}$$

The threshold at which distortion is visible is arbitrarily set at unity in the model. Hence, $\triangle\mathcal{R}$ in equation (2) is 1 at threshold.

One might include all of the normalized sensor responses (all spatial positions, spatial frequencies, orientations, and contrast discrimination bands) in the vectors, $\mathcal{R}_0$ and $\mathcal{R}_1$, and compute a single number representing the overall detectability of differences between the two images. We find it more informative, however, to implement the detection mechanism independently for each local patch (or block) of the images. For each block, we compute the vector difference between $\mathcal{R}_0^j$ and $\mathcal{R}_1^j$, where the superscript $j$ indexes over all the blocks, and where each of the $\mathcal{R}$ vectors includes the normalized sensor responses from that block.

## 3 METHODS

The empirical data used to fit the model were obtained from contrast masking experiments conducted by Foley and Boynton.[9] The task in the experiments was to detect a target pattern superimposed on a masker pattern. The maskers were 2 cycle per degree (cpd) sinusoidal gratings (5 degrees high by 7 degrees wide) of several orientations (0, 11.25, 22.5, 45 and 90 degrees re vertical). The target was a vertically oriented 2 cpd Gabor patch with vertical and horizontal 1/e halfwidths of 0.5 degrees. Background luminance was 32 cd/m$^2$. The target and masker were presented simultaneously as 33 ms pulses and were viewed at a distance of 162 cm. A two-alternative temporal forced-choice paradigm with an adaptive threshold-seeking algorithm was used to measure target threshold contrast as a function of masker contrast: the TvC (threshold vs. contrast) curve.

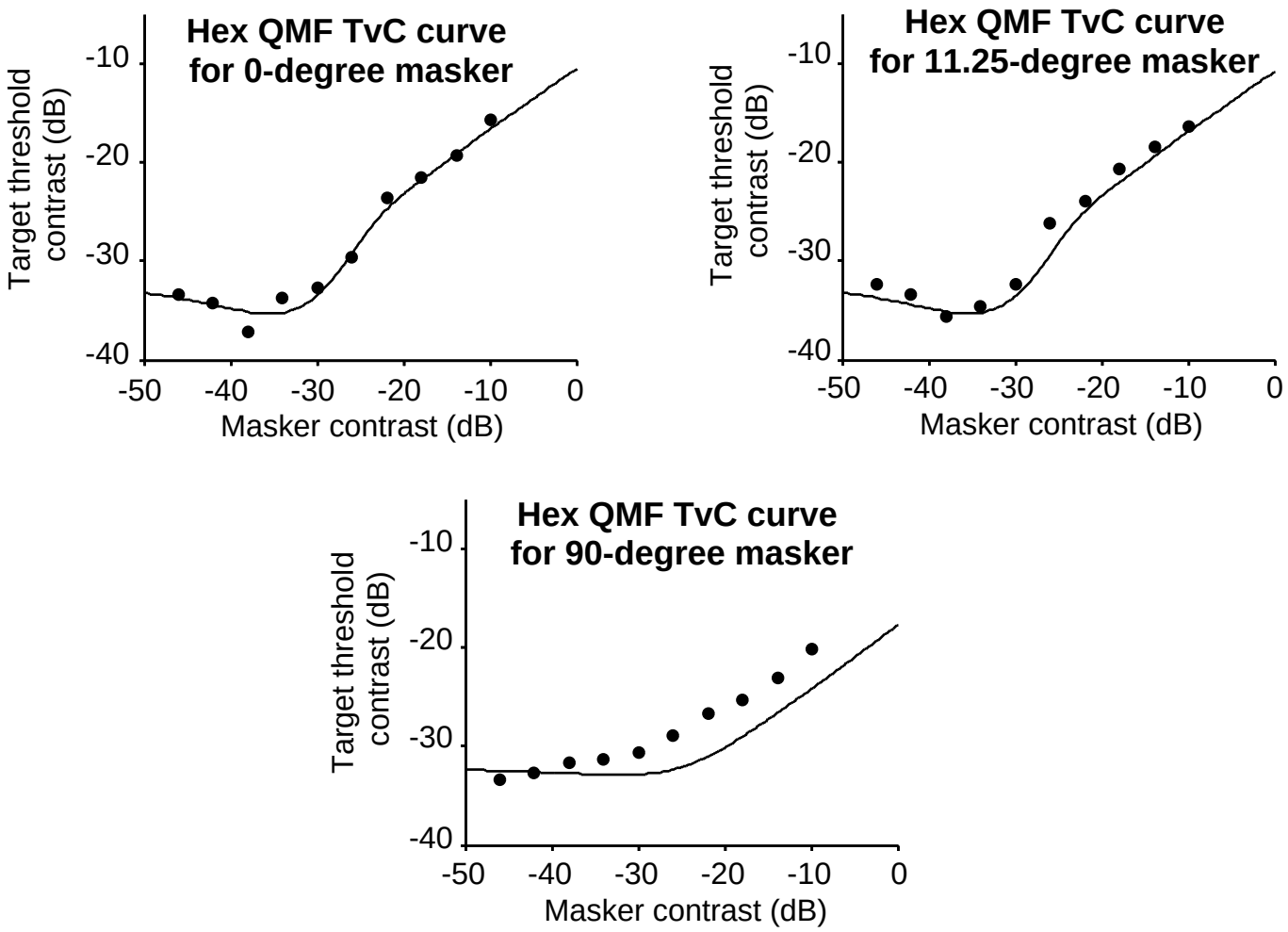


Figure 3: Result of fitting the model to empirical data. Empirical data are denoted by filled circles. Solid curves denote predicted target threshold contrasts.

## 4 RESULTS

**Hexagonal Quadrature Mirror Filters.** The free parameters of the model are the pairs of scaling and saturation constants ($k_i$ and $\sigma_i$). We found experimentally that four was the minimum number of pairs required to fit the data. Since the hex-QMF's have three orientation bands per level, there are a total of twelve normalized sensor responses for each spatial position and spatial frequency band. The task was to pick the $k_i$'s and $\sigma_i$'s to predict the empirical TvC curves for each of the various masker orientations (namely, at 0, 11.25, 22.5, 45 and 90 degrees re vertical).

Figure 3 shows the result of fitting the model to empirical TvC data at masker orientations of 0, 11.25 and 90 degrees. The fit to 0-degree masker orientation TvC data (0-degree TvC data) is extremely good. The overall goodness of fit indicates that four contrast tuning mechanisms are sufficient to reproduce the properties described by the TvC data.

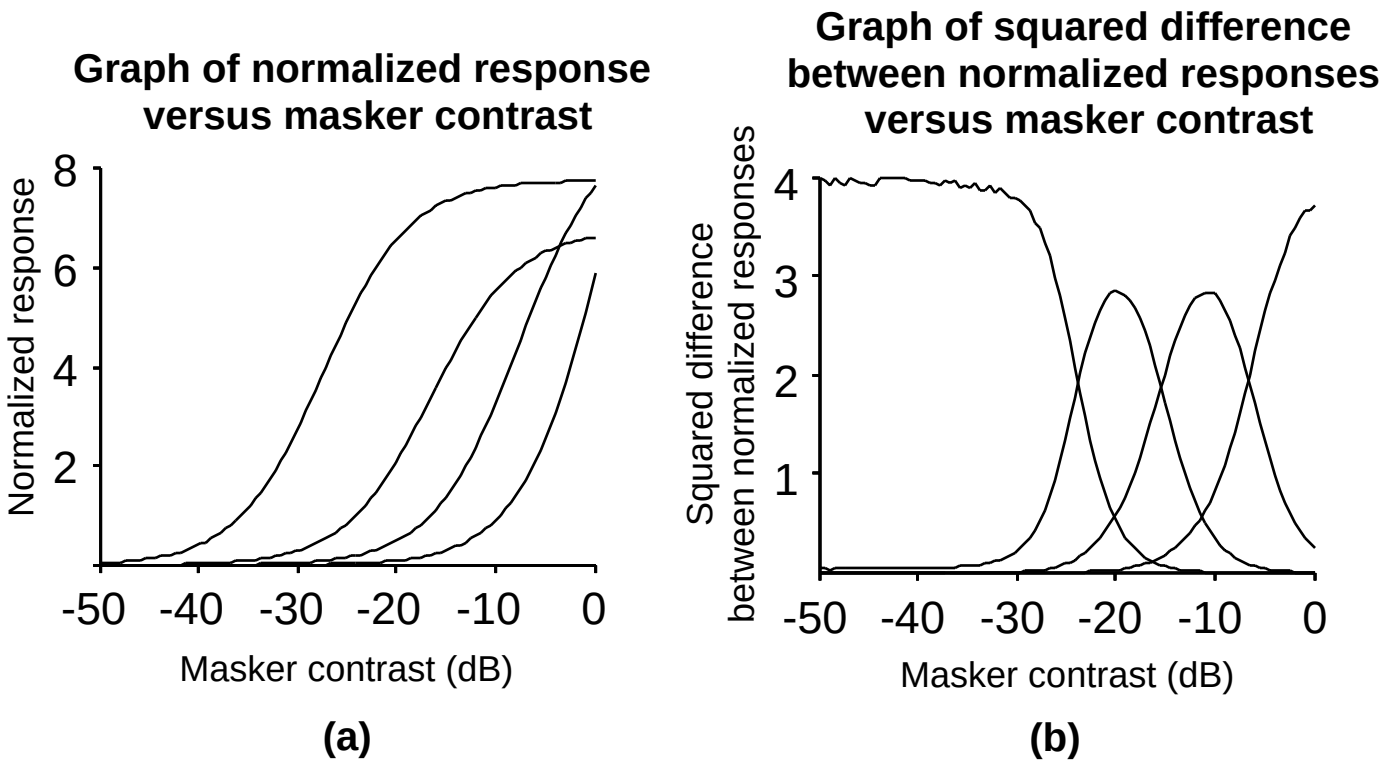


Figure 4: (a) Response of the four normalized sensors as a function of log contrast for a sinusoidal grating image. The responses saturate for high contrasts. The dynamic range (the range of contrasts for which this sensor can discriminate contrast effectively) of each sensor is limited to the rising portion of each curve. (b) Squared-error difference between responses to masker-only and masker with target (at threshold contrast) for each of the four contrast discrimination bands.

Figure 4(b) shows the squared difference between the responses to masker-only and masker with target (at threshold contrast) for each of the four contrast discrimination bands. These graphs sum to 4.0 since we fitted the parameters to predict target threshold contrast at a mean squared difference of 1.0. The graphs also show that each contrast discrimination band is differentially responsive only within a certain range of masker contrasts.

One important characteristic of the TvC data is the presence (or absence) of a "dipper". The presence of a dipper indicates that within that range of masker contrasts, the masker facilitates the detection of the target. A pronounced dipper can be observed in the 0-degree and 11.25-degree TvC data indicating that facilitation occurs at low contrasts for similarly oriented stimuli. The dipper is almost absent in the 22.5-degree TvC data and completely absent in TvC data involving greater orientation differences. The best fitting choice of parameters could generally be determined by simply fitting to the 0-degree TvC data alone. The shapes of the other TvC curves are largely determined by the orientation bandwidth of the front-end filters. This simplifies the task of fitting the model to the data.

While the fits to the 0-degree and 11.25-degree TvC data are impressive, the fits to the other curves are not as good. Although the shapes of the predicted TvC curves are similar to the empirical data,

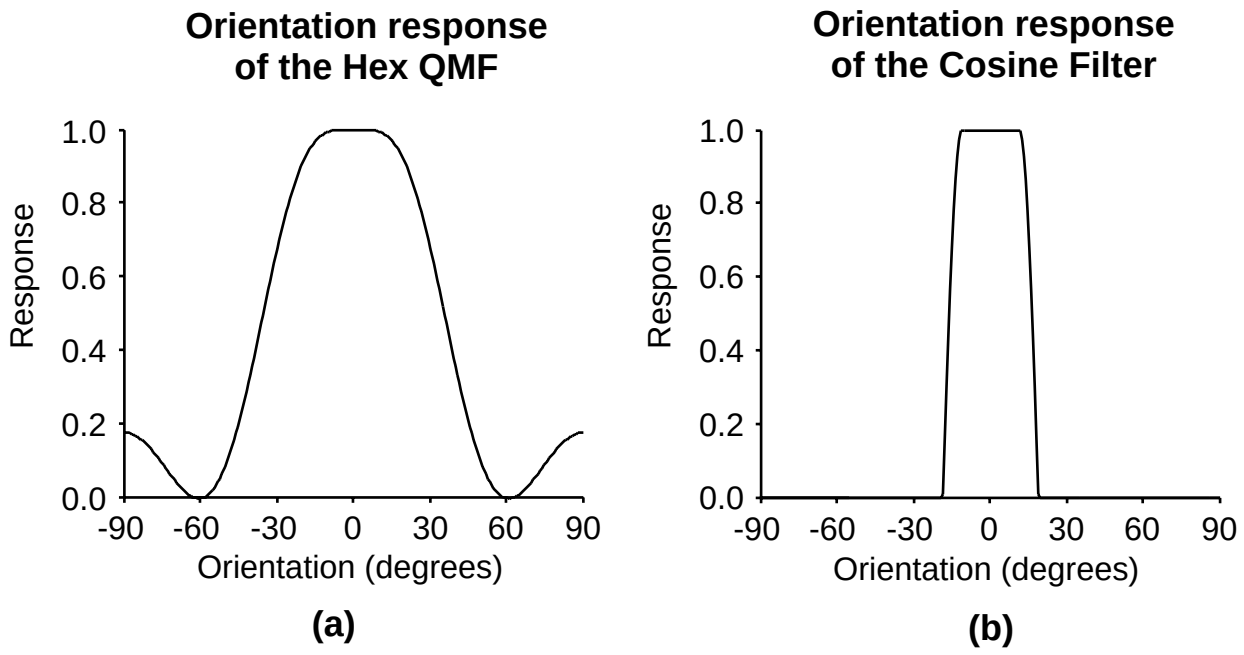


Figure 5: (a) Orientation tuning of the hex-QMF. (b) Orientation tuning of the synthetic Cosine filter.

they are shifted by some amount. We believe that this is caused by the relatively broad orientation bandwidth of the hex-QMF's. Figure 5(a) shows a graph of the orientation selectivity of the vertical hex-QMF. The orientation selectivity of the filter is rather poor; the 3dB bandwidth is about 58 degrees, i.e., 29 degrees re vertical in each direction. Furthermore, there is a non-zero response to maskers at approximately 90 degrees re vertical. We hypothesize that a better fit can be obtained by using filters with narrower orientation bandwidths.

**Synthetic Cosine Filter.** In order to verify this hypothesis, we designed a synthetic orientation tuning curve with a smaller and sharper bandwidth that does not have the anomaly at 90 degrees re vertical. Equation 3 gives the response of a vertically tuned filter to sinusoidal gratings of various orientations.

$$f(\theta) = \begin{cases} 1.0 & \text{if } |\theta| \leq \omega \\ \cos \frac{\pi}{2} \frac{|\theta| - \omega}{\nu} & \text{if } \omega < |\theta| \leq \omega + \nu \\ 0.0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\omega$ is half the width of the pass band and $\nu$ is the width of the transition band beyond the pass band. Because this filter is more narrowly tuned, we require a suite of six filters, tuned to 0, 30, 60, 90, 120 and 150 degrees re vertical, to tile all orientations.

Figure 5(b) shows the orientation tuning of our synthetic cosine filter. With this filter, the fits to all the curves are much better as can be seen in Figure 6.

As before, we found the best fitting choice of normalization parameters simply by fitting to the 0-degree TvC data alone. The other TvC data were predicted from the model based on those parameters.

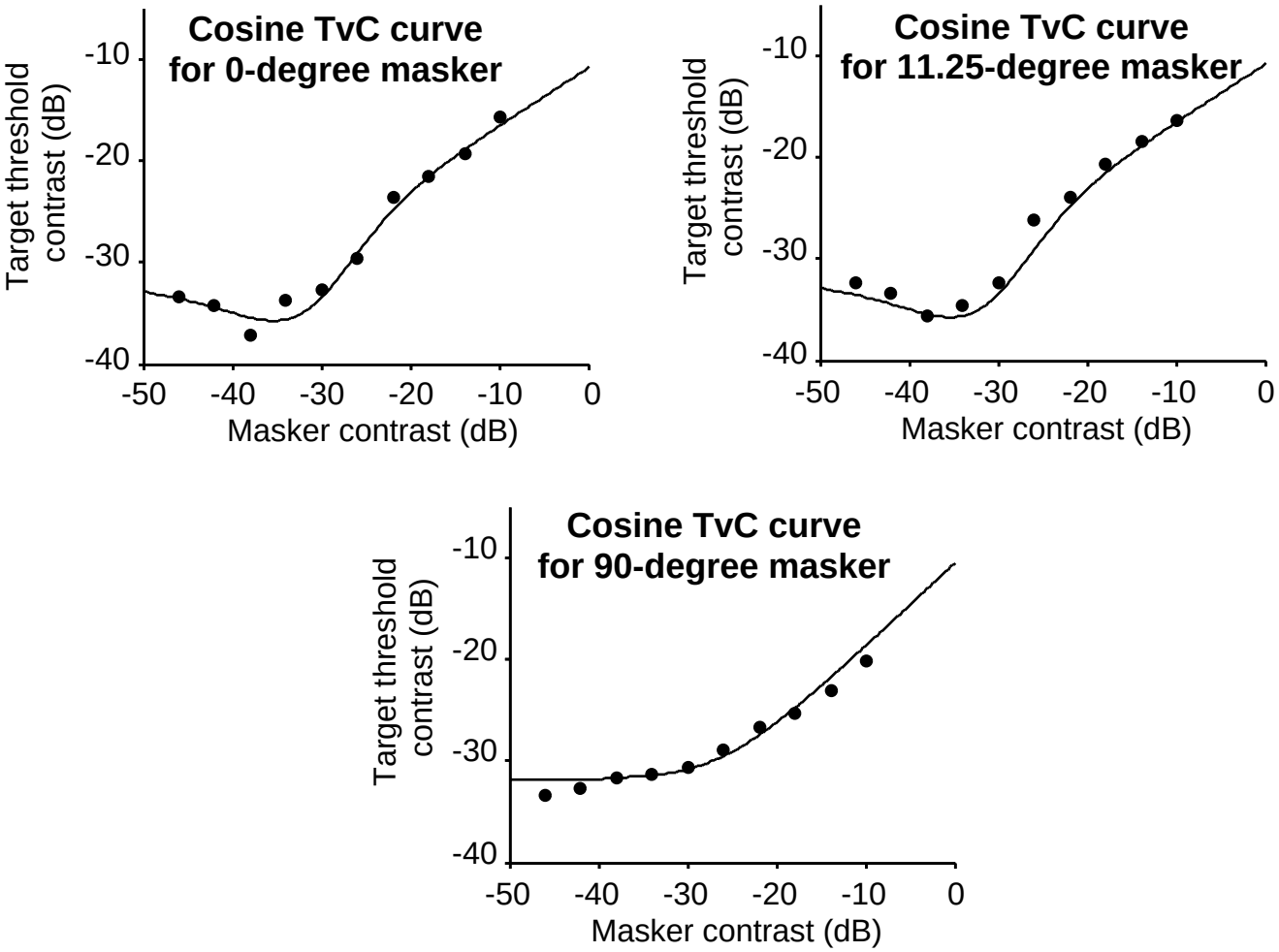


Figure 6: Result of fitting the data using the synthetic cosine filters as the front-end linear transform. Empirical data are denoted by filled circles. Solid curves denote predicted target threshold contrasts.

This supports our previous claim that the shapes of these other TvC curves are governed largely by the orientation tuning of the front-end linear filters.

**Demonstration** In order to assess the model, we added bandpass distortion to a reference image and then computed the perceptual distortion between the original and distorted images. In particular, we distorted the second level of the hex-QMF pyramid in two ways: first to maximize perceptual distortion, and second to minimize it. Figure 7 shows the original Einstein image along with the two distorted images. The perceived distortion is very different, yet the distortion was added so that standard distortion measures (mean squared error and peak signal-to-noise ratio) were very nearly the same. These standard distortion measures are, therefore, poor predictors of the perceived distortion.

Figure 7 also shows the perceptual distortion images for the minimally and maximally distorted Einstein images. Darker regions correspond to areas of lower perceptual distortion while brighter regions indicate areas of greater perceptual distortion.

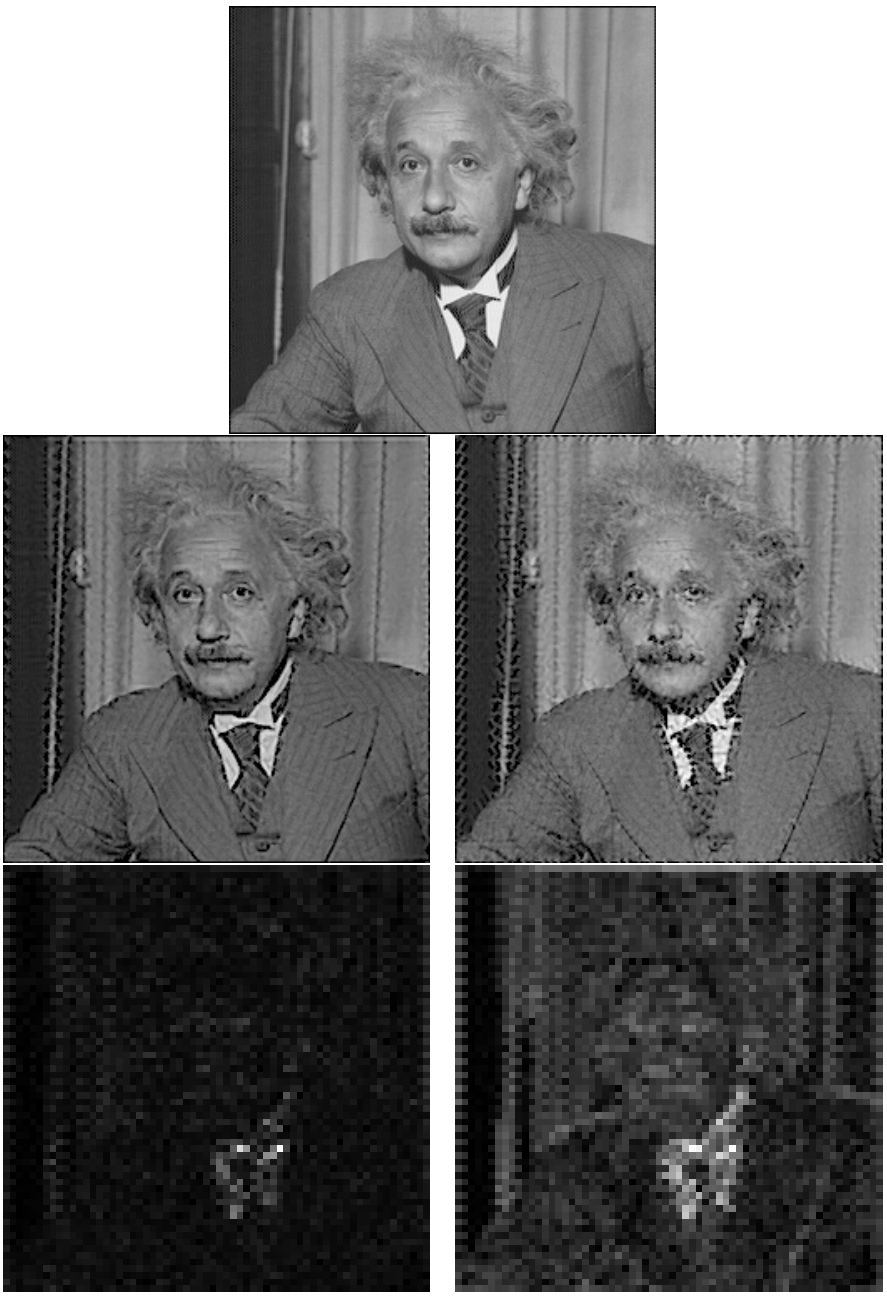

Figure 7: (Top) Original Einstein image. (Middle-left) Image was distorted so as to minimize perceptual distortion (RMSE = 14.5, peak-SNR = 24.9 dB). (Middle-right) Image was distorted so as to maximize perceptual distortion (RMSE = 15.5, peak-SNR = 24.3 dB). Both distorted images have nearly identical mean-squared error and peak-SNR. The overall perceptual-distortion-measures for the left and right images are 0.69 and 1.78 respectively. (Bottom-left) Perceptual distortion measured from the minimally distorted image. Darker regions correspond to areas of lower perceptual distortion while brighter regions indicate areas of greater perceptual distortion. (Bottom-right) Perceptual distortion measured from the maximally distorted image.

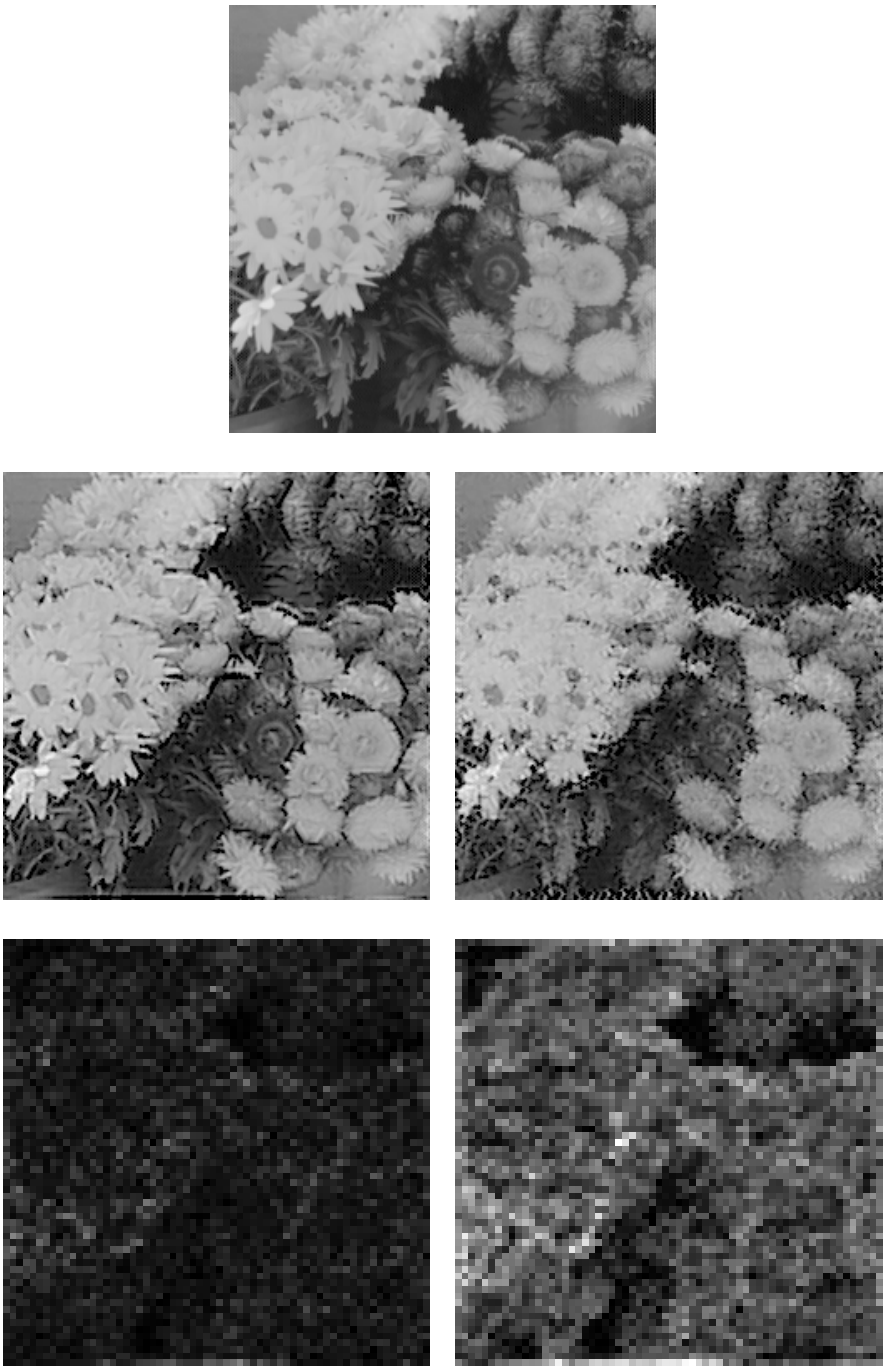

Figure 8: (Top) Original Flowers image. (Middle-left) Image was distorted so as to minimize perceptual distortion (RMSE = 17.4, peak-SNR = 23.4 dB). (Middle-right) Image was distorted so as to maximize perceptual distortion (RMSE = 17.2, peak-SNR = 22.3 dB). Both distorted images have nearly identical mean-squared error and peak-SNR. The overall perceptual-distortion-measures for the left and right images are 0.77 and 2.11 respectively. (Bottom-row) Perceptual distortion measured from the minimally and maximally distorted images.

Figure 8 shows another example of minimally and maximally distorted images with nearly identical MSE.

## 5 CONCLUSION

We have described a model of perceptual distortion that is consistent with empirical findings in the physiology of V1 cells and in spatial pattern psychophysics. In particular, we have shown that the model explains orientation and contrast masking. We plan on using this model further to explain other psychophysical data (e.g. summation experiments) and to explore the advantages of using a better front-end linear transform like the steerable pyramid. In addition to investigating spatial pattern detection phenomena, we also intend to demonstrate the usefulness of the model in more practical applications like image data compression and other image processing tasks.

## 6 ACKNOWLEDGEMENTS

This research was supported by NIMH grant 1-R29-MH50228-01 and by NASA grant NCC2-307.

## 7 REFERENCES

[1] E H Adelson, C H Anderson, J R Bergen, P J Burt, and J M Ogden. Pyramid methods in image processing. *RCA Engineer*, 29:33–41, 1984.

[2] D G Albrecht and W S Geisler. Motion sensitivity and the contrast-response funtion of simple cells in the visual cortex. *Visual Neuroscience*, 7:531–546, 1991.

[3] D G Albrecht and D B Hamilton. Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, 48:217–237, 1982.

[4] A B Bonds. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.

[5] P Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16:20–51, 1981b.

[6] F W Campbell, G F Cooper, and C Enroth-Cugell. The spatial selectivity of visual cells of the cat. *Journal of Physiology (London)*, 203:223–235, 1969.

[7] A Crosier, D Esteban, and C Galand. Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *International Conference on Information Sciences and Systems*, pages 443–446, Patras, 1976.

[8] D Esteban and C Galand. Application of quadrature mirror filters to split band voice coding schemes. In *Proceedings ICASSP*, pages 191–195, 1977.

[9] J M Foley and G M Boynton. A new model of human luminance pattern vision mechanisms: Analysis of the effects of pattern orientation, spatial phase, and temporal frequency. In T A Lawton, editor, *Computational Vision Based on Neurobiology, SPIE Proceedings, volume 2054*, 1994.

[10] D Gabor. Theory of communication. *J.IEE (London)*, 93:429–457, 1946.

[11] N Graham. *Visual Pattern Analyzers*. Oxford University Press, 1989.

[12] G H Granlund. In search of a general picture processing operator. *Computer Vision, Graphics, and Image Processing*, 8:155–173, 1978.

[13] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992a.

[14] D J Heeger. Half-squaring in responses of cat simple cells. *Visual Neuroscience*, 9:427–443, 1992b.

[15] D J Heeger. Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70:1885–1898, 1993.

[16] S G Mallat. A theory for multresolution signal decomposition: The wavelet representation. *IEEE Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[17] D Marr. *Vision*. W. H. Freeman and Co., San Francisco, 1982.

[18] J A Movshon, I D Thompson, and D J Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology (London)*, 283:53–77, 1978a.

[19] T R Reed, V R Algazi, G E Ford, and I Hussain. Perceptually based coding of monochrome and color still images. In *Data Compression Conference*, pages 142–151, 1992.

[20] R J Safranek, J D Johnston, N S Jayant, and C Podilchuk. Perceptual coding of image signals. In *Twenty-fourth Asilomar Conference on Signals, Systems and Computers*, pages 346–350, 1990.

[21] E Simoncelli and E H Adelson. Non-separable extensions of quadrature mirror filters to multiple dimensions. *Proceedings of the IEEE*, 4:652–664, 1990a.

[22] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Transactions on Information Theory, Special Issue on Wavelets*, 38:587–607, 1992.

[23] G Strang. Wavelets and dilation equations: a brief introduction. *SIAM Review*, 31:614–627, 1989.

[24] M Vetterli. Multi-dimensional sub-band coding: some theory and algorithms. *Signal Processing*, 6:97–112, 1984.

[25] A B Watson. Visually optimal DCT quantization matrices for individual images. In *Data Compression Conference*, pages 178–187, 1993.

[26] A P Witkin and J M Tenenbaum. On the role of structure in vision. In A P Pentland, editor, *From Pixels to Predicates*, pages 481–543. Ablex Publishing Co., Norwood, NJ, 1985.