



July 1989

Receptive Fields for the Determination of Textured Surface Inclination

M. R. Turner
University of Pennsylvania

Marcos Salganicoff
University of Pennsylvania

G. L. Gerstein
University of Pennsylvania

Ruzena Bajcsy
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/cis_reports

Recommended Citation

M. R. Turner, Marcos Salganicoff, G. L. Gerstein, and Ruzena Bajcsy, "Receptive Fields for the Determination of Textured Surface Inclination", . July 1989.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-89-48.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/cis_reports/842
For more information, please contact repository@pobox.upenn.edu.

Receptive Fields for the Determination of Textured Surface Inclination

Abstract

The image of a uniformly textured inclined surface exhibits systematic distortions which affect the projection of the spatial frequencies of which the texture is composed. Using a set of filters having suitable spatial, frequency and orientation resolution, the inclination angle of the textured surface may be estimated from the resulting spatial frequency gradients. Psychophysical experiments suggest that, in absence of other cues, humans perceive surface inclination from perspective distortions, suggesting the possibility of a specific neuronal mechanism in the visual system. Beginning with a low level filter model found to be an accurate and economical model for simple cell receptive fields, we have developed both algorithmic machine vision and neural network models to investigate physiologically plausible mechanisms for this behavior. The two models are related through a new class of receptive field formed in the hidden layer of a neural network which "learned" to solve the problem. This receptive field can also be described analytically from the analysis developed for the algorithmic study. This paper, then, offers a prediction for a new type of receptive field in cortex which may be involved in the perception of inclined textured surfaces.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-89-48.

**RECEPTIVE FIELDS for the
DETERMINATION
of TEXTURED SURFACE
INCLINATION**

**M.R. Turner, M. Salganicoff
G.L. Gerstein, and R. Bajcsy
MS-CIS-89-48
GRASP LAB 187**

**Department of Computer and Information
Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104**

July 1989

ACKNOWLEDGEMENTS: Support for this research was provided by ONR N00014-87-0766 and AFOSR 88-0296.

RECEPTIVE FIELDS FOR THE DETERMINATION OF TEXTURED SURFACE INCLINATION

M. R. Turner^{%,#}, M. Salganicoff^{%,#}, G. L. Gerstein[%] and R. Bajcsy[#]

[%]Department of Physiology

[#]Department of Computer and Information Science
University of Pennsylvania
Philadelphia, Pennsylvania 19104

ABSTRACT

The image of a uniformly textured inclined surface exhibits systematic distortions which affect the projection of the spatial frequencies of which the texture is composed. Using a set of filters having suitable spatial, frequency and orientation resolution, the inclination angle of the textured surface may be estimated from the resulting spatial frequency gradients. Psychophysical experiments suggest that, in absence of other cues, humans perceive surface inclination from perspective distortions, suggesting the possibility of a specific neuronal mechanism in the visual system. Beginning with a low level filter model found to be an accurate and economical model for simple cell receptive fields, we have developed both algorithmic machine vision and neural network models to investigate physiologically plausible mechanisms for this behavior. The two models are related through a new class of receptive field formed in the hidden layer of a neural network which "learned" to solve the problem. This receptive field can also be described analytically from the analysis developed for the algorithmic study. This paper, then, offers a prediction for a new type of receptive field in cortex which may be involved in the perception of inclined textured surfaces.

INTRODUCTION

The term "computation" when applied to a task in perception usually signifies a transformation from a set or sequence of stimuli to a model or interpretation of some aspect of the external world. The means by which this is accomplished depends in large part upon the architecture of the information processing device performing the transformation. In most traditional machine perception applications the computation executes on a conventionally programmed, general purpose, digital computer. Based upon an analysis of the mappings between properties of the physical world and the values measured using perceptual sensors, a program is constructed which realizes a computation in a unified algorithmic form.

Most perceptual processing tasks in the world, however, are performed by biological brains with architectures vastly different from those of digital computers. The cortex seems to be organized into regions which specialize in the analysis of particular aspects of sensory inputs. Within each region the computations are spread across large numbers of highly interconnected elements which, in and of themselves, are relatively simple information processing devices. Rather than having the concentrated algorithmic form of the digital computer, the computations of real neural architectures are distributed across an array of elements, each

contributing a portion of the overall transformation.

Since the early experiments of Kuffler (1953) neurophysiologists have attempted to characterize the response properties of neurons by defining their "receptive fields" or the comparable term, "tuning curves." The hope was that the overall computation could be inferred from the pieces of transformation produced by each individual neuron. This, however, is a difficult undertaking without the structure provided by an underlying theoretical model (Hopfield and Tank 1986; Sejnowski et al. 1988). Unfortunately, many computational models are formulated in terms of an idealized algorithm which often has no obvious parallel in the architecture and processes of real neural assemblies.

Although simulated neural network models are often simplified to the point of physiological inaccuracy, they share with their biological counterparts a type of highly parallel distributed processing. The learning algorithms used with many such systems have not been proven optimal in the sense of finding the best set of weights for a given training set. Nevertheless, there is much empirical evidence that they find very good solutions in a variety of tasks. Proceeding on the fundamental methodological assumption that, given similar constraints, evolutionary pressures would lead to similar solutions, we may use these systems to draw predictive inferences for biological systems.

Several researchers have taken this approach and applied learning algorithms in concert with stylized neuronal nets to "hard" problems which the real nervous system solves with ease; among these are shape from shading (Lehky and Sejnowski 1987), oculomotor compensation in the visual field (Zipser and Anderson 1989) and inverse kinematics of arm motion (Kuperstein 1988). In the oculomotor work, it was found that the network solved the problem using receptive fields very similar to those believed to be performing the same task in the brain. Moreover, the network solution to the shape from shading problem utilized biologically realistic units in previously unconsidered ways. Many times, however, the use of neural networks with learning algorithms only pushes the problem back one level. Although the network learns to perform the given task and makes it easy to examine the "receptive" and "projective" fields which form, the problem of identifying a computational transformation from the resulting system still remains.

Consequently, we have applied two entirely separate approaches to a problem in visual perception, that problem being the estimation of a textured surface's inclination from a monocular perspective view. Both conventional algorithmic and "back-propagation" neural network models were studied. A connection between the two types of models was then found in a new class of receptive field formed in the hidden layer of a neural network which "learned" to solve this problem and which can also be described analytically from the equations developed for the algorithmic solution. Although such receptive fields have not yet been described in the physiological literature, it is known that humans (and, presumably, animals) have the ability to solve this same class of visual problem.

Psychophysical studies have demonstrated that humans can perceive surface inclination from the distortions which occur in the perspective transformation (Gibson 1950b; Gruber and Clark 1956; Flock and Moscatelli 1964) It is clear that in most cases the visual system works with a multitude of available surface cues to improve the accuracy and reliability of surface inclination estimates. Nevertheless, when other cues are eliminated the ability to estimate surface inclination from perspective distortions of textured surfaces is retained, suggesting the

possibility of a specific neuronal mechanism for this purpose in the visual system. The models described in this paper, then, offer a prediction for a new receptive field in cortex which might be involved in the perception of inclined textured surfaces.

The paper is organized as follows. The first section describes the properties of perspective imaging. Section 2 introduces the low level filter, the 2D Gabor filter, applied to the perspective images. Using the imaging and low level filter models, section 3 describes the relationship between the amplitudes of 2D Gabor filters and inclined textured surfaces. The amplitude distributions which arise from application of a systematically varying set of filters to perspective images of textured surfaces form a characteristic pattern which may, as shown in section 4, be considered a model for a receptive field. Section 5 presents a neural network model which, in learning to solve this problem, formed receptive fields much like those described in section 4. The final section discusses certain differences between the "learned" receptive fields and their idealized analytic description, directions for future work, and the physiological implications of this new receptive field class.

PERSPECTIVE PROJECTIONS AND TEXTURE GRADIENTS

A perspective transformation with a pinhole camera (as shown in figure 1) is a much simplified but adequate approximation to that occurring in the eye. When a textured planar surface having an oblique inclination to the viewer's line of sight is perceived with a perspective transformation, several systematic distortions occur to the projected texture elements across the image. First, there is a distance effect which decreases the texture element size and increases element density as the plane recedes into the distance. Elements on the planar surface which are farther away from the viewer appear smaller. In addition, there is a second type of distortion known as the "foreshortening effect". At points farther away on the texture plane the angle between the line of sight and the tangent to the plane decreases. This tends to shrink the texture element size along the direction having greatest orientation to the viewer. Perspective distortions result from a combination of these two effects.

The systematic distortions which affect the size and density of texture elements in perspective projections also affect the projection of the spatial frequencies of which the texture is comprised. As a simple example, the image of a surface containing a uniform sine wave grating slanted 60 degrees from the vertical (see figure 3a) will exhibit systematic local changes in frequency. By identifying these frequency shifts at different locations in the image and with some knowledge of the parameters of the imaging system (e.g. focal length of the camera) we may estimate the slant of the original surface. This requires a suitable spatially and spectrally local measurement of the image content. We have used 2D Gabor functions because of their physiological relevance as simple cell receptive field models; nevertheless, there are a number of other filter choices which could serve the same purpose. Within a machine vision context, for example, Bajcsy and Lieberman (1976) used windowed Fourier power spectra to measure the frequency shifts of inclined textured surfaces.

2D GABOR FUNCTIONS

The model begins at the level of simple cells in V1 in the visual cortex. 2D Gabor functions have been found to be an accurate and economical description of these simple cell receptive

fields (Jones and Palmer, 1987a, 1987b, Jones et al., 1987). A 2D Gabor function may be realized as the product of a sinusoidal plane wave of some frequency and orientation and a two dimensional Gaussian (see figure 2). They provide, therefore, a measurement of image content which is both frequency and orientation selective as well as being spatially localized. There is an uncertainty relation which limits the simultaneous spatial and spectral selectivity possible with a linear filter. Daugman (1985) has shown that 2D Gabor functions are, in some sense, optimal in that they achieve the maximum possible joint resolution in the spatial and frequency domains allowed by the uncertainty relation.

The set of Gabor filters used for this study are shown in figure 2. This set contains 12 frequencies and 2 orientations. For each frequency and orientation specification a pair of filters is generated which differ in phase by 90 degrees (approximately quadrature). A filter is applied to a particular location in the image by taking the inner product of the filter with the image material centered at that location. A phase insensitive amplitude is then computed by taking the Pythagorean sum of the values obtained using a pair of filters with identical frequency, orientation and image location, differing only in phase. (Further details are available in Turner 1986 and Turner et al. 1989a)

GABOR FILTERS AND TEXTURE GRADIENTS

When applied to the image of an inclined grating (such as figure 3a) the amplitudes obtained using a particular filter pair will vary from location to location in the image depending upon the local match between the projected spatial frequency and the frequency -- orientation of the Gabor filter phase pair. This may be visualized by displaying the amplitudes as an array of small squares, each one centered at the image location from which the amplitude was measured and with a gray level intensity proportional to that amplitude. A set of six such displays, each for a different filter frequency, are shown in row b of figure 3 for the sine wave grating of 3a. The bright band in each represents the spatial region in which the frequency of that particular Gabor filter pair best matches the projected spatial frequency of the inclined sine wave grating.

With a change in the inclination of the planar surface or the frequency of the grating, the distribution of amplitudes in each display would change, with the high amplitude bands appearing at different locations and/or with different widths. (As an extreme example, if the grating were parallel to the image plane the amplitudes of all filters would be uniform across the image.) The patterns of amplitudes which arise across different filter frequencies and image locations are determined by the surface inclination and texture frequency being projected. These patterns afford a means by which to identify the inclination of textured surfaces having strong spectral peaks.

An analytic characterization of the amplitude patterns is presented in the appendix to this paper. Using equations for the perspective transformation, a new set of equations is derived which, given a particular surface inclination and uniform texture spatial frequency, describe the frequency shifts which occur at different locations in the image plane. The amplitude of a particular Gabor filter pair measuring this projected spatial frequency may then be evaluated using the Gabor filter's Fourier transform. Thus, given the salient properties of the imaging system, those of the Gabor filter pair used in the measurement and the texture frequencies being projected, this equation may be used to predict the amplitudes which would be

measured at different locations in the image plane.

We had previously developed an algorithm which uses this explicit formalism in estimating the inclination of a textured surface (Turner et al. 1989a). An inclination estimate is made by adjusting a set of parameters to reduce the difference between amplitudes calculated using the predictive equation and those measured from an image using a set of Gabor filters. While the algorithm may in some ways be suggestive as a model for texture surface perception in the visual system (Turner et al. 1989b), it does not provide much insight at the level of a neuronal mechanism.

A RECEPTIVE FIELD FOR THE PERCEPTION OF TEXTURE FREQUENCY GRADIENTS

Nevertheless, the analysis developed for the algorithmic model may also be used to predict the form of a receptive field. The set of amplitudes shown in row b of figure 3 may be regarded as slices through a higher level receptive field which takes Gabor amplitudes as input and signals by its output the presence of a particular inclination angle and texture frequency in the image. Viewed in this way the higher amplitude regions (the bright band across each display) represent the excitatory regions of the receptive field and the darker areas inhibitory regions. However, with 2 spatial dimensions (the x and y locations where the Gabor filter pairs are applied) and 2 frequency dimensions (the frequency and orientation of the Gabor filters in the array) the general form of such a receptive field is 4 dimensional. Since a 4 dimensional object is difficult to represent, the simplification achieved by limiting the receptive field to one spatial and frequency dimension, though less general, facilitates investigation and presentation of the receptive field structure.

Accordingly, the center column of each of the displays in row b has been extracted in row c. Each of the 6 columns contains the amplitudes measured from the center strip of the inclined grating image (y varies, x is held constant at the image center) using a Gabor filter of a particular frequency and orientation. From column to column the frequency of the Gabor filter used is systematically varied (the orientation of the filters is held constant). Note that the brightest, high amplitude square appears in a different place in each of the 6 columns of row c. The overall structure of the receptive field becomes evident when the columns are concatenated together along the frequency dimension. This is shown in figure 3d at a much finer resolution than the 6 displays shown in rows b and c. Figure 3d contains 128 points along the frequency dimension (rather than the 6 columns in row c) by 512 points along the y dimension (only 9 spatial locations are shown in each column of row c). More importantly, however, figure 3d was not produced by applying a much larger, finer resolution set of Gabor filters to the inclined grating image. Rather, an amplitude for each point in figure 3d was directly calculated by systematically varying the spatial and frequency variables of the prediction equation developed for the algorithmic model.

The receptive field which emerges can be described as an elongated, oriented excitatory band flanked by inhibitory surrounds. This would, in fact, be a familiar characterization of simple cells in visual cortex if the coordinates were purely spatial. However, in this case the coordinates are space and spatial frequency. Thus, these receptive fields are detectors of different gradients of particular spatial frequencies in the visual image. The orientation of the excitatory band corresponds to the rate at which the spatial frequency preference changes with spatial location, and hence the inclination angle of the textured surface to which the neuron is tuned

(see figure 4a and b for examples sensitive to different inclinations). These 2 dimensional receptive fields may be considered slices across the oriented excitatory slabs of their more general 4 dimensional counterparts.

The analysis described above, initially done for the machine vision algorithm, pointed to a plausible receptive field model for texture gradient perception. The neural network learning algorithm afforded us a completely different tool with which to investigate the same visual problem. If these receptive fields also arose as a part of the neural network solution, the commonality of this feature between the two very different approaches, combined with the evidence that neural networks often arrive at physiologically realistic mechanisms, would suggest the possibility that such receptive fields might contribute to a cortical solution to the same perceptual task.

To this end we created a 3 layer neural network simulation whose inputs were the phase insensitive amplitudes of a number of Gabor functions of different frequencies, orientations and spatial locations in the image. The training corpus consisted of the amplitudes measured from a set of 192 textured images, each at 6 angles of slant. Neurons in the output layer represented the different angles of slant as computed by the network. A back-propagation algorithm was used to adjust the connection weights. After training with examples the system was studied with particular attention to the hidden layer receptive fields.

THE NEURAL NETWORK SIMULATION

A three layer back-propagation network was created using the Rochester Connectionist simulator (Goddard et al. 1988) executing on a SUN4/260 workstation. Each input unit was associated with one of the 24 filter frequency-orientation pairs (12 frequencies x 2 orientations) at one of the six image sample points, making a total of 144 input units (see figure 5). These input units were fully connected to the middle layer units via modifiable "synaptic" weights which could either be excitatory or inhibitory. The middle layer units were then fully connected to the six output units via another level of modifiable weights.

Desired outputs corresponding to the known inclination angle of the plane were encoded in the six output units. The units were tuned for 0, 10, 20, 30, 40 and 50 degrees respectively. Each output unit had a Gaussian tuning curve with a standard deviation of 10 degrees. These desired output values were presented coincident with the Gabor amplitude values during training runs to form an association pair. This standard deviation provided overlap in the tuning curves of output units; this is a useful mechanism for disambiguating the output of a group of neurons and occurs in several physiological systems (Erickson 1963, 1974).

The number of middle layer units in the network was varied to see if it affected the ultimate accuracy of the network. This did not seem to be the case, as roughly similar performance levels were achieved in networks with 4,5,6,7,16 and 32 middle layer units; only marginally better performance was noted with the 32 middle layer unit network. Interestingly, the number of middle units with high gradient receptive fields between the input and middle layer in any given network instantiation after training was approximately 7. For example in the 32 middle layer unit network, the majority of the receptive fields were flat, i.e. nonselective, except for a small minority of about 7 that were highly selective. Thus, the network seems to converge in a relatively consistent way, invariant of the number of middle layer units.

July 20, 1989

The Learning Algorithm

A gradient-descent "back-propagation" learning algorithm was used to determine the interconnection weights between the layers of the network. At each iteration this method attempts to minimize the squared error between the output computed by the network for a given input, and the desired or correct output presented simultaneously to the network by a "teacher". The system performs a gradient descent search, adjusting the interconnection weights in the direction which most rapidly decreases that error. Details of this algorithm may be found in (Rumelhart et al. 1986). Although it does not have a strong physiological basis, back-propagation was selected mainly because of its relative ease of implementation and its speed in identifying interconnection weights to solve the given visual problem with relatively low error.

Training

The raw images for the training set consisted of 2 natural and 3 synthetic textures. These included a metal grate with a hexagonal pattern, a brick wall, a horizontal sinusoidal grating, a vertical sinusoidal grating, and a sum of horizontal and vertical sines. Both the natural textures were taken from Brodatz (1966) via Weber (1986). These textures were selected specifically because they contain the isolated spectral peaks that would allow clear determination of frequency gradients. A total of 32 input texture variants were generated from this base by picking various vertical strips at different parts of the natural textures for filter application or by changing the frequencies slightly in the synthetic ones. Using standard computer graphics techniques these textures were mapped onto inclined planar surfaces and projected to the image plane via a standard perspective projection matrix. Six angles were used (0, 10, 20, 30, 40 and 50 degrees, 0 being perpendicular to the viewing plane) with the textured surfaces progressively inclined away from the viewer at the top of the image. Using Stevens' (1983) notation for surface inclination these were positive slant values with a constant tilt of 90 degrees.

The input data to the neural network was generated by the application of a set of filters to the raw image set. The filters were selected to cover the bulk of the orientation and frequency peaks in the learning set and would therefore provide good coverage of the local spectral content of the images at those points. With the dimensional simplification described in the previous section, filters were applied in a vertical strip down the center of the image. The values of the Gabor amplitudes were spatially averaged into 6 representative region values spanning the top to the bottom of the image. Each Gabor amplitude distribution of the training corpus was presented to the input layer simultaneously with the correct inclination angle encoded on the output layer. The learning algorithm was then iterated through one cycle. This process was repeated with the entire learning set of 192 stimulus-response pairs presented a total of 500 times.

In order to prevent the network from specializing on the absolute values of the input and output array, each input array data set was multiplied by a positive random constant which varied from .35 to .95. This corresponds to a random variation in the contrast of the raw input image scene. It was hoped that this would force the network to search for features in the amplitude distribution that would be invariant of illumination on the scene. This random fluctuation also served to prevent the learning algorithm from settling on local minima in the error space, since

the random nature of the illumination could be considered a perturbation of the networks position in the weight space, and might serve to remove it from this local minimum.

The network encoded the training set quickly and with excellent error results. In all cases, the network seemed to converge to an asymptotic level within approximately 200 presentation of the learning corpus (see figure 6). Error was computed by performing a normalized correlation between the actual network output and the desired known slant value, as encoded using the Gaussian tuning curves. This type of correlation can be essentially thought of as an normalized inner product between the desired and actual output vector, and is a good measure of the ratio encoding between the output units which specify the output value indicated by the network in a way which is invariant of the absolute output potentials of the output units. The total corpus correlation was computed by taking the average of all the correlations in the learning set. The network shown in the figure (with 7 middle units) reached a maximum corpus correlation of .975 with the learning set at random illumination levels. This value is representative of the performance also obtained with other numbers of middle layer units.

Hidden Layer Receptive fields

The "receptive field" of a hidden layer unit is defined by the connection weights of the input layer units which project to it. These may be organized along the same axes as the analytic receptive field plots to facilitate comparison, i.e. with respect to the spatial location and frequency of the Gabor filter associated with each input unit. Several different types of receptive fields from the simulation runs are shown in figure 7, each for the weights of input units corresponding to a single Gabor filter orientation. The first three (a) have the same basic structure as the analytic receptive fields, with an excitatory diagonal flanked by inhibitory regions. Note that one of these fields has a "reversed" slope, suggesting that it is used in a NOT sense. Receptive fields (b) are arranged in the complementary manner to (a), with an inhibitory diagonal and excitatory surrounds. In fact, this is the predominant form of the receptive fields with oriented structure. For reasons described in Rumelhart et al. (1986, p.351) the learning algorithm tends to select inhibitory receptive field types. Most of the simulation runs contained some receptive fields like the final examples (c), with no clearly evident structure. These may represent attempts by the neural net to identify specific features of individual stimuli in the learning set, or may be used for contrast normalization.

DISCUSSION

Some of the receptive fields formed in the neural network hidden layer have an oriented structure similar in appearance to those defined from the algorithmic analysis. Elements with this structure may be viewed as "grandmother" cells for particular inclination angles and texture frequencies across arrays of Gabor filters of particular frequencies and bandwidths. Yet each of these higher level elements has a bandwidth of its own, responding in a graded manner as the inclination and other parameters of the stimulus deviate from its preference. Therefore, it is likely that surface inclination is better identified by the assembly response of ensembles of these elements (Georgopolis 1988), rather than by the identification of a single element with the greatest output within an ensemble. This appears to be happening in the neural network system.

In the simulations done so far, elements in the output layer have been tied to a number of hidden layer units in ways which have made it difficult to identify the exact transformation being used by the system at this stage. A project is currently underway to explicitly study this aspect of the computation using equation generated sets of hidden layer elements. Furthermore, it is not known what constitutes an "optimal" set of these oriented receptive field elements either in general or for specific textured material. The neural network seems to have used 7 or fewer hidden layer elements for the training corpus, but not all of these had the oriented signature of the analytic receptive fields.

This points to a significant difference between the two approaches. While many hidden layer units had an oriented structure, none was as cleanly or precisely defined as the analytic versions. Each hidden layer receptive field possessed some degree of irregularity. The algorithmic analysis provided an idealization of the receptive field which, in practice, the neural network never attained. While the learning set may contain an insufficient number of texture types to allow perfect abstraction of the problem, it is also possible that the ideal receptive fields would never develop no matter how large the training set because the irregularities contribute to the transformation in an as yet undetermined manner.

The neural network study has been confined to textured planar surfaces with one angle of inclination (the surface slant is varied while the tilt is fixed). It is obviously appropriate to extend the problem to surfaces of arbitrary inclination, and to receptive fields in their more general four dimensional form. The higher dimensionality will pose obvious display problems, necessitating additional methods for examining slices through the multidimensional receptive field structures. Nevertheless, the system may still be studied using a similar three layer, back propagation network.

It is also appropriate to consider receptive fields with systematic structure in other higher dimensional feature spaces. One example is the spatio-temporal filtering model of Adelson and Bergen (1985) for motion perception. Their model differs somewhat from the one presented here in that, rather than using separate receptive fields to identify structure in low level filter distributions, the low level filters themselves are oriented in the higher dimensional space. Their work has served as a theoretical foundation for physiological experiments which have identified cells with spatio-temporal tuning properties in visual cortex (McLean et al. 1987).

Similarly, a major consequence of our observation of elongated, oriented, excitatory-inhibitory receptive fields in a y - f space is a prediction for experimentation. We should look in the mammalian visual system for neurons that have a systematically varying preferred spatial frequency across their receptive field. In other words, we are raising the suggestion that somewhere in the visual system there could be large receptive fields that exhibit clear gradients of frequency tuning properties across their extent. As far as we know, neurons such as these have neither been sought nor reported. Experimental determination of such tuning properties is time consuming, but perfectly feasible. Neurons with the type of receptive fields suggested could be detected by the tachistoscopic projection of synthesized textures with specific spatial frequency and orientation gradients. Essentially, we would then exhaustively parameterize the x , y , orientation and frequency space at some sampling interval, searching for unique combinations of these that might lead to heightened activity within a given higher order neuron or assembly of neurons. If a family of such selective neurons is found this would substantiate the existence of this type of second order neuronal receptive field process. Possibly a

more efficient approach to such experiments might be through the use of a randomly shifting plane combined with backwards averaging from the observed neuronal spike train (Eggermont et al. 1983).

In summary, both the algorithmic and neural network models have suggested a mechanism for the perception of texture gradients which can, in principle, be easily attained by physiologically realistic weighted summations of simple cell like elements. The model predicts some receptive field properties that could be sought among real neurons in cortex.

ACKNOWLEDGEMENTS

Support for this research was provided by ONR N00014-87-K-0766 and AFOSR 88-0296. We thank Jayashree B. Gokhale for critical reading of the manuscript.

APPENDIX

In this section an equation is derived for estimating the amplitude of a 2D Gabor filter pair applied to the perspective projection of a slanted planar surface having a single or dominant spatial frequency. By systematically varying the space and frequency variables, this equation may also be used to describe receptive fields for the identification of certain texture gradients in distributions of Gabor filter amplitudes. Although defined here for textured surfaces with one axis of inclination (slant with constant tilt), the more general form for surfaces of arbitrary inclination is derived in Turner et al. (1989a).

Projection Function

The projection of points from the planar surface to the image plane is obtained by the rotation of the textured plane around its horizontal axis by the angle of inclination (slant), followed by a perspective transformation, giving us (see figure 1)

$$x = \frac{f u}{d + v \sin \theta} \quad (1)$$

$$y = \frac{f v \cos \theta}{d + v \sin \theta} \quad (2)$$

Inverting these for (u,v) gives

$$u = \frac{x d \cos \theta}{f \cos \theta - y \sin \theta} \quad (3)$$

$$v = \frac{y d}{f \cos \theta - y \sin \theta} \quad (4)$$

Projection of Texture Spectra

The frequency projected to the image plane at the point (x,y) is the product of the horizontal and vertical components of the texture frequency (F_u, F_v) and the Jacobian of the projection function

$$\begin{bmatrix} F_x & F_y \end{bmatrix} = \begin{bmatrix} F_u & F_v \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \quad (5)$$

Where

$$\frac{\partial u}{\partial x} = \frac{f d \cos^2 \theta - y d \sin \theta \cos \theta}{(f \cos \theta - y \sin \theta)^2} \quad (6)$$

$$\frac{\partial u}{\partial y} = \frac{x d \sin \theta \cos \theta}{(f \cos \theta - y \sin \theta)^2} \quad (7)$$

$$\frac{\partial v}{\partial x} = 0 \quad (8)$$

$$\frac{\partial v}{\partial y} = \frac{f d \cos \theta}{(f \cos \theta - y \sin \theta)^2} \quad (9)$$

2D Gabor Filters

The filters used in this work are discrete realizations of the following function:

$$A_{\phi}(x,y) = \exp\left[-\frac{(x - x_c)^2 + (y - y_c)^2}{2 \sigma^2}\right] \sin \omega[x \cos \alpha - y \sin \alpha + \phi] \quad (10)$$

where

x_c and y_c are the center locations of the Gabor filter Gaussian envelope.

σ is the standard deviation of the Gaussian envelope.

ω is the frequency of the sinusoidal plane wave.

α is the orientation of the plane wave.

ϕ is its phase.

With the projected texture frequency (F_x, F_y), the amplitude of a filter pair may be calculated using the Fourier transform of the Gabor function resulting in

$$A' = C 2\pi\sigma^2 \exp\left[-2\pi^2 \sigma^2 \left[(g_x - F_x)^2 + (g_y - F_y)^2 \right]\right] \quad (11)$$

where

C is the contrast of the texture frequency.

σ is the standard deviation of the Gaussian envelope of the 2D Gabor filter pair.

g_x and g_y are the horizontal and vertical components of the Gabor filter modulating frequency.

F_x and F_y are the horizontal and vertical frequency components from equation (5) projected to the image plane.

The projected texture frequency in the image plane is described above as a point function.

However, the 2D Gabor filter effectively occupies a finite spatial area. Thus, for function (11) to accurately approximate the Gabor filter amplitude distributions, the spatial size of the filters must be kept reasonably small to minimize the frequency variation over the effective area of the Gabor filter. For the examples in this paper, the filter bandwidth has been kept constant at 1.5 octaves, well within the range of physiological estimates (Kulikowski et al., 1982).

REFERENCES

- [1] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* , 2:284-299, 1985.
- [2] R. Bajcsy and L. Lieberman. Texture gradient as a depth cue. *Computer Graphics and Image Processing* , 5:52-67, 1976.
- [3] P. Brodatz. *Textures, a photographic album for artists and designers*. Dover, New York, 1966.
- [4] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* , 2:1160-1169, 1985.
- [5] J.J. Eggermont, P.I.M. Johannesma and A.M.H.J. Aertsen. Reverse-correlation methods in auditory research. *Quarterly Reviews of Biophysics*, 3: 341-414, 1983.
- [6] R.P. Erickson. Sensory neural patterns and gustation. In: Y. Zotterman, editor, *Olfaction and Taste* pages 205-213. MacMillan, New York, 1963.
- [7] R.P. Erickson. Parallel "population" neural coding in feature extraction. In: F.O. Schmitt. and F.G. Worden, editors, *The Neurosciences: Third Study Program* pages 155-169. MIT Press, Cambridge, Mass., 1974.
- [8] H. Flock and A. Moscatelli. Variables of surface texture and accuracy of space perceptions. *Perceptual and Motor Skills* , 19:327-334, 1964.
- [9] A.P. Georgopolis, R.E. Kettner and A.B. Schwartz. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J. Neurosci.* , 8:2928-2937, 1988.
- [10] J.J. Gibson. The perception of visual surfaces. *Am. J. Psychol.* , 63:367-384, 1950.
- [11] N.H. Goddard, K.J. Lynne and T. Mintz. Rochester connectionist simulator. *Technical Report TR233*, The University of Rochester, Computer Science Department, March 1988.
- [12] H.E. Gruber and W.C. Clark. Perception of slanted surfaces. *Perceptual and Motor Skills* , 6:97-106, 1956.
- [13] J.J. Hopfield and D.W. Tank. Computing with neural circuits: a model. *Science* , 233:625-633, 1986.
- [14] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiology* , 58:1233-1258, 1987.
- [15] J.P. Jones and L.A. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiology* , 58:1187-1211, 1987.
- [16] J.P. Jones, A. Stepnowski and L.A. Palmer. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J. Neurophysiology* , 58:1212-1232, 1987.
- [17] S.W. Kuffler. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiology* , 16:37-68, 1963.
- [18] J.J. Kulikowski, S. Marcelja and P.O. Bishop. Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biol. Cybern.* , 43:187-198, 1982.
- [19] M. Kuperstein. Neural model of adaptive hand-eye coordination for single postures. *Science* , 239:1308-1311, 1988.
- [20] S.R. Lehky and T.J. Sejnowski. Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* , 333:452-454, 1987.
- [21] J. McLean, S. Raab and L. Palmer. Spatiotemporally oriented simple receptive fields: local linear motion detectors. *Society for Neuroscience Abstracts*, 13: 1623, 1987.
- [22] D.E. Rumelhart, G.E. Hinton and R.J. Williams. Learning internal representations by error propagation. In: D.E. Rumelhart, J.L. McClelland, editors, *Parallel Distributed Processing: Exploring the Microstructures of Cognition.*" pages 318-364. MIT Press, Cambridge, Mass., 1986.

- [23] T.J. Sejnowski, C. Koch and P.S. Churchland. Computational Neuroscience. *Science* , 241:1299-1307, 1988.
- [24] K. Stevens. Slant-tilt: The visual encoding of surface orientation. *Biol. Cybern.* , 46:183-195, 1983.
- [25] M.R. Turner. Texture discrimination by gabor functions. *Biol. Cybern.* , 55:71-82, 1986.
- [26] M.R. Turner, R. Bajcsy and G.L. Gerstein. Estimation of textured surface inclination by parallel local spectral analysis. (submitted) 1989.
- [27] M.R. Turner, G.L. Gerstein and R. Bajcsy. Underestimation of texture slant by human observers: a model. (submitted) 1989.
- [28] A.G. Weber. Image data base. *Technical Report USC SIPI Report 101* , University of Southern California, Signal and image processing Institute, February 1986.
- [29] D. Zipser and R.A. Anderson. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* , 331:679-684, 1989.

FIGURE CAPTIONS

Figure 1: Perspective imaging model. A textured planar surface with coordinate system (u,v) is distance d from a pinhole camera with a flat image plane, coordinate system (x,y) and focal length f . The inclination of the textured surface is indicated by the angle θ .

Figure 2: The set of 48 2D Gabor filters in the top half of the figure are the ones used in calculating inclination of all the image examples which follow in the paper. The set contains 12 frequencies (8 - 30 pixel wavelengths in 2 pixel increments), 2 orientations (0, 90 degrees) with 2 phases (0 and 90 degrees) for each frequency and orientation specification. All filters have a bandwidth of 1.5 octaves, thereby making them scalings and rotations of the two larger filters (shown to provide a more detailed view of the smaller filters actually used) in the bottom half of the figure.

Figure 3: (a) The image of a surface containing a sine wave grating slanted 60 degrees from the vertical will exhibit systematic local changes in spatial frequency. (b) The amplitude distributions of a set of six 2D Gabor filter pairs each with a different frequency, applied to the sine wave image of (a). (The filters used here are a subset of the filters shown in figure 2, spanning 8 to 28 pixel wavelengths in increments of 4 pixels with an orientation of 90 degrees. Note that these filters are linearly spaced in wavelength.) The amplitude distributions for each filter pair may be visualized as an array of small squares (9 spatial locations per dimension are shown in each display). Each square is centered at the image location from which the amplitude was measured with a gray level intensity proportional to that amplitude. The high amplitude region (the bright band) in each display represents the spatial region in which the frequency - orientation of that particular Gabor filter pair best matches the projected frequency of the grating in the image. The set of amplitude distributions may be viewed as a model for a receptive field which takes amplitudes of a number of Gabor filter pairs as input and signals by its output the presence of a particular inclination angle and texture frequency in the image. The bright portions represent excitatory regions and the dark portions inhibitory regions. In general form this receptive field is 4 dimensional. (c) Nevertheless, a dimensional simplification may be made by limiting the receptive field to one spatial and one spatial frequency dimension. Accordingly, the center column of each display in (b) has been extracted. (d) When concatenated together along the frequency dimension with a much higher resolution than the 6 frequencies and 9 spatial locations of (c), the receptive field has an elongated, oriented structure with an excitatory band flanked by inhibitory surrounds. Since the coordinates of this receptive field are space and spatial frequency, these receptive fields are detectors of different gradients of particular spatial frequencies in the visual image.

Figure 4: Two receptive fields like the ones shown in figure 3d for different surface inclinations of the horizontal grating shown in figure 3a. Figure 4a is for a slant inclination of 40 degrees while 4b is for 70 degrees. The orientation and the width of the excitatory band vary with surface inclination.

Figure 5: Schematic of the neural network. The activity level of each input unit is proportional to the amplitude of one of the 2D Gabor filter pairs applied to the central strip of the image. Each unit in the input layer has a weighted connection to every unit in the

middle layer. Similarly, every unit in the middle layer has weighted connections to each unit in the output layer. The output layer encodes surface inclination angle by the combination of simultaneous activity levels of output units, each associated with a particular angular value (0, 10, 20, 30, 40, 50 degrees). Although there are only 6 output units, the overlap of inclination tuning curves from unit to unit allows a continuum of angles to be represented.

Figure 6: Typical learning curve for the adaptive neural network. In all cases, the network converged to an asymptotic level within approximately 200 presentations of the learning corpus. Error is computed by performing a normalized correlation between the actual network output unit activity levels and the tuning curve encoding of the desired inclination value.

Figure 7: Representative "receptive fields" for hidden layer units. Input unit weights are shown for a single orientation of the Gabor filters, and are arrayed with respect to y (spatial location on original picture material) and f (the filter spatial frequency). Size of square increases with larger weight; empty square is excitatory, solid black square is inhibitory. (a) Receptive fields with the same basic structure as the ones shown in figures 3d and 4: an excitatory band flanked by inhibitory surrounds. (b) Receptive fields which are the inhibitory complement of (a) and which occurred more frequently than the excitatory versions. (c) Most of the simulation runs also developed receptive fields without clearly evident structure like the ones shown in (a) and (b). The significance of such fields for the computation is not yet understood.

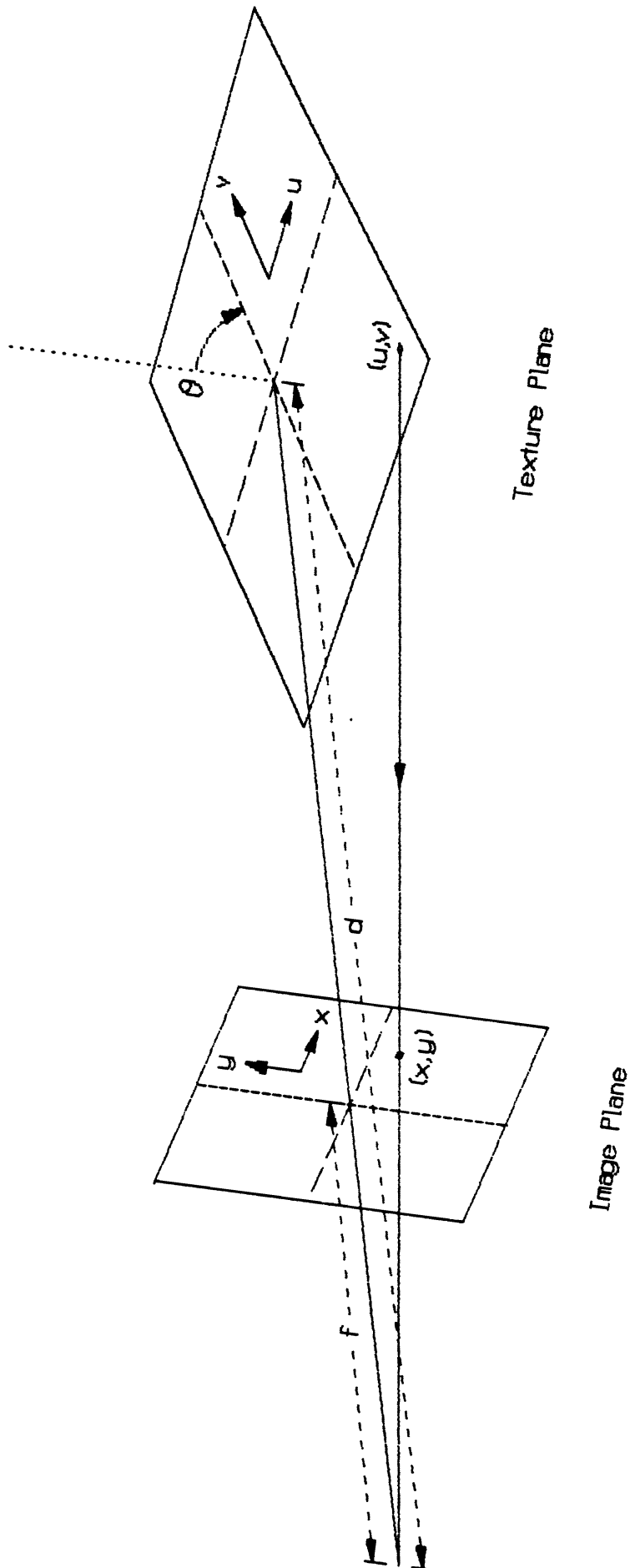


Figure 1

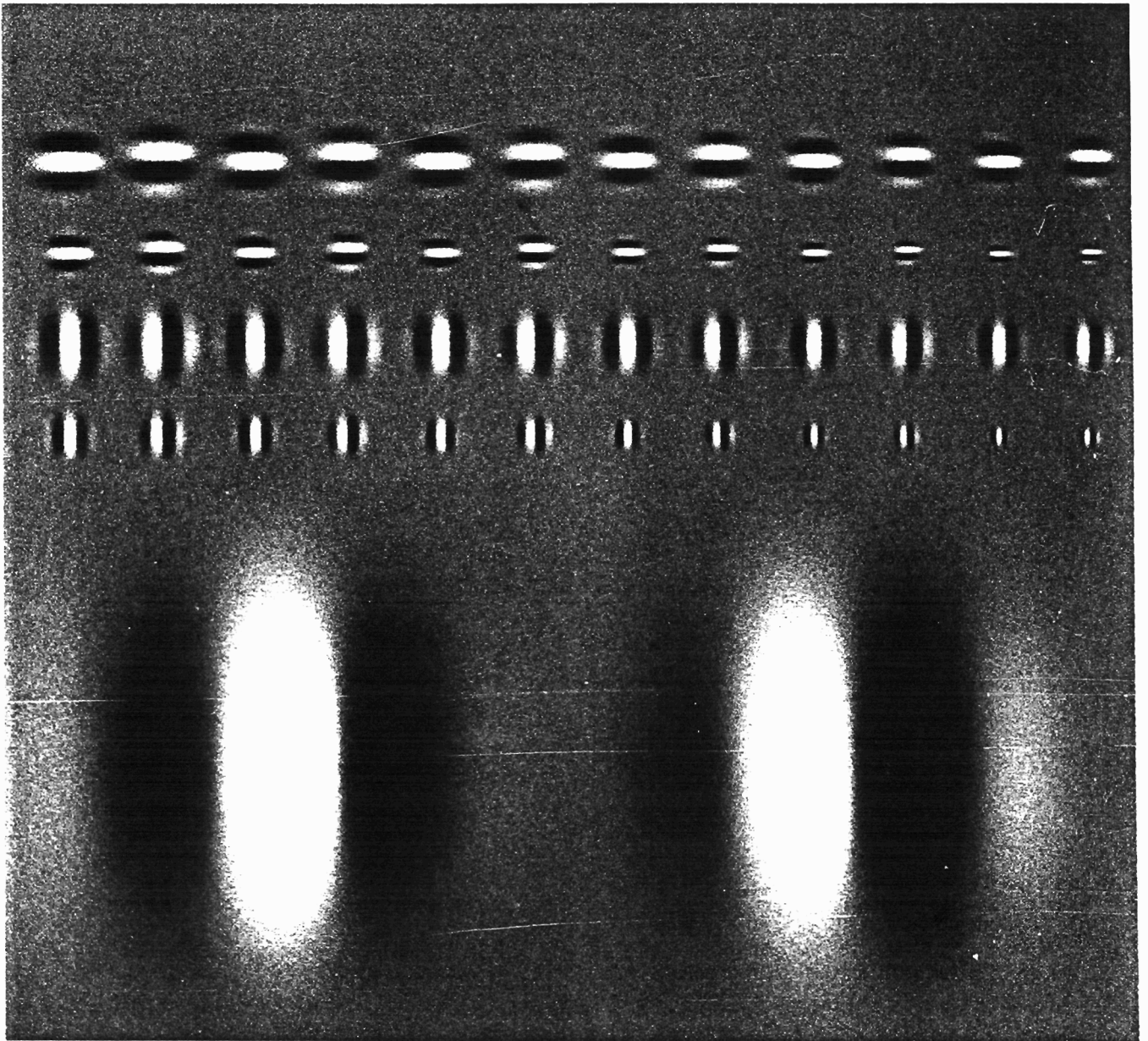


FIGURE 2

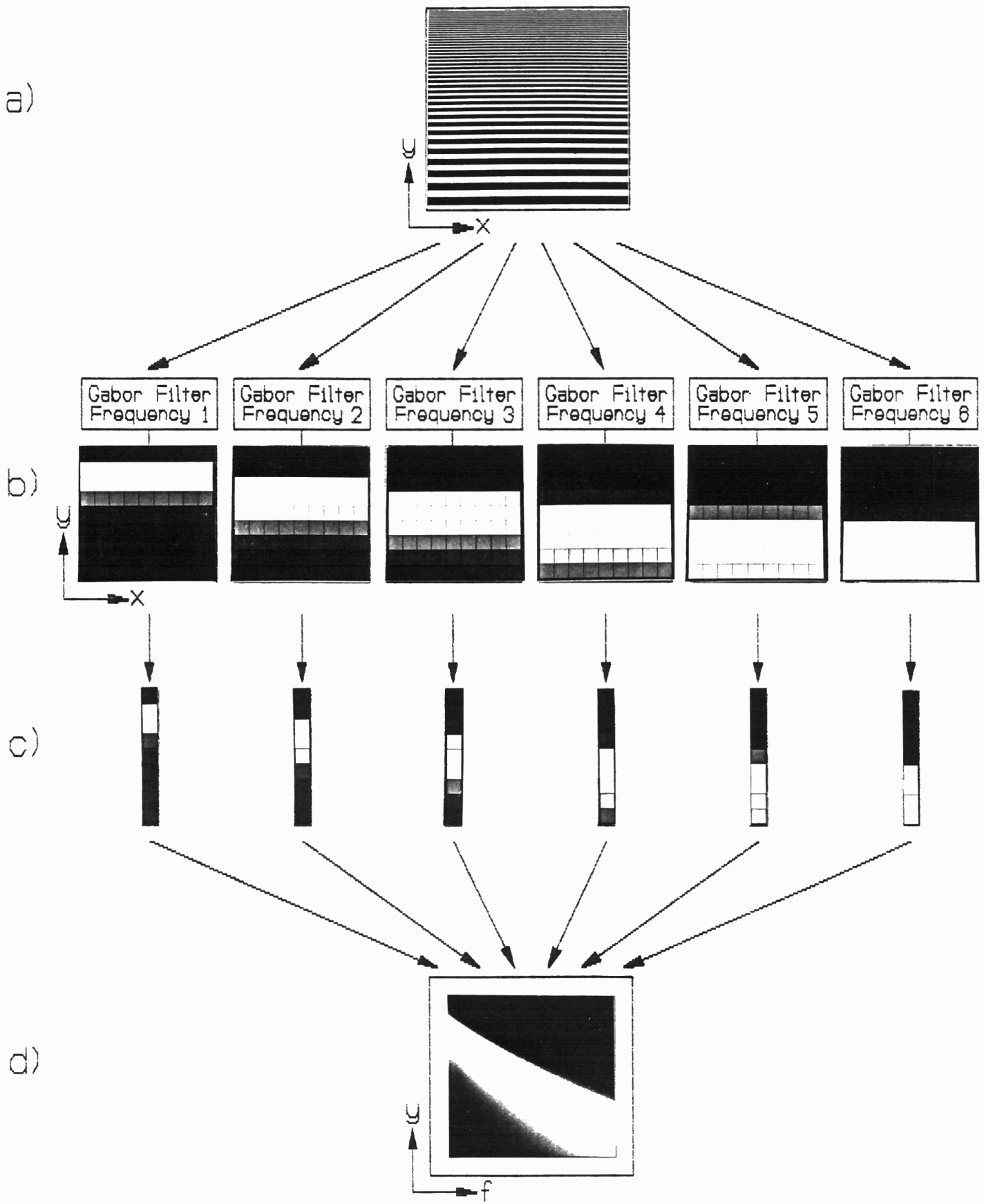
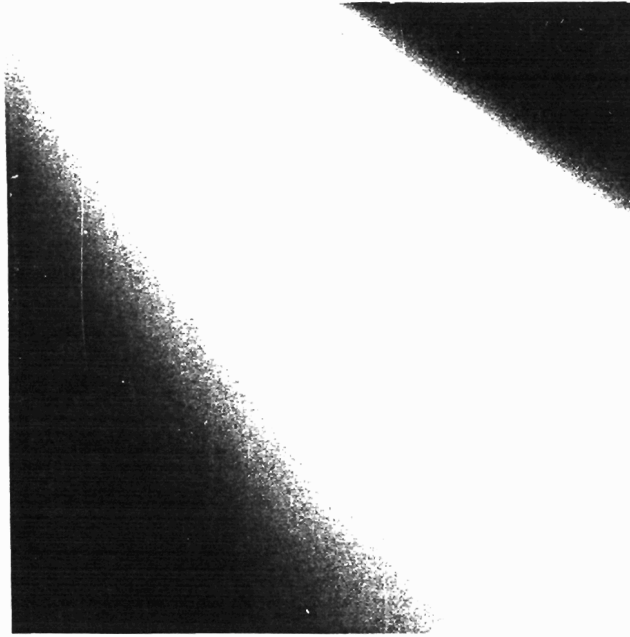
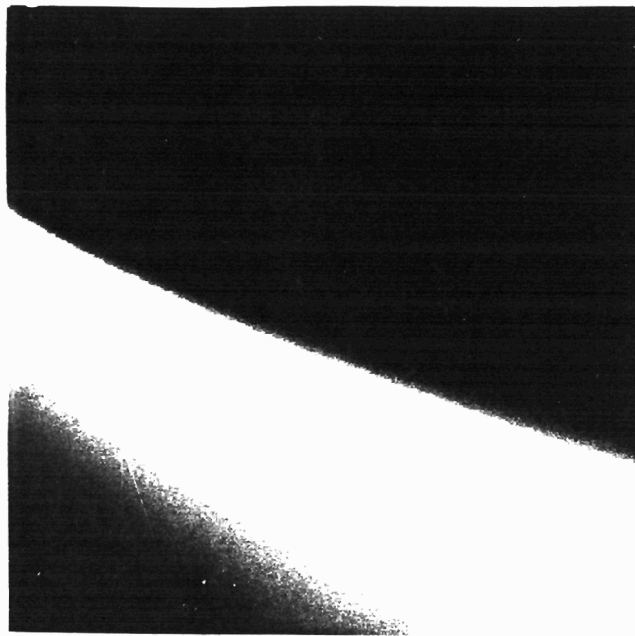


FIGURE 3



(a)



(b)

FIGURE 4

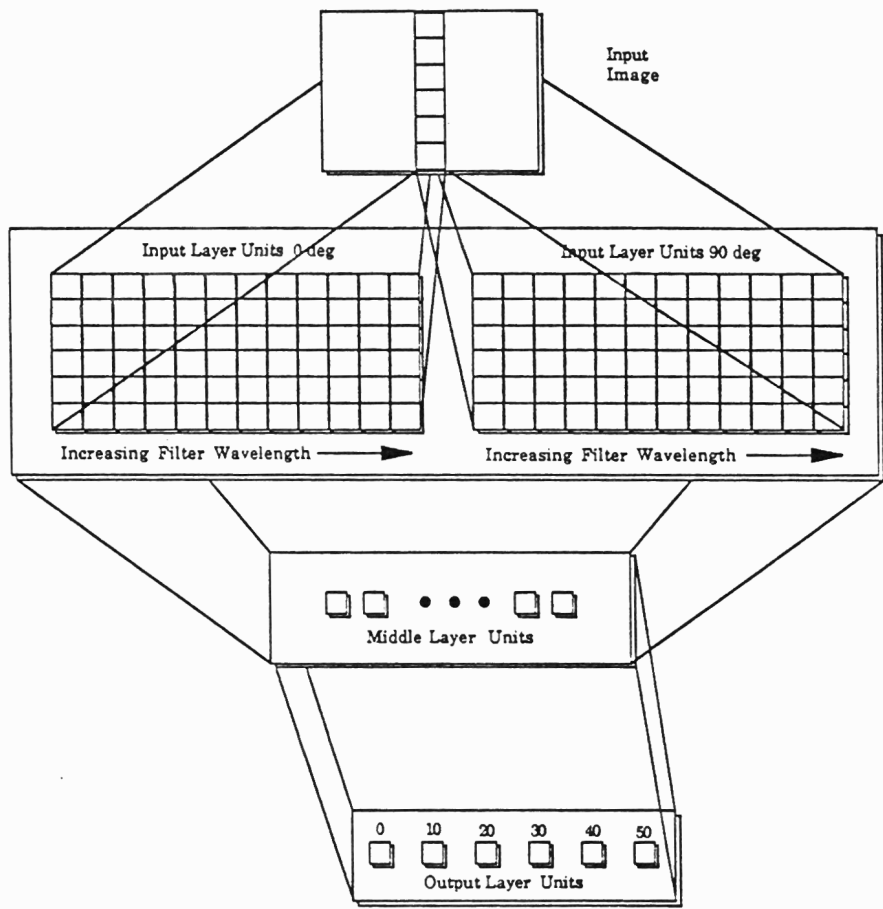


FIGURE 5

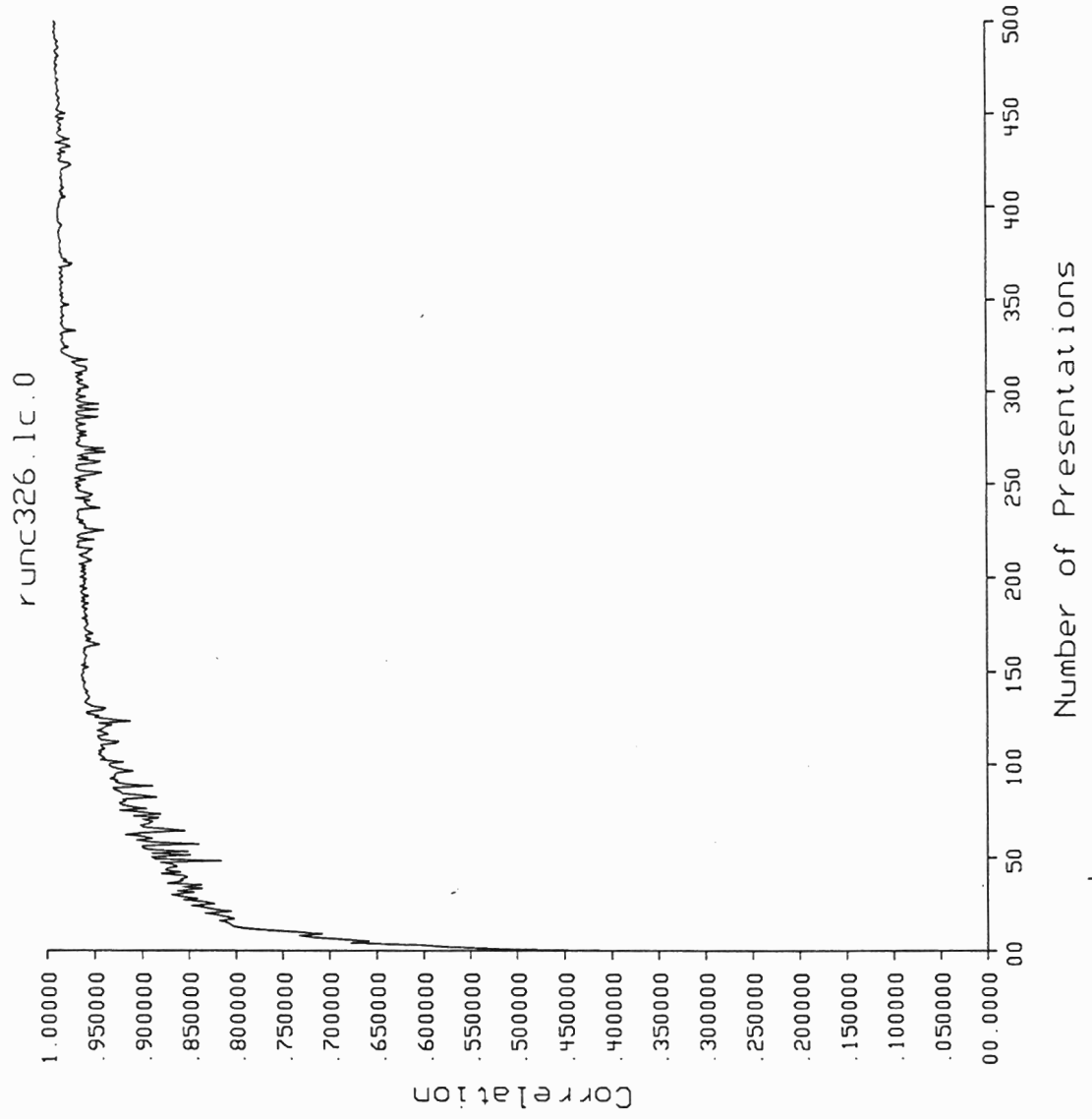


FIGURE 6

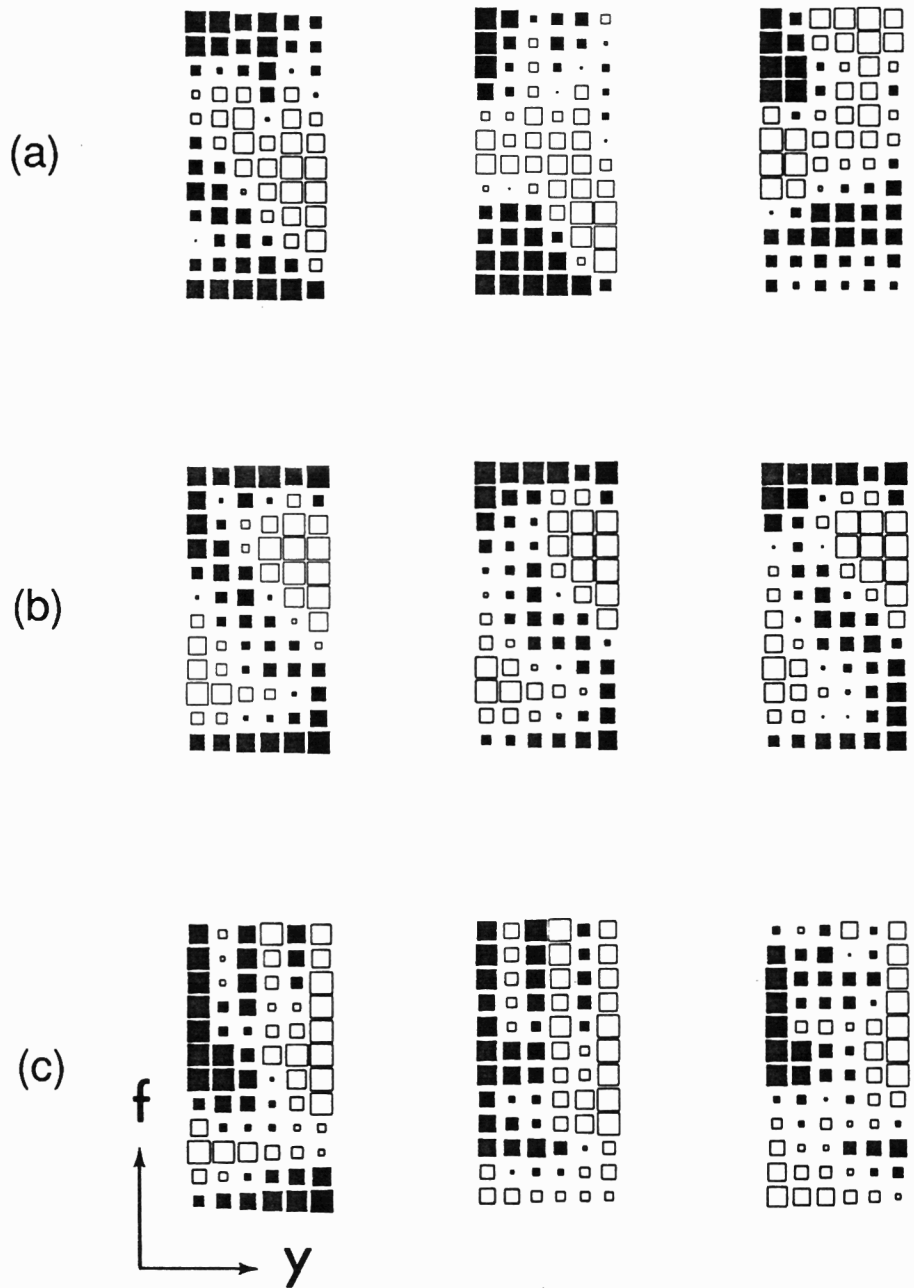


FIGURE 7