Designing An Orientation Finding Algorithm Using Human Visual Data

by

Mojgan Monika Gorkani

Bachelor of Applied Science
Electrical Engineering
University of Toronto, 1991

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1993

Signature of Author .....
Program in Media Arts and Sciences,
School of Architecture and Planning,
May 17, 1993

Certified by .....
Rosalind W. Picard
Assistant Professor
Perceptual Computing Group in the Media Lab
Thesis Supervisor

Accepted by .....
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students

Designing An Orientation Finding Algorithm Using Human Visual Data

by

Mojgan Monika Gorkani

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 17, 1993, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

An algorithm for detecting orientation in textures is developed and compared to results of humans detecting orientation in the same textures. The algorithm is based on the steerable filters of Freeman and Adelson (1991), orientation-selective filters derived from derivatives of Gaussians. The filters are applied over multiple scales and their outputs nonlinearly contrast normalized. The human data was collected from forty subjects who were asked to identify "the minimum number of dominant orientations" they perceived, and the "strength" with which they perceived each orientation. Test data consisted of 111 graylevel images of natural textures taken from the Brodatz album, a standard collection used in computer vision and image processing. Results on comparing the computer and humans' test data indicate they each chose at least one of the same dominant orientations on 95 of the natural textures. Of these textures, 74 were also in 100% agreement on the location of all the dominant orientations chosen by both human and computer. Individual cases which disagree are analyzed, and possible causes are discussed. Some apparent limitations in the current filter shapes and sizes are illustrated, as well as some effects possibly due to semantic influences and gestalt grouping.

Thesis Supervisor: Rosalind W. Picard
Title: Assistant Professor
Perceptual Computing Group in the Media Lab

Designing An Orientation Finding Algorithm Based On Human Visual Data

by

Mojgan M. Gorkani

Thesis for the degree of Master of Science

at the

Massachusetts Institute of Technology

May 1993

Thesis Reader .....

William Freeman
Research Scientist
Mitsubishi Electric Research Laboratories, Inc.

Thesis Reader .....

Josef Bigün
Research Associate
Signal Processing Laboratory of the Swiss Federal Institute of Technology in Laussane

Contents

Acknowledgments

First of all, I gratefully acknowledge the support of my advisor Roz Picard. She has always been encouraging and very patient with me (especially when I barge into her office and start talking non-stop about a problem). She is a great role model and someone that I have looked up to and will continue to do so.

Bill Freeman, my reader, has helped me tremendously with his steerable filter code. His help has been invaluable to me. Josef Bigün was kind enough to be my reader and over long distance gave me great insight in the complicated task of finding orientations.

Oliver Yip helped me in developing and conducting the human visual test. Brad Horowitz offered his X-window expertise to develop the interface for the visual test. Ted Adelson gave me helpful comments about the human visual test and his four beautiful pictures which I have used in my introduction are greatly appreciated. I would like to thank Lee Campbell and Stephen Intille for their patient reading of my thesis and their helpful remarks. Irfan Essa, Stan Sclaroff and Ali Azarbayejani offered me their Latex and graphical tools expertise. Eero Simoncelli over the ethernet helped me with some theoretical questions. The Texture Heads : Fang Liu, Chris Perry, Kris Popat, Steve Mann, and Alex Sherstinsky gave me support and encouragement. Jeff Bilmes forced me to lift weights so that I would be strong enough to finish this thesis. John Wang helped me a lot with the Obvius software.

I would like to also thank all the subjects who took my visual test. Without their data, this thesis would not exist.

Thanks to my fellow officemates Sourabh Niyogi and George Chou for their stimulating conversations. And to the rest of Percom members including Bea Bailey, Laureen Fletcher and Judy Bornstein who one way or another helped me to finish this thesis. Thanks to Linda Peterson for being so understanding and giving me lots of encouragement.

Thanks to my patient and understanding roomate, Debbie Becker. Others who supported me along the way: Gilbert Chiang, Ilya Lyubomirsky, Eric Ly, Raj Gupta, Judy Chen, Wenjie Hu, Mary Beshai, Siegfried Fleischer, Grace Foster and many more.

Most importantly, I would like to thank my parents Mahmoud and Fakhry and my brothers Payman and Arman. My brother Arman has been my major source of support. It has been great living in the same city with him!

This research was made possible by the sponsorship of BT (British Telecom).

Chapter 1

Introduction

Imagine the future when vast amounts of image data such as movies, photo albums, journals and other visual collections will be available in databases and can be accessed online. Accessing this information fast and efficiently is a big challenge and offers a rich, new research area. BT (British Telecom) is funding a research project at the MIT Media Lab to develop tools which can search for a picture just like one can search for a keyword or a text description in a database. One major problem with image databases is that a short description attached to an image will not be sufficient for all queries. A more likely scenario is where a user shows a computer a pre-existing pattern and asks the computer to search for patterns similar to it. It is important to find visual features which describe the visual content in an image and can be used to judge the similarity between images.

One such visual feature is the dominant orientations of an image. For example, in Figure 1-1, the number of dominant orientations in each image is sufficient to discriminate the four textured images. There is psychophysical evidence that humans use orientation as a cue for discriminating textures [1][2][3]. Further, results from physiological experiments suggest the

existence of orientation selective mechanisms in the human visual system [4][5].

Extraction of orientation over a large scale can be used for rotating images to align before comparing them pixel by pixel - a process possibly performed by humans [6]. Shape and perspective information can also be obtained from relative orientation [7, 8].

As orientation is one of the most important features for human attention and for texture matching, it is important that algorithms which extract orientation be developed. One of the key difficulties in extracting orientation information is that it has to be gathered over multiple scales [9][10]. Also, orientation estimation is complicated by effects such as contrast, gestalt organization, and the semantic meaning associated with a pattern.

The research in this thesis focuses on the interaction of orientation estimation with scale and contrast. In particular, an algorithm is developed to find dominant orientations in textured images combining the orientation information over four different scales and using contrast normalization. To evaluate the results of this algorithm in detecting dominant orientations, a human visual test was conducted. The human data was collected from forty subjects who were asked to identify "the minimum number of dominant orientations" they perceived, and the "strength" with which they perceived each orientation. Test data consisted of 111 gray level images of natural textures taken from the Brodatz album [11], a standard collection used in computer vision and image processing. The positions and the numbers of dominant orientation found by the algorithm were compared with the human data. Some of the parameters in the algorithm were then varied to obtain better agreement between the human data and the algorithm.

Chapter 2 gives background information on orientation estimation techniques and compares several popular methods for finding orientations. Chapter 3 outlines problems in estimating

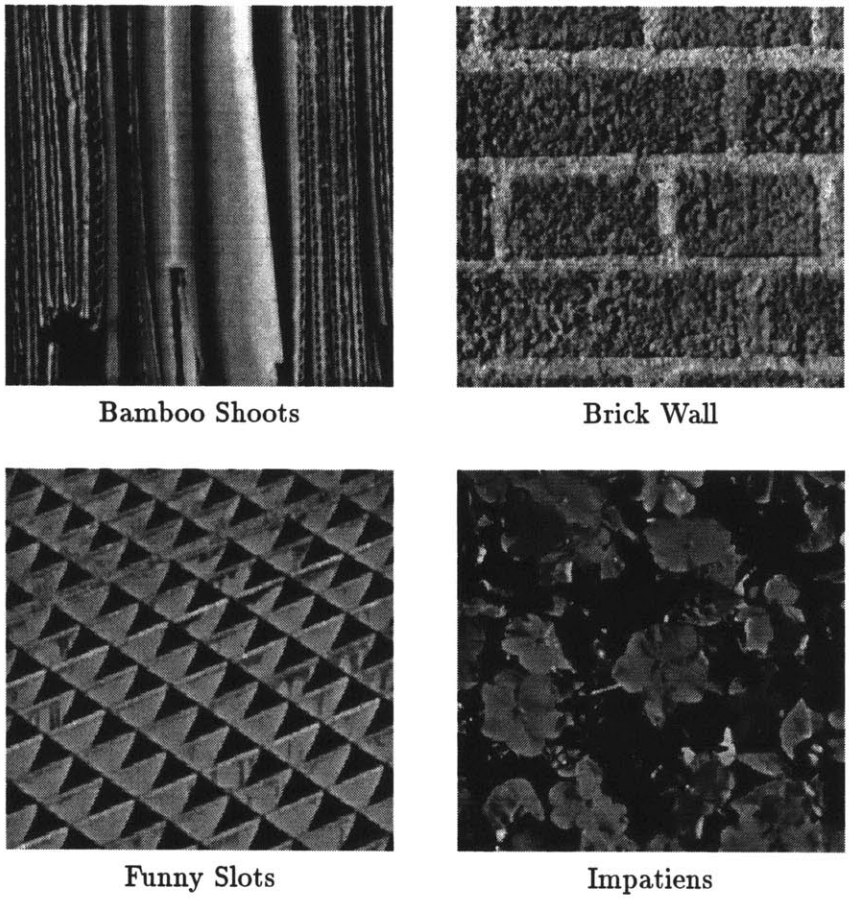


Figure 1-1: Four textured images (Printed with permission from E. H. Adelson).

orientations at different scales and reviews several multi-scale techniques for finding local orientations. Chapter 4 outlines the orientation finding algorithm used in this thesis for finding dominant orientations in textured images. Chapters 5 and 6 describe the human visual test and outline the comparison done between the human visual data and the computer data respectively. Conclusions and further work is discussed in Chapter 7.

Chapter 2

Background

An oriented pattern's Fourier transform tends to cluster along lines in the Fourier domain perpendicular to the orientations in the spatial domain. This clustering can be detected by summing energy in the appropriate region of the power spectrum and examining how the sum is affected by rotations. Most of the methods proposed for calculating main orientations are in essence finding the orientation along which this clustering occurs. In this chapter, I first talk about methods which operate directly in Fourier domain, then those based on filters in the spatial domain, and finally some other methods of finding local orientation.

2.1 Estimating Orientation In the Fourier Domain

Orientation can be estimated directly in the Fourier domain. Bajcsy [12] showed in her paper that directionality can be measured by calculating the histogram of the Fourier power along lines through the origin (see Figure 2-1). If $F(w_1, w_2)$ is the Fourier transform of image I, then the Fourier power spectrum $P(w_1, w_2)$ approximately equals $|F(w_1, w_2)|^2$. Bajcsy transformed

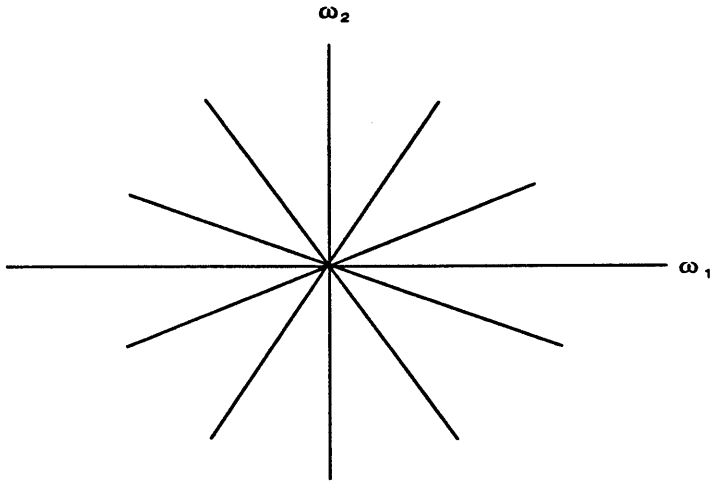


Figure 2-1: The power spectrum is summed along each of these lines

---

$P(w_1, w_2)$ from cartesian coordinates to $P(r, \phi)$ in polar coordinates, where for each frequency r, $p_r(\phi)$ is a one-dimensional function. She determined the main orientations by looking at the peaks of the histogram of directional power $P(\phi)$. $P(\phi)$ is calculated for quantized values of ϕ, i.e., $\phi = 0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \ldots$ by summing the radial power $p_r(\phi)$,

$$P(\phi) = \sum_{r=1}^{\frac{W_r}{2}} p_r(\phi) \tag{2.1}$$

where $\frac{W_r}{2}$ is the window size. Bajcsy applied this method on several textures of size 32 x 32 where the quantized interval for ϕ was $\frac{\pi}{16}$ and $W_r = 16$. Bajcsy does not explain why this particular quantized interval for ϕ was used.

This method has several drawbacks. One problem involves the inherent difference between continuous and finite discrete Fourier transforms. Transforming from cartesian coordinates to

polar coordinates is tricky since it involves interpolation and the results are inaccurate near the origin. Bajcsy first high-pass filtered the images since the definition of angle is very coarse at low-frequency points due to the smaller number of samples available on the rectangular grid. Also low-frequency components are affected more by illumination changes than surface coloration [13]. However, false peaks were still found in the histogram of $P(\phi)$ due to smooth changes in shading which are perceptually discounted by humans [12]. Bajcsy argued that more information about the scene (its illumination, continuity and context) is needed to remove this kind of slow change. If the oriented energy falls between two bins, the energy contribution from each bin will be small and will not be prominent in the histogram.

Chaudhuri, et. al. [14], extract orientation information by using sector filters on the Fourier transform of the image instead of summing the oriented power in the polar coordinates as Bajcsy. The filter rejects all the Fourier components outside the desired angle band. An ideal sector filter causes ringing and smearing artifacts. To overcome such affects, they multiplied the sector filter with a raised cosine window $W_1(\varphi)$:

$$\begin{aligned} W_1(\varphi) &= cos^{\alpha}(\frac{\varphi - \varphi_m}{B}\pi) \qquad \varphi \in [\varphi_1, \varphi_2] \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{2.2}$$

where φ is defined as,

$$\varphi = \arctan(\frac{w_2}{w_1}) \tag{2.3}$$

and φ_1 and φ_2 are the angular limits of angle band, $\varphi_m = \frac{\varphi_1+\varphi_2}{2}$ is the center of the desired

angle band and $B = \varphi_2 - \varphi_1$ is the chosen angular bandwidth. The value α is chosen to make sure that the window coefficients are not too low to miss the components near the two limits of the angle band.

Before applying the sector filter, the computed Fourier transform of the image is bandpass filtered which removes the very high and low frequency components to reduce the effects of aliasing, noise, illumination effects and having fewer sampling points near the origin. The cutoff frequency to remove the low frequency components depends on the nature of the input image. Depending upon the scale of a local structure, the cutoff frequency might suppress the contribution of the Fourier components of these structures. Chaudhuri, et. al. ran the sector filters with different φ_1 and φ_2 on images with structures at different orientations and inverse Fourier transformed the resulting filtered images. Ideally the inverse Fourier transformed images should contain the structures in the original image with orientations corresponding to the angular band of the filters. The distribution of orientations corresponding to the structures was found by counting the pixels of the binarized inverse Fourier transformed output images corresponding to the different angle bands. One problem they found was that for structures of finite width in the original image, the components in the frequency domain do not lie completely within the defined sector. As the thickness of the structures increases, there is an increased spreading of the Fourier components, leading to a reduced angular sensitivity. This results in poor contrast in the filtered components and certain artifacts.

2.2 Filtering: Trade-offs And Desirable Properties

Another way to detect the direction along which the clustering in the Fourier domain occurs is by filtering in the spatial domain. Filters in the spatial domain can be used to find local orientation information more easily than operating directly in the Fourier domain. Also there is psychophysical and physiological evidence that some of the primate receptive fields can be modeled by linear filters [15]. Filters can be chosen which select only the frequency components in a particular direction. The local orientation in an image can be found by running a set of directionally sensitive filters. If there is one main orientation then there will be a large response in the filters whose direction coincides most closely with the main orientation and the responses of all the other directional filters will be small. If a large fraction of filters give a comparable response then the neighborhood has structures covering all orientations.

The selection of the set of directional filters involves certain trade-offs. As mentioned before very low spatial frequency components should not be used for the orientation estimation since they are affected more strongly by illumination effects than surface coloration. Further, the definition of angle is very coarse at low-frequency points due to the smaller number of samples available on the rectangular grid [14]. On the other hand, high spatial frequencies are sensitive to noise and aliasing so they too are inappropriate for estimating orientation. To measure orientation some kind of bandpass filtering is preferable. Another trade-off occurs between angular and spatial resolution. If the filter is too orientation-specific then a large spatial neighborhood is needed to make reliable measurement of energy [13]. Conversely, if the filter responds over a wide range of orientations then it will be difficult to localize the orientation accurately. It makes sense then to choose filters which exhibit good local properties both in spatial and fre-

quency domains. Since Gaussians are optimally localized in space and frequency, derivatives of Gaussians are often used as directional filters.

2.2.1 Number of Filters Required for Orientation Estimation

It seems initially that many filters at different orientations are needed to estimate the local orientation. It would be desirable to know how many filters are needed for an accurate measurement of orientation.

If a directional filter can be rotated to any orientation by a finite number of basis filters then the basis filters can be used to estimate the local orientation. If a two-dimensional filter in the spatial domain $f(x, y)$ with polar representation $f(r, \phi)$ has a finite polar Fourier series expansion then:

$$f(r, \phi) = \sum_{n=-N}^{N} a_n(r) \exp(in\phi). \tag{2.4}$$

where

$$r = \sqrt{x^2 + y^2} \tag{2.5}$$

$$\phi = \arctan(\frac{y}{x}) \tag{2.6}$$

and $i = \sqrt{-1}$, then filter f rotated through an arbitrary orientation θ can be expressed as:

$$f^{\theta}(r, \phi) = \sum_{n=-N}^{N} a_n(r) \exp(in\phi) \exp(-in\theta). \tag{2.7}$$

In this expression the terms $a_n(r) \exp(in\phi)$ can be viewed as the basis filters and the terms

$\exp(-in\theta)$ as the coefficients needed to linearly combine them to synthesize filter g at any orientation [16]. The number of basis filters needed is determined by the number of positive and negative frequencies present in the polar frequency expression shown in (2.7). If all the $a_n(r)$ are non-zero for all $-N \leq n \leq N$ then $2N+1$ basis filters have to be used.

The basis filters can also be rotated versions of the filter. Freeman and Adelson in [16] derive the conditions needed to synthesize a filter f rotated to angle θ as a linear combination of rotated versions of f:

$$f^{\theta}(x,y) = \sum_{l=1}^{K} k_l(\theta) f^{\alpha_l}(x,y). \tag{2.8}$$

Filter f is defined to be a steerable filter if it satisfies (2.8) for a finite K. To determine the interpolation functions $k_l(\theta)$ and the number of basis filters needed, K, both sides of (2.8) are substituted by their polar Fourier series expansion given in (2.7), and both sides of the resulting equation are projected on the complex exponential $\exp(im\theta)$:

$$a_m(r)\exp(im\theta) = \sum_{l=1}^{K} k_l(\theta) a_m(r)\exp(im\alpha_l), \quad -N \leq m \leq N \tag{2.9}$$

In (2.9) since $-N \leq m \leq N$, there will be $2N+1$ linear equations. The equations can be stated in the following format:

$$\begin{pmatrix} 1 \\ \exp(i\theta) \\ \vdots \\ \exp(iN\theta) \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ \exp(i\alpha_1) & \dots & \exp(i\alpha_K) \\ \vdots & \vdots & \vdots \\ \exp(iN\alpha_1) & \dots & \exp(iN\alpha_K) \end{pmatrix} \begin{pmatrix} k_1(\theta) \\ k_2(\theta) \\ \vdots \\ k_K(\theta) \end{pmatrix} \tag{2.10}$$

where only the expressions for positive m are shown. To calculate the coefficients k_l, however, expressions for positive and negative m have to be used. If $a_m(r) = 0$ for any m, then the corresponding (mth) row of the left hand side and of the matrix of the right hand side of (2.10) should be removed. Freeman and Adelson show that if T is the number of positive or negative frequencies where $a_m(r)$ is non-zero for $-N \leq m \leq N$, then the required number of basis filters K has to at least equal T. The T basis function orientations α_l must be chosen so that the columns in the matrix in (2.10) are linearly independent. One way to ensure this is to choose these angles to be equally spaced between 0 and π. An observation is that the interpolation functions $k_l(\theta)$ are independent of the values of the non-zero coefficients $a_m(r)$; therefore, these coefficients are not necessarily unique for a particular filter. This is true since the interpolation is only dependent on the angular tuning of the filters.

Freeman and Adelson also show in [16] that for filters of the form $f(x, y) = W(r)P_N(x, y)$, where $W(r)$ is an arbitrary windowing function and $P_N(x, y)$ is an Nth order polynomial in x and y whose coefficients may depend on r, a linear combination of $2N + 1$ basis functions are sufficient to rotate filter f to any orientation. For an odd or even order $P_N(x, y)$, only $N + 1$ basis filters are shown to be needed to steer a filter. They also derive the interpolation functions $k_l(\theta)$ for a set of separable basis functions $W(r)R_l(x)S_l(y)$ so that the rotated filter f^θ can be expressed as a linear combination of separable basis filters:

$$f^\theta(x, y) = W(r) \sum_{l=0}^{K} k_l(\theta) R_l(x) S_l(y) \tag{2.11}$$

If a filter does not have a finite polar Fourier series representation then it can not be exactly steered by a finite number of basis filters. One solution is to approximate a kernel with an

adequate higher order polynomial multiplied by a radially symmetric function. To approximate the filter response from a limited number of basis filters, certain methods can be used to minimize the error in the synthesized filter as suggested by Perona in [17].

2.2.2 Quadrature Filters

Another requirement on the choice of the set of directional filters has to do with the phase dependency of the orientation estimate. If the directional filter is even about the axis of its main orientation then it will give a zero response along this axis when convolved with a signal which is odd along that axis. For example, a second derivative of a Gaussian filter outputs zeros along an edge oriented the same direction as the filter. Similarly, a directional filter which is odd about the axis of its main orientation such as the first derivative of a Gaussian will give a zero response along this axis when convolved with a signal which is even along that axis.

This means that even when the direction of the filter coincides with the local orientation of the image, the output of the filter can be zero. Two filters are said to be in quadrature if they have the same frequency response but differ in phase by 90^o. In other words they are Hilbert transforms of each other [18]. Where one filter response shows zero crossings, the other one will show extremes. Let $E(\theta)$ be the sum of the squared outputs of an oriented filter f^θ and its Hilbert transform h^θ. $E(\theta)$ is an estimate of the spectral power of the image along the direction θ:

$$E(\theta) = (f^\theta * I)^2 + (h^\theta * I)^2 \tag{2.12}$$

where I represents the image. $E(\theta)$ will be independent of the local phase of the input [19].

When the direction of filter f coincides with the local orientation of the input, this energy will be maximal even if the input's phase and the phase of filter f differ by 90°.

If we have a filter f then its Hilbert transform h along the main orientation of filter f can be described in the following way:

$$H(w_1, w_2) = i\ sign[\cos(\varphi - \phi_k)]F(w_1, w_2) \tag{2.13}$$

where $F(w_1, w_2)$ and $H(w_1, w_2)$ are the Fourier transforms of f and h respectively, and ϕ_k is the main direction of filter f [19].

The fact that a filter can be steered by a linear combination of basis filters does not mean that its Hilbert transform is also steerable. For example, the Hilbert transform of a second derivative of a Gaussian is not steerable by a finite number of basis filters. However, an approximation to the Hilbert transform can be found which is steerable. Freeman and Adelson in [16] formed an approximation to the Hilbert transform of a second derivative of a Gaussian by finding the least squares fit to a polynomial times a Gaussian which is steerable.

2.2.3 Vectorial Representation of Filter Responses

The responses of the directional filters can be represented as orientation vectors: the magnitude of the vector corresponds to the magnitude of the filter response, while the direction is given by the filter orientation doubled [19]. The local orientation can be computed by the vector addition of these orientation vectors. If there is a neighborhood containing structures in all directions, then the vector sum of the filter responses should vanish since there is no one dominant orientation. This does not happen if the direction of the orientation vectors corresponds

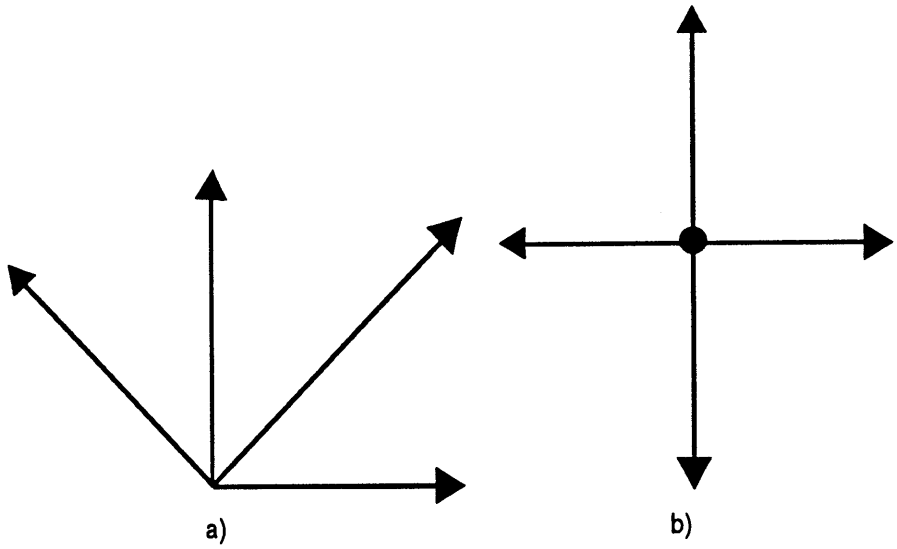


Figure 2-2: Computation of local orientation by vector addition of the four filter responses. Shown is an example where the neighborhood is isotropic with respect to orientation: all filter responses are equal. a) The angles of the vectors are equal to the filter directions. b) The orientations shown are double those in (a).

---

to the filter directions since the filters are usually equally spaced between 0 and π and there is an equal contribution from each response. For example, in the case of the four directional filters, the sum will only vanish for this neighborhood if the angles of the orientation vectors are double the filter orientations as shown in Figure 2-2. This makes sense since by doubling the angles of the filters, the orientation vectors will be uniformly distributed between 0 and 2π and will then cancel out if they have equal magnitudes.

In [19] Knutsson and Granland suggest directional filters f_k which are polar-separable in the Fourier domain:

$$F_k(\rho, \varphi) = V_1(\rho)V_2(\varphi) \tag{2.14}$$

where φ is defined in (2.3) and ρ is defined:

$$\rho = \sqrt{w_1^2 + w_2^2} \tag{2.15}$$

$F_k(\rho, \varphi)$ is the polar representation of the Fourier transform of f_k, $F_k(w_1, w_2)$. $V_1(\rho)$ and $V_2(\varphi)$ are functions of ρ and φ respectively. $V_2(\varphi)$ for the f_k are defined as:

$$V_2(\varphi) = \cos^{2p}(\varphi - \phi_k) \tag{2.16}$$

where ϕ_k is the main orientation of filter f_k and the value p determines the angular resolution of the filter. If there are K directional filters, the main orientations ϕ_k are equally distributed between 0 and π where $\phi_k = \frac{k\pi}{K}$ where $k = 0, \ldots, K-1$. For a local neighborhood ideally oriented in an arbitrary orientation θ_o (the Fourier transform is concentrated on a line perpendicular to θ_o), the response of each filter can be expressed as an orientation vector. Since the $V_1(\rho)$ term is the same for all the filters f_k, only the angular term of the response of the filters are considered in the expression:

$$\hat{f}_k(\theta_o) = \exp(i2\pi k/K)V_2(\theta_o) \tag{2.17}$$

where $\hat{f}_k$ is the vector representation of the output of f_k. Note the factor 2 in the complex exponential results from doubling the filter orientation since the direction of the vector is double the filter orientation.

Knutsson and Granland show in [19] that the sum of these orientation vectors result in a

vector whose angle is double the local orientation:

$$\sum_{k=0}^{K-1} \hat{f}_k(\theta_o) = \frac{K}{2^{2p}} \begin{pmatrix} 2p \\ p+1 \end{pmatrix} \exp(i2\theta_o) \tag{2.18}$$

This is true provided that $p > 0$ and $K > p + 1$.

Note that the condition for (2.18) to be true is very similar to the criterion for the number of basis filters needed to synthesize a filter at any orientation. Recall that the number of basis filters needed corresponds to the number of frequencies for which the $a_n(r)$ coefficients of the Fourier polar series representation are non-zero as shown in Section 2.2.1. The polar component $V_2(\varphi)$ can be represented as a Fourier polar series representation and there will be $p+2$ non-zero Fourier polar coefficients.

2.2.4 Examples of Directional Filtering Methods Used for Orientation Estimation

A number of methods have been suggested to measure the local orientations in textures using directional filters. The choice of the directional filter is difficult because it is hard to satisfy all the requirements discussed in the previous section. Also there is a desire to choose filters which yield frequency and orientation bandwidths which are in reasonable agreement with the physiological and psychophysical estimates of the early human visual processes. The following examples are representative of some of the methods currently used and do not cover all the possible ways of measuring orientation. It should be mentioned beforehand that the local orientation θ_o estimated in the methods below is perpendicular to the perceived orientation.

Kass and Witkin's method [13]

Kass and Witkin find the main directionalities in a texture by examining the energy in the output of an orientation-selective linear filter f_{kw}. Their choice of the orientation-selective filter is:

$$\begin{aligned} f_{kw}^{\theta}(x,y) &= (\cos\theta, \sin\theta)\cdot\nabla q \\ &= \cos\theta(x(\sigma_1^{-4}\exp(\frac{-(x^2+y^2)}{2\sigma_1^2}) - \sigma_2^{-4}\exp(\frac{-(x^2+y^2)}{2\sigma_2^2})) \\ &\quad + \sin\theta(y(\sigma_1^{-4}\exp(\frac{-(x^2+y^2)}{2\sigma_1^2}) - \sigma_2^{-4}\exp(\frac{-(x^2+y^2)}{2\sigma_2^2})) \end{aligned} \tag{2.19}$$

where σ_1 and σ_2 are the variances of isotropic Gaussian filters. This filter is bandpass and its frequency tuning is identical to the first derivative of a Gaussian. The filter f_{kw}^{θ} is also steerable with the partial derivatives of q as basis filters. The oriented energy, $V(\theta)$ is calculated in the following way:

$$\begin{aligned} V(\theta) &= m_g \star (f_{kw}^{\theta} \star I)^2 \\ V(\theta) &= m_g \star (cos(\theta)C_x + sin(\theta)C_y)^2 \end{aligned} \tag{2.20}$$

where $m_g(x,y)$ is Gaussian weighting function, I is the image, and C_x and C_y are the outputs of the image convolved with the basis filters. This oriented energy is very similar to $E(\theta)$ shown in (2.12) except that the Hilbert transform of the oriented filter was not calculated by Witkin and Kass. The weighting function m_g is used to average the results in order to reduce the noise in the calculation of the energy. Averaging the squared output of the directional filter response also makes the energy $V(\theta)$ independent of the input phase. However, using the quadrature

method discussed in Section 2.2.2 provides a better spatial resolution since averaging with m_g will blur the oriented energy $V(\theta)$.

If (2.20) is expanded, $V(\theta)$ will be in terms of $\cos 2\theta$ and $\sin 2\theta$; therefore, it will be periodic with a period of π and will have two extrema π apart. Also, since $V(\theta)$ is in terms of $\cos(2\theta)$ and $\sin(2\theta)$, the orientation found from analyzing $V(\theta)$ will be double the desired local orientation θ_o. The orientation θ_o is found by calculating the peak of $V(\theta)$ in the following way:

$$\begin{aligned} \theta_o &= \frac{\arctan(\frac{\int_0^\pi V(\theta)\sin(\theta)d\theta}{\int_0^\pi V(\theta)\cos(\theta)d\theta})}{2} \\ &= \frac{\arctan(\frac{m_g * 2C_xC_y}{m_g*(C_x^2-C_y^2)})}{2}. \end{aligned} \tag{2.21}$$

If complex variable P_{kw} is defined to be

$$P_{kw} = (C_x + iC_y)^2 \tag{2.22}$$

then angle θ_o and the measure of reliability of the orientation (also frequently called "coherence") χ_{kw}, can be expressed in terms of P_{kw}:

$$2\theta_o = arg(P_{kw} * m_g) \tag{2.23}$$

$$\chi_{kw} = \frac{|P_{kw} * m_g|}{(|P_{kw}| * m_g)}. \tag{2.24}$$

As can be seen, P_{kw} is a combination of the basis filter outputs and can be regarded as an orientation vector where $2\theta_o$ would be the direction of an average of these vectors in a local neighborhood. The measure of reliability for the orientation calculated for pixel position (x, y)

is reflected in how well the P_{kw} vectors are lined up in a local region around position (x, y).

Bigün and Granlund's method [20]

Even though this method initially does not appear to use directional filtering, its implementation in the spatial domain results in using directional filters. Bigün and Granlund define an image to be "linear symmetric" if it is made up of a set of parallel lines. The Fourier transform of a linear symmetric image is shown to be exactly concentrated to a line through the origin. Of course, most real-life textures are not linear symmetric. However, if they have a strong orientation their Fourier transforms should be spread about a line through the origin. Bigün and Granlund find the main orientation in the least-squares sense by fitting a line through the Fourier transform of the texture. The optimally fitted line in LSE sense represents the axis of least inertia of the Fourier transform and can be found by finding the smallest eigenvalue of the 2×2 inertia matrix J,

$$J = \begin{bmatrix} \int_{E_2} \int w_1^2 |F(w_1, w_2)|^2 dE_2 & -\int_{E_2} \int w_1 w_2 |F(w_1, w_2)|^2 dE_2 \\ -\int_{E_2} \int w_1 w_2 |F(w_1, w_2)|^2 dE_2 & \int_{E_2} \int w_1^2 |F(w_1, w_2)|^2 dE_2 \end{bmatrix}, \tag{2.25}$$

where $F(w_1, w_2)$ is the Fourier transform of image I and E_2 is the Euclidean space with $dE_2 = dw_1 dw_2$. The magnitude of the Fourier transform is used so that the orientation estimation is independent of the input Fourier phase.

J can be calculated in the spatial domain for a local neighborhood and its approximate discrete representation consists of:

$$J_{11} = \frac{1}{4\pi^2} (\frac{\delta I}{\delta y})^2 * m_g$$

$$J_{22} = \frac{1}{4\pi^2}(\frac{\delta I}{\delta x})^2 * m_g$$
$$J_{12} = J_{21} = -\frac{1}{4\pi^2}\frac{\delta I}{\delta x}\frac{\delta I}{\delta y} * m_g. \quad (2.26)$$

where $\frac{\delta I}{\delta y}$, $\frac{\delta I}{\delta x}$ are the discrete Gaussian derivatives of the image I , m_g is a Gaussian filter and J_{mn} is the (m,n)th component of J.

The optimal line $\bar{s}$ is the orthonormal eigenvector corresponding to the smallest eigenvalue of J called λ_o. The other eigenvalue λ_1 corresponds to the line perpendicular to $\bar{s}$. The main direction represented by angle θ_o is calculated to be

$$\tan(2\theta_o) = \frac{2J_{12}}{J_{11} - J_{22}} \quad (2.27)$$

where J_{mn} is the (m,n)th component of J. Two measures of the reliability of the estimation of the orientation, χ_{bg1} and χ_{bg2}, are calculated :

$$\chi_{bg1} = (\frac{\lambda_1 - \lambda_o}{\lambda_1 + \lambda_o})^c \quad (2.28)$$
$$\chi_{bg2} = \lambda_1 - \lambda_o \quad (2.29)$$

where λ_1 and λ_o are the two eigenvalues for J (2×2) and c is a constant used to control the dynamic range of χ_{bg1} so that it is less dependent on the contrast of the image. Bigün and Granland do not discuss how c should be chosen for different images. The values $c = 1$ and $c = 6$ were used to illustrate how the value of c affects the coherence measure.

If the eigenvalues have a multiplicity greater than 1, i.e., $\lambda_o = \lambda_1$ then there is no one

optimal fitted line. This occurs if the energy in the Fourier domain is distributed in such way that there are many lines which give the same least square error. (See Figure 2-2(b)). In this case the values for χ_{bg1} and χ_{bg2} will be close to zero since the difference between the eigenvalues is close to zero. The least eigenvalue λ_o will be zero and will have a multiplicity of one if and only if image I is linear symmetric. In this case χ_{bg1} will ideally be equal to 1 and χ_{bg2} will be equal to the value of the biggest eigenvalue, λ_1. If the value for λ_1 is large then χ_{bg2} will be large. A high χ_{bg2} indicates that the orientation found is reliable which makes sense since λ_o must be very small compared to λ_1. If both eigenvalues are zero then the image is constant (has no main orientations). Combining (2.27), (2.28), (2.26) and using the variable P_{bg} defined as:

$$P_{bg} = (\frac{\delta I}{\delta x} + i\frac{\delta I}{\delta y})^2, \tag{2.30}$$

the values for θ_o and χ_{bg1} and χ_{bg2} can be expressed in the following way:

$$2\theta_o = arg(P_{bg} * m_g) \tag{2.31}$$

$$\chi_{bg1} = (\frac{|P_{bg} * m_g|}{(|P_{bg}| * m_g)})^c \tag{2.32}$$

$$\chi_{bg2} = \frac{1}{4\pi^2}|P_{bg} * m_g|. \tag{2.33}$$

For each point (x, y) in image I, the local orientation θ_o and the measures of reliability χ_{bg1} and χ_{bg2} can be calculated.

Rao and Schunck's Method [21]

Rao and Schnunk used a directional filter, f_{rs} which is a first derivative of a Gaussian. The first derivative of a Gaussian is a steerable filter:

$$f_{rs}^{\theta} = \cos(\theta) g_1^{0^o} + \sin(\theta) g_1^{90^o} \tag{2.34}$$

where $g_1^{0^o}$ and $g_1^{90^o}$ are the first derivative of Gaussians oriented at 0^o and 90^o. Since f_{rs} is a first derivative of a Gaussian, convolving image I with $g_1^{0^o}$ and $g_1^{90^o}$ is the same as taking the gradient of the image in the x and y direction:

$$\begin{aligned} I * g_1^{0^o} &= \frac{\delta I}{\delta x} \\ I * g_1^{90^o} &= \frac{\delta I}{\delta y} \end{aligned} \tag{2.35}$$

where $\frac{\delta I}{\delta x}$ and $\frac{\delta I}{\delta y}$ are discrete Gaussian derivatives of I.

The local orientation θ_o is then calculated in the following way:

$$2\theta_o = \arg(P_{bg} * m_b) \tag{2.36}$$

where P_{bg} is defined in (2.30) and m_b is a box filter. The orientation measure is analogous to Bigün and Granland's method (2.31) except that a box filter is used instead of a Gaussian filter. The same result as (2.36) can be obtained by minimizing the oriented energy $E(\theta)$ with respect to θ in equation 2.12 with $f^{\theta} = f_{rs}^{\theta}$. In this case, a Hilbert transform is not calculated for f_{rs}^{θ} and $E(\theta)$ is smoothed with a box filter analogous to Kass and Witkin's and Bigün and

Granland's smoothing with the Gaussian filter m_g.

The measure of the reliability or coherence measure of the orientation estimated for position (x_o, y_o) χ_{rs} is:

$$\chi_{rs} = R(x_o, y_o)\frac{\sum_{(i,j)\in L} \|R(x_i, y_i)\cos((\theta_o(x_o, y_o) - \theta_o(x_i, y_i))\|}{\sum_{(i,j)\in L} G(x_i, y_i)} \tag{2.37}$$

where $R(x_o, y_o)$ is the magnitude of the gradient at position (x_o, y_o) , L is the local region where this measurement is calculated and $\theta_o(x_i, y_i)$ is the orientation found at position (x_i, y_i). The coherence χ_{rs} is a measure of how well the gradient vectors line up with respect to the gradient vector at position (x_o, y_o). The $R(x_o, y_o)$ component is used as a weight to make sure that χ_{rs} is high at points where there is high visual contrast. It is not clear how the size of the local region L is chosen.

Freeman and Adelson's method [16]

Freeman and Adelson use the second derivative of a Gaussian to estimate the local orientation. The second derivative of a Gaussian, g_2 is then defined as:

$$g_2(x, y) \quad = 0.92132(2x^2 - 1)\exp(-(x^2 + y^2)) \tag{2.38}$$

and is normalized so that the square of its integral equals to one over all space. The corresponding representation is:

$$g_2(r, \phi) = 0.92132r^2\exp(-r^2)\cos(2\phi) + (r^2 - 1)\exp(-r^2) \tag{2.39}$$

where r and ϕ are described by (2.4). The expression in (2.39) can then be represented as the Fourier polar series of (2.7) with:

$$a_o = 0.92132(r^2 - 1)\exp(-r^2)$$
$$a_2 = a_{-2} = 0.46066r^2\exp(-r^2)\,. \tag{2.40}$$

Since there are three non-zero coefficients $a_n(r)$, three basis filters are needed to steer g_2 as discussed in Section 2.2.1. For computational efficiency it is desirable to use x-y separable basis filters rather than rotated versions of g_2. Freeman and Adelson in [16] show that there exist three x-y separable filters whose linear combination can steer g_2 to any orientation. The x-y separable basis filters for g_2 will be denoted g_{20}, g_{21} and g_{22}:

$$\begin{aligned} g_{20} &= 0.9213(2x^2 - 1)\exp(-(x^2 + y^2)) \\ g_{21} &= 1.843xy\exp(-(x^2 + y^2)) \\ g_{22} &= 0.9213(2y^2 - 1)\exp(-(x^2 + y^2)) \end{aligned} \tag{2.41}$$

As mentioned in Section 2.2.2, the Hilbert transform of g_2 is not steerable. Freeman and Adelson approximate this Hilbert transform with a least squares fit to a polynomial times a Gaussian which is steerable. They found an approximation, h_2, using a 3rd order, odd parity polynomial:

$$h_2 = (-2.205x + 0.9780x^3)\exp(-(x^2 + y^2))\,. \tag{2.42}$$

The x-y separable basis filters for h_2 will be denoted h_{20}, h_{21}, h_{22} and h_{23}.

To find the main orientation in a local neighborhood, the oriented energy described in (2.12) will be calculated for image I. Since g_2^θ and h_2^θ can be expressed as a linear combination of their basis functions, $E(\theta)$ can be expressed in the following way:

$$\begin{aligned} E(\theta) &= (g_2^\theta * I)^2 + (h_2^\theta * I)^2 \\ &= ((\sum_{l=0}^{2} k_{gl}(\theta) g_{2l}) * I)^2 + ((\sum_{n=0}^{3} k_{hn}(\theta) h_{2n}) * I)^2 \\ &= (\sum_{l=0}^{2} k_{gl}(\theta) g_{2l} * I)^2 + (\sum_{n=0}^{3} k_{hn}(\theta) h_{2n} * I)^2 \end{aligned} \tag{2.43}$$

where k_{gl} $0 \leq l \leq 2$ are the interpolation functions to steer g_2 and k_{hl} $0 \leq n \leq 3$ are the interpolation functions to steer h_2. As can be seen in (2.43) to calculate $E(\theta)$, image I just has to be convolved with the basis filters for g_2 and h_2. The interpolation functions for a particular θ can then be used to give the oriented energy. For g_2 and h_2, the interpolation functions are [16]:

$$k_{gl} = (-1)^l \begin{pmatrix} 2 \\ l \end{pmatrix} \cos^{2-l}(\theta) \sin^l(\theta), \qquad 0 \leq l \leq 2 \tag{2.44}$$

$$k_{hn} = (-1)^n \begin{pmatrix} 3 \\ n \end{pmatrix} \cos^{3-n}(\theta) \sin^n(\theta), \qquad 0 \leq n \leq 3\,. \tag{2.45}$$

Freeman and Adelson in [16] show that the interpolation functions are products of cosines and sines. Because of the squaring operation in (2.43), $E(\theta)$ can then be simplified to a Fourier series in angle, where only even frequencies are present:

$$E(\theta) = C_1 + C_2 \cos(2\theta) + C_3 \sin(2\theta) + \textit{ higher order terms} \tag{2.46}$$

where C_1, C_2 and C_3 are a combination of the basis filter outputs.

The dominant orientation θ_o can be approximated by using the lowest frequency terms and by maximizing $E(\theta)$ with respect to θ. Freeman and Adelson show that the solution for θ_o will be:

$$\theta_o = \frac{\arg(C_3, C_2)}{2}. \tag{2.47}$$

The strength of the orientation estimation S is defined as:

$$S = \sqrt{C_2^2 + C_3^2}. \tag{2.48}$$

The approximation stated in (2.47) is exact if there is only one dominant orientation locally. The dominant orientation measure and its strength S is measured at each pixel position (x, y). If there is more than one dominant orientation locally, then the approximation is not correct and other methods have to be used to find the orientations.

Discussion of the Four Methods

The calculation of the local orientation θ_o in Rao and Schunck's method is identical to Bigün and Granlund's method except that Rao and Schunck use a box filter for local averaging and Bigün and Granland use a Gaussian filter. It is not clear which filter is more appropriate for making the orientation estimation independent of input phase and reducing the noise. The two-dimensional box filter is a poor low-pass filter and its transfer function is not isotropic, i.e., it depends, for a given frequency, on the direction of the frequency. Also, since its Fourier transform is a sinc function, the attenuation in frequency domain tends to oscillate instead

of increasing monotonically with the frequency [22]. The orientation estimation of Kass and Witkin's is identical to Bigün and Granlund's method except that the directional filters are different. The angular tuning of the two methods, however, are identical.

The "coherence" measure of Bigün and Granlund χ_{bg1} when $c = 1$ is expressed the same as Kass and Witkin's coherence measure χ_{kw}. Rao and Schunck's coherence measure χ_{bg1} is different from Kass and Witkin's and Bigün and Granlund's. Rao and Schunck show for one textured image that their coherence measure gives a better result than Kass and Witkin's measure. One of the reasons for the difference in the results may be that Rao and Schunck use the magnitude of the gradient, $R(x_o, y_o)$, to emphasize high visual contrast. The inclusion of the $R(x_o, y_o)$ component may enhance the fine details in the image more than the other two coherence measures. On the other hand, a high contrast structure with no dominant orientation might have a false coherence measure due the $R(x_o, y_o)$ component.

The strength measure in (2.48) used in Freeman and Adelson's method is the absolute strength of an estimated orientation as determined by the energy output of the directional filters. The coherence measures of the other three methods, on the other hand, give an indication of how dominant an orientation is in a local region. If there are many different orientations of equal absolute strengths in a local neighborhood, then the coherence measure for any of these orientations will be very small even though their absolute strength could have been large. Choosing the right size for the local region used in the coherence measure is not clearly explained in any of the methods. Because the coherence measures are ratios of the outputs of the filters, two orientations of differing contrasts can have the same coherence measures unless the absolute strength of orientation is also incorporated as is done in Section 2.2.4 by Rao and Schunck.

Also in Freeman and Adelson's method quadrature filters are used to make orientation

estimation independent of input phase. The other three methods average the energy output of the directional filters. This averaging is effective for making the orientation estimation independent of input phase; however the spatial resolution in the orientation estimation is lowered because of blurring from averaging.

All of the methods use steerable filters. The first derivative of a Gaussian is used in both Bigün and Granlund's method and Rao and Schunck's method. The orientation tuning of Kass and Witkin's filter is identical to the first derivative of a Gaussian. Freeman and Adelson use a second derivative of a Gaussian. The second derivative of a Gaussian has a finer orientation tuning than the first derivative of Gaussian; however, it needs three basis filters instead of two to steer. There is psychophysical and physiological evidence supporting the use for both of these derivatives and higher order Gaussian derivatives to model cortical receptive fields [15].

2.3 Other Methods of Finding Orientations

Andersson in [23] uses a non-steerable set of filters with a much finer orientation tuning than the filters in the methods discussed above. If there is only one dominant orientation then there is not a big difference between Andersson's method and the methods discussed above although Andersson does not prove that that the number of filters he used (between 18 to 24 filters) is sufficient to estimate all orientations uniformly. If there are multiple orientations, such as lines crossing at different orientations, then near the point of intersection the filters in the methods discussed above may give false orientation estimates because of their broad angular tuning. Andersson's filters on the other hand will give a more accurate result since they have a narrower angular tuning. Higher order derivatives of Gaussians can also be used to detect finer

angular spacing because of their narrow angular tuning. These derivatives are also steerable; however, the higher the order of the derivative, the greater the number of basis filters needed to steer it.

The local orientation can also be estimated by finding the direction of the principle axis of image I in a local region using moments [24]. Bigün and du Buf in [25] show that the second orders complex moments of the local Fourier power of image I can be used to find local orientation. They also show that higher order of complex moments can be used to detect N-folded symmetries. The latter may also have important applications for image comparison; the focus in the thesis however will be on detection of perceptually dominant orientations, symmetric or not.

Chapter 3

Scale Problem

In an image there can be textures at different scales like Figure 3-1 (Brodatz image $D11$). Measuring texture features at the right scales is still a big problem in texture analysis. Several multiresolution or multichannel methods for texture analysis have been developed such as Gabor decomposition [26] [25] and Tree-structured wavelets [27]. Orientation detection is also dependent on scale. Several methods can be used to measure orientation at different scales. These methods will be reviewed and compared in this section.

3.1 Choosing the Right Scales

One strategy for resolving the scale problem is to measure the dominant orientations over different scales. The two big questions with this strategy are:

1. How many scales should be evaluated and what should they be?

2. How should the data from multiple scales be combined?

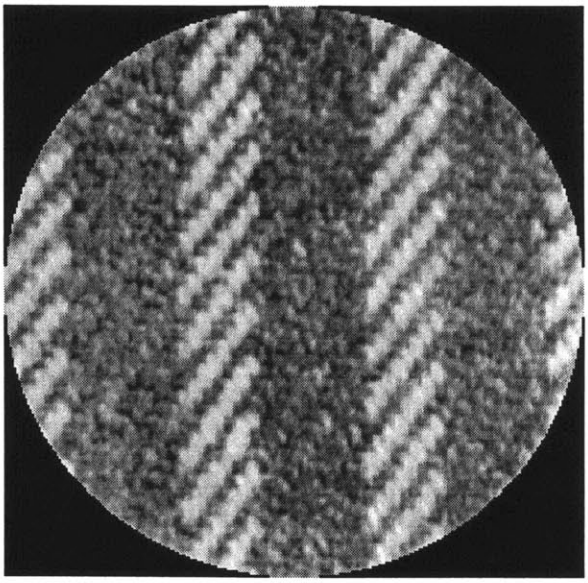


Figure 3-1: Image with orientations at different scales.

Due to the computational costs required to generate filters over a continuum of scales and storage of the filtered images, usually, a discrete, small set of scales is employed.

3.1.1 Laplacian Pyramid

A computationally efficient way to obtain a discrete, small set of scales is by computing a pyramid. The Gaussian pyramid consists of a series of images obtained by repeatedly smoothing and subsampling an image [28]. The pyramid allows large scales to be brought into the range of a local neighborhood. Overall, the computation of this pyramid requires only four thirds of the operations necessary for the first level [28].

Running a Gaussian pyramid on textured images might not be appropriate since some textures are quasi-periodic signals whose dominant frequency channels are located in the middle frequency range [27]. For these textured images, a bandpass decomposition appears more

appropriate. The Laplacian pyramid is an effective scheme for this kind of decomposition. It can be obtained by subtracting the smoothed image from the unsmoothed image on each level of a Gaussian pyramid [28]. Applying the Laplacian pyramid on an image results in frequency channels whose scale doubles from level to level of the pyramid, while the center frequency of the passband is reduced by an octave. The Laplacian pyramid is a complete representation of an image in a sense that one can perfectly reconstruct the original image given the coefficients in the pyramid. Completeness is an important property since it guarantees that no information is lost, i.e., if two images are different then their representations will be different [29].

Bigün in [30] uses the Laplacian pyramid for multiscale analysis of orientations. His orientation finding algorithm involves calculating P_{bg} described in (2.30) at each level of the pyramid. However, applying the first derivatives of Gaussians as required for the calculation of P_{bg} causes the coefficients of the Laplacian pyramid to change. The coefficients can no longer be used to perfectly reconstruct the original image. This will cause some Fourier components to be emphasized more than others and the orientation estimation will no longer be independent of the frequency content of the estimated orientation.

Also, if the texture appears in a scale which coincides exactly with one of the bandpass frequency channels, i.e., located near the maximum response of the channel then its orientation information will be obtained in the corresponding level of the pyramid. However, if the scale does not coincide near the maximum response of the channel then the frequency components associated with this scale will be attenuated and the texture will appear in low contrast in the resultant pyramid image. As can be seen, the problem of contrast and scale are not independent.

3.1.2 Steerable Pyramid

In [31], a multiscale structure called the steerable pyramid is constructed allowing orientation estimation at different scales. The steerable pyramid corresponds to a series of images obtained by repeatedly bandpass filtering, low-pass filtering and subsampling the image. Figure 3-2 shows a block diagram of two stages of the steerable pyramid. $L_o(\omega_1, \omega_2)$ and $L_1(\omega_1, \omega_2)$ are low-pass kernels and $B(\omega_1, \omega_2)$ is a bandpass kernel. The input image is convolved with the bandpass kernel $B(\omega_1, \omega_2)$ and a low-pass kernel $L_1(\omega_1, \omega_2)$. Running the steerable pyramid on an image results in a spectral decomposition shown in Figure 3-3. The overall response of the pyramid is flat in the frequency domain. The original image can be reconstructed by adding the overall response of the pyramid with a high-pass residue image.

The band-pass filter $B(\omega_1, \omega_2)$ in the steerable pyramid is constrained by the flat low-pass response of the pyramid. A symmetric 15-tap bandpass filter was used by Simoncelli et. al. [31] to approximate this bandpass filter. To use this bandpass filter for estimating orientations, an angular frequency response $A(\theta) = i\cos^3(\theta)$ is multiplied by the radial response. The result is a directional filter. $A(\theta)$ can be expressed as :

$$A(\theta) = i\left(\frac{1}{4}\cos(3\theta) + \frac{3}{4}\cos(\theta)\right). \tag{3.1}$$

In Fourier polar representation (3.1) corresponds to four non-zero polar coefficients $a_n(r)$; therefore, four basis filters are needed to steer the bandpass filter to any direction. Figure 3.1.2 shows mesh plots of the desired angular symmetric frequency response and the imaginary component of the frequency responses of the basis filter set. The angular tuning of this directional filter is finer than the directional filters discussed in Chapter 2 and is identical to a third

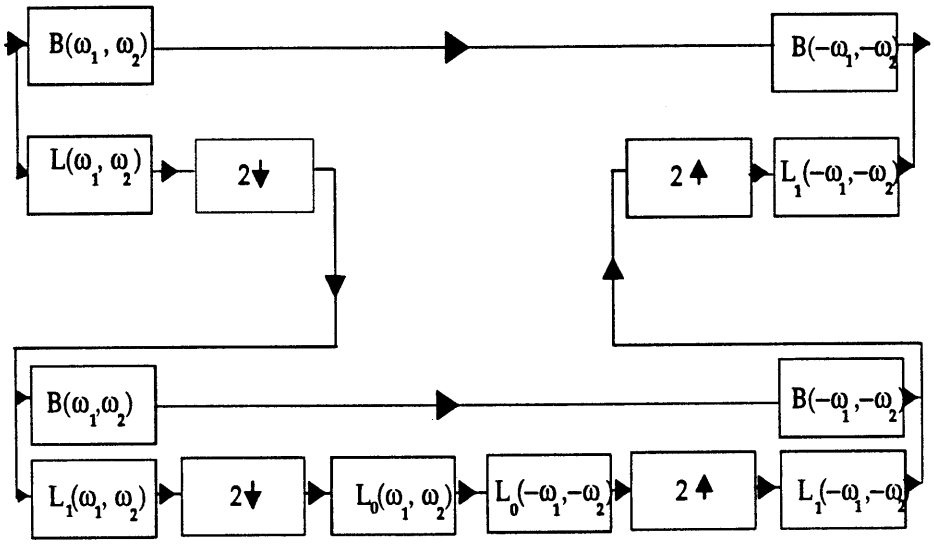


Figure 3-2: Block diagram of two stages of the steerable pyramid (adopted from [31]).

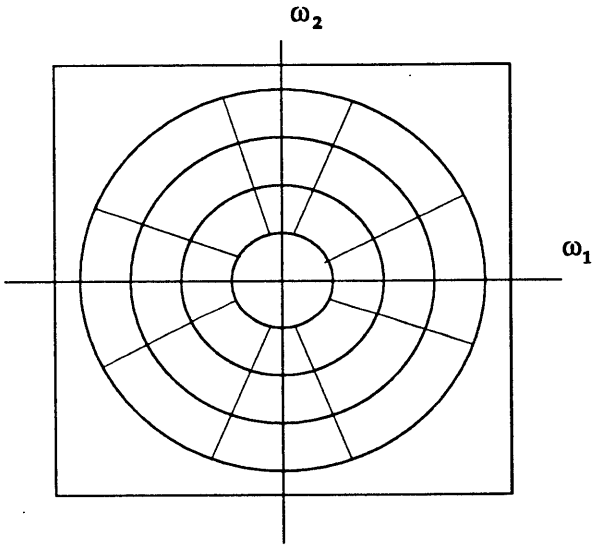


Figure 3-3: Illustration of the spectral decomposition performed by the steerable pyramid (adopted from [31]).

derivative of a Gaussian. At each level, these four basis filters are applied to the image. When the basis filters are applied again at each level, the filtered version (original image minus the high-pass residue) can be recovered.

As discussed in Chapter 2, to make the orientation estimation independent of the input phase, the Hilbert transform of the directional filter is found. An approximate Hilbert transform is found which is steerable with 5 basis filters. For multiscale orientation estimation, the oriented energy as described by (2.12) is maximized at each level of the pyramid. The image size over which the orientation is estimated decreases with the number of levels because of the subsampling. The estimated orientation and its strength, (2.47) and (2.48) can be found for each pixel (x, y) for each level. For the steerable pyramid, C_2 and C_3 in (2.48) are combinations of the outputs of the basis filters for the directional filter and its Hilbert transform. The number of levels for the pyramid is dependent on the size of the image and the size of the directional filter kernel.

Because the bandpass filters have to satisfy the flat low-pass response of the pyramid, the orientation estimation will be independent of the frequency content of the estimated orientation. However, as in the case of the Laplacian pyramid, if a texture appears in a scale which does not correspond to the maximal response of the filters associated with the pyramid then the frequency components corresponding with that scale will be attenuated and the output of the directional filter will be small. It may be difficult to detect the orientation existing in that scale.

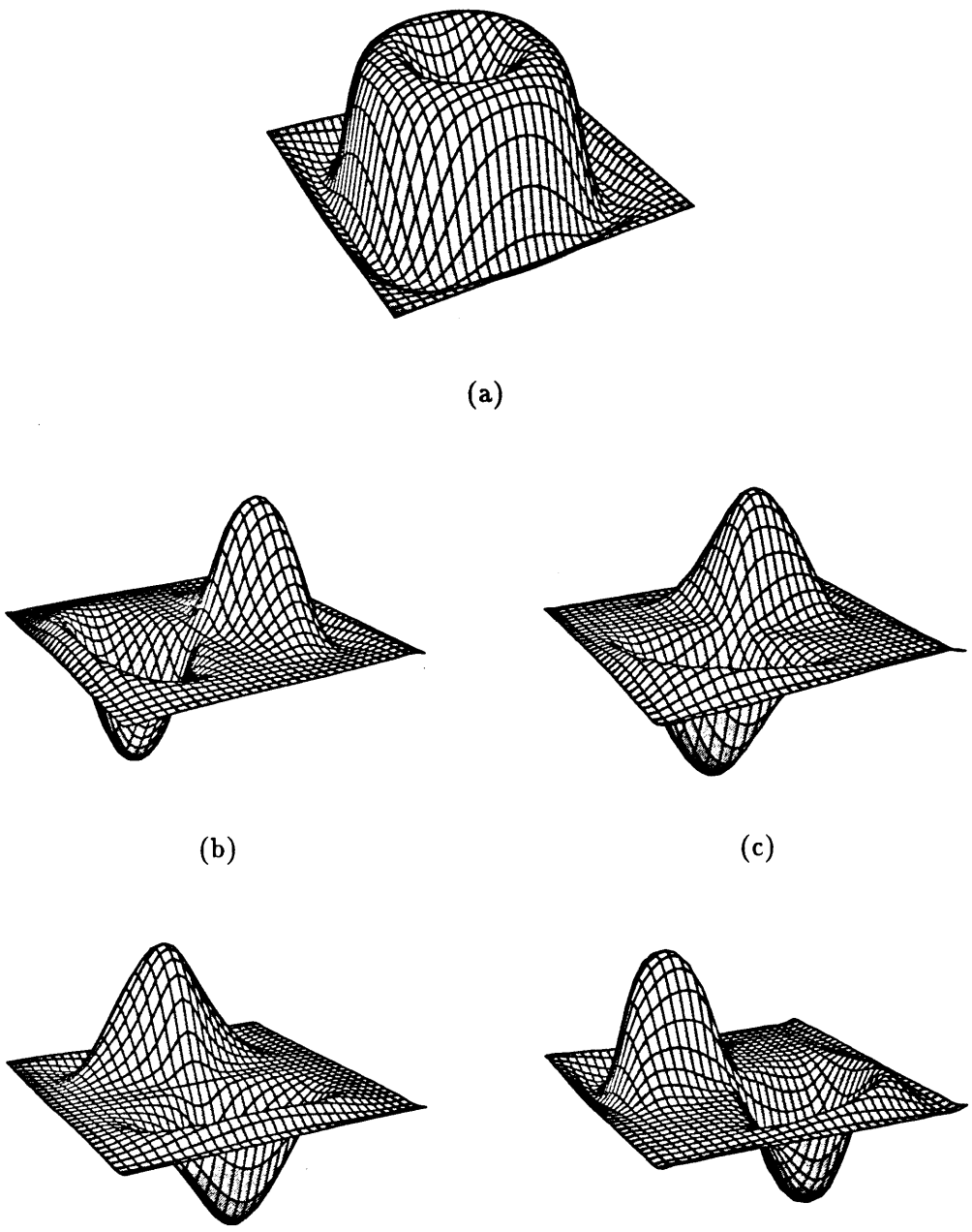


Figure 3-4: Frequency domain response plots: a) Desired angular symmetric two-dimensional frequency response. (b)-(e) Imaginary component of the frequency responses of the resulting steerable basis set. The functions shown here are the same as in [31].

3.1.3 Adaptability in Scale

Another way to estimate the orientation at the right scales is to have filters which are adaptable in orientation and scale: a minimum number of filters can be used to interpolate the local scale and orientation. The dominant orientation can then be found by first maximizing the filters over scale and then over orientation. The requirements for a filter to be adaptable over scale can be easily illustrated for a filter f which is polar separable in the Fourier domain:

$$F(\rho, \varphi) = V_1(\rho)V_2(\varphi) \tag{3.2}$$

where $F(\rho, \varphi)$ is the polar coordinate representation of the Fourier transform of f and $V_1(\rho)$ and $V_2(\varphi)$ are functions of ρ and φ respectively. It is desirable to synthesize $V_1(\rho)$ at any scale as a linear combination of basis filters:

$$V_1(\rho) = \sum_{i=1}^{K} S_i(\rho)k_i(\rho) \tag{3.3}$$

where S_i are the basis filters. Observe that (3.3) looks very similar to (2.7) where filter f was oriented to orientation θ as a linear combination of basis filters. However, $V_1(\rho)$ is not periodic and exists for all (positive) frequencies; hence, it can not have a Fourier series representation. It is then not possible to generate a continuum of scales spanning the whole positive line [17]. This may not be necessary since the range of scales of interest is never the entire real line. An interval of scales:

$$0 < \rho_l \leq \rho \leq \rho_h < \infty \tag{3.4}$$

is reasonable: if one takes the human visual system as an example, the range of frequencies to which it is most sensitive goes from approximately 2 to 16 cycles per degree of visual angle, i.e., a range of 3 octaves [17]. In this case, $V_1(\rho)$ will have a compact support and a Fourier series expansion of this region can be applied as shown in [31].

$V_1(\rho)$ can be fitted with one period of an n-term Fourier series expansion. The value of n depends on how much error can be tolerated between the actual signal and its approximate Fourier series. If the basis filters are shifted versions of $V_1(\rho)$:

$$S_i = V_1(\rho - \rho_i) \tag{3.5}$$

then (2.10) can be used to determine how many basis filters are needed [31].

One major problem with adaptation in scale is that the condition for it is in direct conflict with the Nyquist theorem. In order to avoid aliasing due to subsampling of a signal, filter f should have a limited bandwidth in the frequency domain. On the other hand, adaptation in scale requires a compact region of support in the spatial domain since a finite number of Fourier series components is desired. It is impossible to satisfy both of these conditions [31]. One solution would be to maintain full resolution in one of these parameters. Another solution is design an approximate adaptive scale representation as described by Perona [17] and Simoncelli et al. [31] where a certain amount of joint aliasing is introduced.

Another difficulty has to do with the fact that there is physiological evidence that the bandwidth of the frequency response of simple cells in the visual cortex of vertebrates is more or less proportional to the center frequency of the actual cell [31]. The frequency responses are then shifted copies of each other on the logarithmic axis. This suggests that the basis

filters described in (3.5) should be shifted copies of $V_1(\rho)$ in log frequency. This warping of the frequency (using the log frequency) does not affect the conditions for scale adaptation. However, when $\rho = 0$, $\log(\rho)$ is not defined and the log function must be modified near the origin to take care of this problem [31]. This means that the basis filters will not just be the result of translation and compression of the original filter. Also, at low frequencies a large region of support is required in the spatial domain. The computational cost can then be high due to both the number of basis filters and the size of the filter kernel. Also adaptation in scale combined with adaptation in orientation can very cumbersome.

3.1.4 Other Multi-scale Analysis

It would be nice if a small set of scales could be found which are sufficient for the orientation analysis. Wright and Jernigan [32] show that if filters are polar separable in frequency domain, i.e., are of the form expressed in (3.2) then along the radial frequency direction, six overlapping Gaussian-shaped filters are effective for coding texture information. Along the φ direction, they indicated their results were not conclusive. It appeared to them that at least seven Gaussian-shaped filters with orientations uniformly spaced along the closed range ($-90^o < \varphi < 90^o$) are required to code white noise along this dimension irrespective of the radial spatial frequency content of image. The textures that they tested were all given by polar separable Gaussian random Fields only differing significantly in their local power spectra. However, their study is a significant first attempt to find the relevant scales needed to characterize textures.

Malik and Perona in [10] for their model of preattentive texture perception used 6 directionally tuned DOOG (difference of offset of Gaussians) and 2 isotropic DOOGs with their center frequency corresponding to all integer values of the frequency between 3 and 14 cycles/deg.

This resulted in 96 filters which they claimed to adequately sample the spatial-frequency range around the peak of luminance-contrast-sensitivity function. They proposed a method of combining the outputs of these filters for texture discrimination based on a model of nonlinear inhibition. Their model is computationally expensive and is ad hoc in certain steps. Also it appears to have only been demonstrated on small number of synthetic textures. Nonetheless, it is significant since it is based on psychophysical and biological motivations.

Chapter 4

Orientation Finding Algorithm

To find dominant orientations at different scales we used the steerable pyramid described in Section 3.1.2. It provides an efficient way of filtering in the spatial domain and has the nice property of having a flat response in the Fourier domain. Also, since the angular tuning of the steerable pyramid filters is identical to the third derivative of a Gaussian, the orientation tuning is finer than that of the first and second derivative of Gaussians used in the methods in Section 2.2.4.

The local orientation θ_o and its strength S described in (2.47) and (2.48) are calculated for each pixel position (x, y) in the image at the different levels of the pyramid. The variables C_2 and C_3 in (2.48) are now combinations of the outputs of the basis filters for the steerable pyramid directional filter and its quadrature pair. Since the image sizes are 256 x 256, four levels of the pyramid level = $0, 1, 2, 3$ were used, with the subsampled image at level 3 being 32 x 32. Because of the kernel size of the filters 17 x 17, any level higher than 3 will not give additional useful information. To find the dominant orientations in the image an orientation histogram is calculated for each level of the steerable pyramid.

4.1 Orientation Histogram

A histogram H was calculated by quantizing the θ_o (expressed in degrees) estimated for every position (x, y) in the image and summing up their corresponding strengths S:

$$H(k) = \frac{N_\theta(k)}{\sum_{i=0}^{b-1} N_\theta(i)}, \quad k = 0,1,\ldots,b-1 \tag{4.1}$$

where $N_\theta(k)$ is the sum of the strengths associated with angles of all points in the local region of image I that are within the interval: $-90^o + \frac{k180^o}{b} \leq \theta_o < -90^o + \frac{(k+1)180^o}{b}$, and b is the number of bins in 180^o. For the thesis, $b = 158$ with the discretization interval in the histogram being 1.14^o. Since the angular bandwidth of the filter is much larger than 1.14^o, the choice of b is not unreasonable.

The normalization in the denominator of (4.1) enables the histograms of the different images over different levels of the steerable pyramid to be comparable. Also in the case where there are many orientations of equal strength but there is no overall dominant orientation, normalizing the strengths will give small values to these orientations. The form of $H(k)$ and the value of b which best agree with the human performance are of interest to us and pursued in chapter 6.

The definition of H is very similar to the orientation histogram defined by Tamura et al. [2] and Rao and Schunck [1]. Tamura et al. imposed a minimum threshold on the strengths before accumulating them and set $b = 16$. They followed this with a strict algorithm that permits at most two peaks. Rao and Schunk set $b = 180$ and summed strengths similar to the method here, but they did not normalize the histograms to sum to 1 for comparison. They did not discuss any algorithm for finding peaks, but appear to have done this visually. A new general method for finding any number of peaks is given in Section 4.2.

4.1.1 Test Images

For the analysis done here, the test images are the digitized textured images in the Brodatz Album, cropped to size 256 x 256. The histograms H were calculated for each image at each level of the steerable pyramid. There are in total 111 test images named $D1, D2, \ldots, D112$[1]. Four orientation histograms H were calculated for each image corresponding to the steerable pyramid levels $0, 1, 2, 3$.

For the human visual test, the test images shown to subjects were multiplied by a disc of radius 128 pixels to ensure that the subjects were not influenced by the vertical and horizontal image boundaries. Since the filters were run on square images, they "see" some information in the corners that was hidden from the subjects. This can lead to conflicting results between humans and the computer as discussed for one problem image in Section 6.3.3.

Unfortunately it is non-trivial to run a steerable pyramid on a round image. Nonetheless it would be interesting to compare the human results to the results of the algorithm using the "circular" data as well as the results using the "square" data. How the results are affected by small changes in the data is part of our continuing research.

4.2 Analyzing the Orientation Histograms

It is easy to look at the orientation histograms and visually pick the peaks which look prominent. However, it is more difficult to teach the computer to pick out these prominent peaks in the histograms.

A peak finding algorithm is proposed to pick out peaks corresponding to perceptually dom-

[1] $D45$ was not analyzed since it was missing from the tape of digitized images we received.

inant orientations. The first step in finding the prominent peaks associated with the dominant orientations involves smoothing the histogram data since they are very noisy.

4.2.1 Histogram Smoothing

A one-dimensional Gaussian filter was used to smooth the data. Solving analytically for the "optimal" amount of smoothing can be done if we have a huge training set of known "correct" smoothed distributions. Since this is not the case here, then different sizes and standard deviations were tried on a large amount of data, 111 test images, to find the minimum number of parameters for the filter which would effectively smooth the data yet keep the prominent peaks. Success was judged visually over the whole database. It was found that an eleven point Gaussian filter applied twice to the histogram smoothed most of the noise but still retained the shape of the histogram. Figure 4-1 shows an example of the smoothing with the original orientation histogram for test image $D3$[2] (look at *test*3 in Figure 6-7) at level 0 of the steerable pyramid and the two smoothed versions. Note that the orientations associated with the peaks are perpendicular to the perceived orientation.

The one-dimensional Gaussian filter, g_x, has the following form:

$$g_x = \frac{(1.02019)}{\sqrt{2\pi\sigma^2}} \exp(\frac{-x^2}{2\sigma^2}) \tag{4.2}$$

where $\sigma^2 = 16$ and $x = 0, \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$ resulting in an eleven point Gaussian filter. The 1.2019 factor was used to make sure that the coefficients of the filter would add up to one. The twice smoothed histogram H will be called H_s.

[2]Throughout thesis, all images $D1 - D112 = test1 - test112$ multiplied by a circular disc

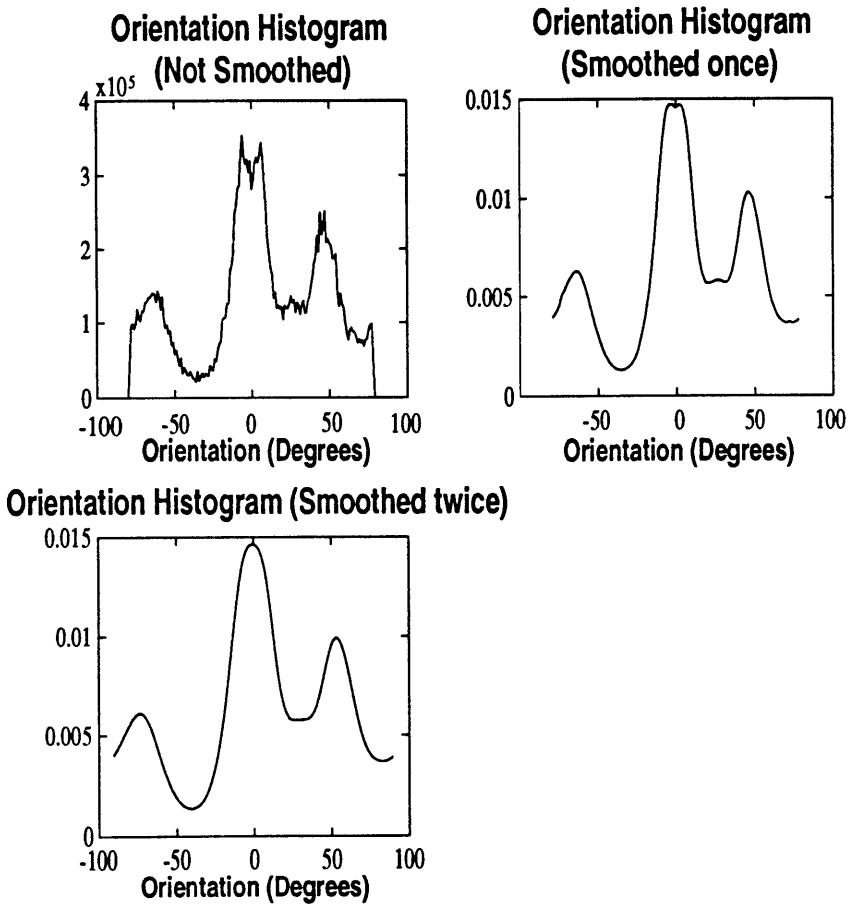


Figure 4-1: Smoothed orientation histograms for D3.

4.2.2 Finding the Prominent Peaks

There are many ways to "find peaks" in a histogram. One way is to fit some functional form, e.g., Gaussians, to the histogram, and associate a Gaussian with each peak. This method requires assumptions on how many Gaussians to fit, and what range their widths should be. Also, Gaussians are of infinite duration so they would need to be truncated. A heuristic study of the 111 histograms indicated that not all peaks of interest are symmetric like Gaussians. A Gaussian fitting method will tend to fit two or more Gaussians to such peaks, which will not typically correspond to the number of different dominant orientations perceived. The method used below is therefore based on derivatives of the histogram, which do not require any kind of symmetry assumptions. Even after the smoothing there can be still a lot of perturbations in the data not necessarily corresponding to dominant peaks. Fitting this kind of data with Gaussians can cause the detection of a lot of false peaks. The histogram shown in Figure 4-2 corresponds to the orientation histogram of the level 0 of the steerable pyramid for $D30$. $D30$ is clearly non-directional (see *test*30 in Figure 6-13); however, very sharp and narrow Gaussians can be fitted to the data. It is necessary to follow the fitting, or to follow the derivative-based method below with higher level decision making to find peaks.

To find the prominent peaks, the maximum and the minimum points of the smoothed orientation histogram were found. The maximum and minimum points can be found by finding the first order difference of H_s. Letting d_H represent the first order difference of H_s :

$$d_H(k) = H_s(k+1) - H_s(k) \quad 0 \leq k \leq b-2 \tag{4.3}$$

$$d_H(b-1) = H_s(0) - H_s(b-1) \tag{4.4}$$

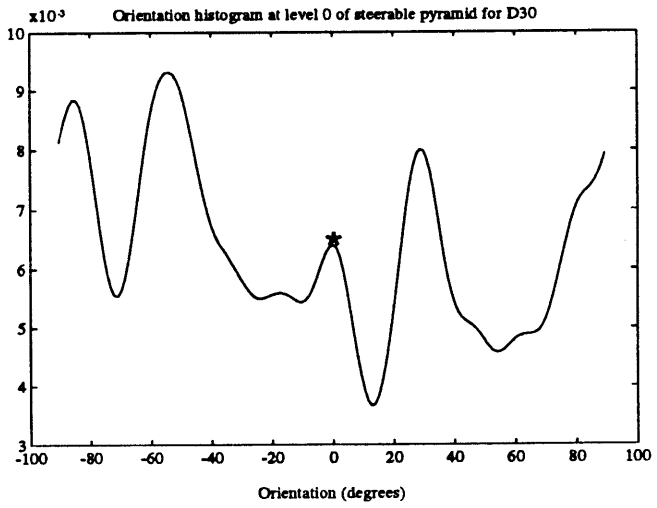


Figure 4-2: Orientation Histogram for D30 at level 0 of the pyramid.

---

where (4.4) is true since H_s is periodic, i.e., $H_s(b+1) = H_s(0)$. The zero crossings of d_H indicate the maxima and minima of H_s.

For a strong oriented pattern made up of line segments along one direction whose width and spacing corresponds to the size of the directional filter and its frequency tuning, there will be one peak in the orientation histogram. The peak will have a high value and will be narrow. This would be considered a prominent peak. If the pattern is not strongly oriented, or the spacing and size of the structures do not correspond with the parameters of the directional filter or there are structures at many orientations then the peak will be much wider and its magnitude much smaller.

Since the form of the distribution of the histograms is not known, a method assuming a general form for the distribution was implemented. This method gives a higher "salience" measure as the peaks become narrower, sharper and bigger in magnitude and a smaller "salience"

measure as the peaks become broader and smaller in magnitude. An approximate measure of the sharpness of a peak can be determined by approximating its height and width, and taking the ratio of these two values. To estimate these values, first the inflection points on either side of the peaks in H_s are found. Figure 4-3 shows the orientation histogram calculated for $D3$ and the corresponding graph of d_H. As can be seen, the zero crossings correspond to the maximum and minimum points, and the inflection points correspond to the largest slopes (positive or negative) on either side of the maximum or minimum points. A positive inflection point before a negative inflection point indicates that the zero crossing corresponds to a maximum point in H_s.

The difference of values between the corresponding inflection points is affected by the magnitude of the peak in H_s and the largest slopes on the two sides of the peak. The distance between the inflection points gives an indication of the narrowness of the peak. This information about the inflection points can then be used to give a measure of the magnitude and sharpness of the peak. Figure 4-4 gives an example of a prominent peak P_1 in an orientation histogram (corresponding to the level 0 of the steerable pyramid for $D1$) and its corresponding inflection points. The inflection points (max_1,min_1) corresponding to peak P_1 are very close. $D1$ can be seen in Figure 6-5. Figure 4-5 gives an example of a much less prominent peak P_1 in an orientation histogram (corresponding to the level 0 of the steerable pyramid for $D9$) and its corresponding inflection points (max_2,min_2). $D9$ can be seen in Figure 6-1. Notice that for the prominent peak P_1 in Figure 4-4, the inflection points are much closer and the difference of their values much higher than the ones of the less prominent peak P_1 in Figure 4-5.

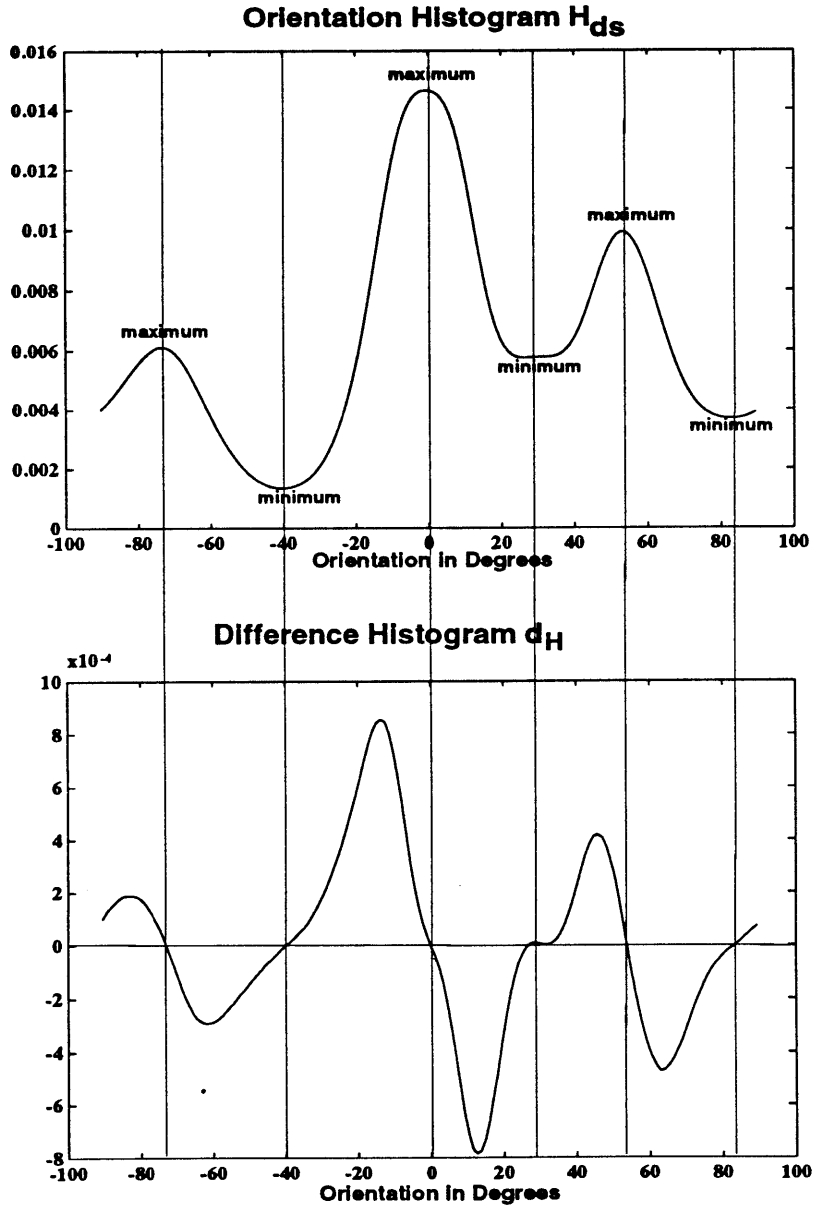


Figure 4-3: Orientation histogram and the correponding d_H graph for $D3$.

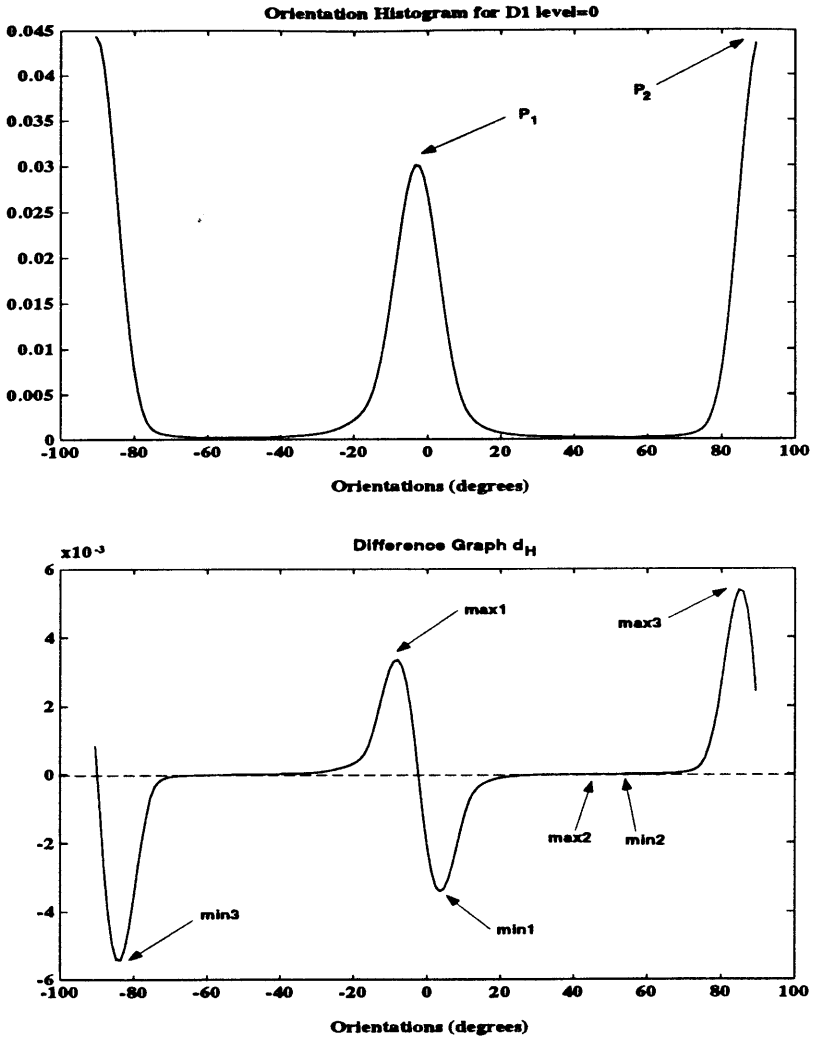


Figure 4-4: Orientation histogram and the correponding d_H graph for $D1$.

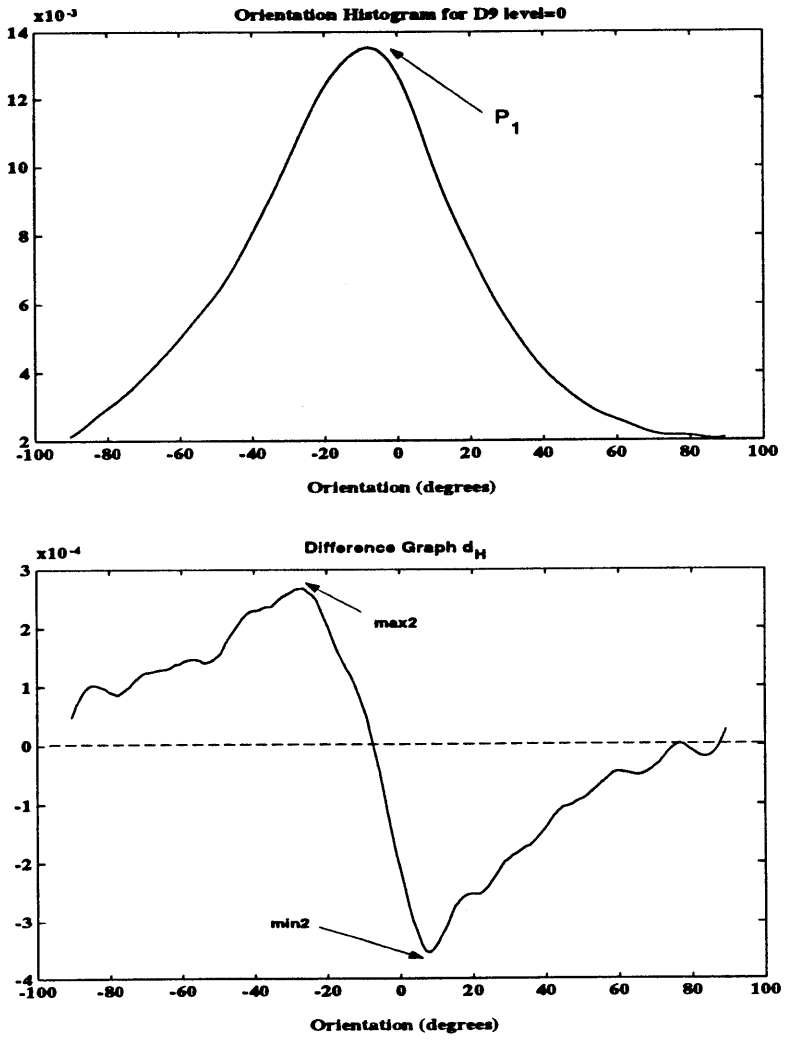


Figure 4-5: Orientation histogram and the correponding d_H graph for $D9$.

We propose a salience measure γ for finding prominent peaks,

$$\gamma = w_m \, H_s(\theta_p) \, \frac{V_{dist}}{H_{dist}} \, w_{br} \tag{4.5}$$

where θ_p is the orientation associated with a peak. The vertical difference V_{dist} between the inflection points gives a measure of the magnitude of the peak in H_s and of the steepness of its sides. The horizontal distance H_{dist} between the inflection points indicates the narrowness of the peak. These distances are marked on Figure 4-6.

The motivation for w_m will be discussed below and the motivation for w_{br} will be discussed in Section 6.3.1. The weighting function w_{br} was not incorporated for the first set of comparisons described in Section 6.3. Even though γ is dependent on the magnitude of the peak, to make sure that the γ for small valued peaks is much smaller than large valued peaks, the peak magnitude $H_s(\theta_p)$ is included in the calculation of γ as shown in (4.5).

There are two cases which have to be accounted for when calculating γ. Case 1 is discussed below. Case 2 is discussed in Appendix B.

Case 1: Peaks in the orientation histogram where the value of the maximum point in the histogram is very close in value to one or both of its neighboring minimums. Figure 4-2 shows an example of this case with the peak marked by *. In these cases, the ratio $\frac{V_{dist}}{H_{dist}}$ might falsely indicate a strong peak (especially if the peak drops off sharply) but the peak should not be considered because it is caused by a perturbation in the histogram. For a non-oriented image, ideally H_s should be flat for all orientations; however, because of noise and differing contrasts, some orientations will be weighted more than others. In Figure 4-2,

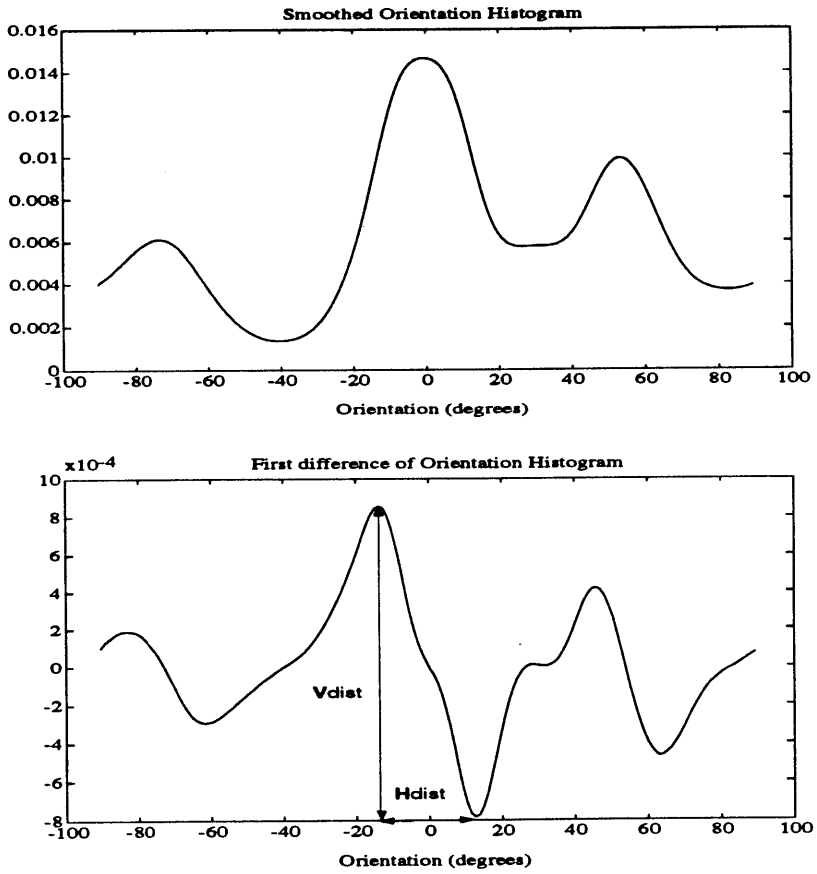


Figure 4-6: Graph of d_H for $D3$ showing the H_{dist} and V_{dist} measures.

the maximum in the histogram denoted with * will have a large ratio $\frac{V_{dist}}{H_{dist}}$ value because it is sharp on one side but it is clearly not a prominent peak (as can be seen in D30). The proposed function w_m can take care of these cases. The goal is to incorporate the ratio of the largest minimum value (there are two minimums on either side of the peak) and the maximum value. If the value of the maximum point is denoted as MAX_{val} and the largest minimum value is denoted as MIN_{val} then the weighting factor, w_m can be expressed as:

$$w_m = 1 - \frac{MIN_{val}}{MAX_{val}} . \tag{4.6}$$

If the minimum value is large relative to the maximum value then w_m is close to 0 which will make γ very small. If the minimum value is much smaller than the maximum value, then w_m will be close to 1 and γ will stay almost the same.

This peak finding method is plainly heuristic and therefore has very little worth unless it is proven to work on a large variety of data. Our results indicate that for the 111 different Brodatz images with 4 levels of resolution for each, the peaks found by the algorithm in the 444 histograms corresponded very well with the perceptual strength of the peaks. These results are compared to the perception of humans in Chapter 6.

4.3 Dealing with Contrast

In both the Fourier methods and filtering methods discussed in Chapter 2, the strength of the estimated orientation is dependent on the contrast of the textured image and increases linearly with contrast. The coherence measures described in Chapter 2 are less dependent on contrast since they are ratios of the outputs of the filters. Because of contrast dependency of

the methods, a low contrast oriented structure will have a small strength associated with its orientation. As a result, the contribution of this orientation will be small since its strength is smaller. Humans, on the other hand, can detect orientations at low contrast as shown in the results of the human visual test described in chapter 6. How can we make the orientation estimation less dependent on contrast?

4.3.1 Contrast

First what is meant by contrast? When studying the visibility of objects such as disks, or bars, or rectangles on a background, as in [33], contrast C is defined as:

$$\begin{aligned} C &= \frac{(L_o - L_B)}{L_B} \\ &= \frac{\Delta L}{L_B} \end{aligned} \tag{4.7}$$

where L_o is the luminance of the object and L_B the luminance of the background. For periodic spatial patterns like sine gratings, C can also be defined as:

$$C = \frac{(L_{max} - L_{min})}{(L_{max} + L_{min})} \tag{4.8}$$

where L_{max} and L_{min} are the maximum and minimum illuminances for the signal [33].

One big problem with orientation analysis using directional filter outputs is that the responses of filters increase monotonically with the input image contrast. A non-oriented pattern with high contrast can give a response very similar to as an oriented pattern with small contrast. It would be desirable to have a measure for orientation analysis which is less dependent on input contrast. But at the same time it is important to keep the actual stimulus contrast

information since a high contrast oriented structure is more strongly perceivable than a low contrast oriented structure. There are several methods which are used to enhance the low contrast oriented structures.

4.3.2 Non-linear Transformation of the Magnitude of Orientation

Since the estimation of local orientation θ_o is a ratio of the responses in the filters as shown in (2.47) it will not be influenced by the contrast of the image. As long as there is a change in the grey level in a certain direction in a neighborhood the orientation measure will capture this direction. However, the strength measure (2.48) will be dependent on the contrast of the image. Since the orientation histogram H_s is calculated using the normalized sums of the strengths, the magnitude of its peaks will also be dependent on the local image contrast. An example of this problem is shown in Figure 4-7(a) where the peak in the orientation histogram H_s for $D6$ (see *test*6 in Figure 6-3) corresponding to the horizontal orientation (peak at 90^o in the histogram) is hardly noticeable even though visually perceivable. In the visual test, this horizontal orientation was given a large strength by the human subjects. When analyzing the orientation histograms it will be very difficult to determine whether this is a prominent peak or due to noise.

To enhance these peaks, the logarithm of the strength S can be taken for each pixel position (x, y). The logarithm transformation enhances the small values and compresses the larger values. Since the logarithm of the strength values less than one will be negative, a constant value of one must be added to the strength values before taking the logarithm. The addition of this constant will also eliminate the occurrence of Log of 0. Figure 4-7(c) shows the H_s histogram for D6 after the logarithm transformation of its orientation strength. The peak for

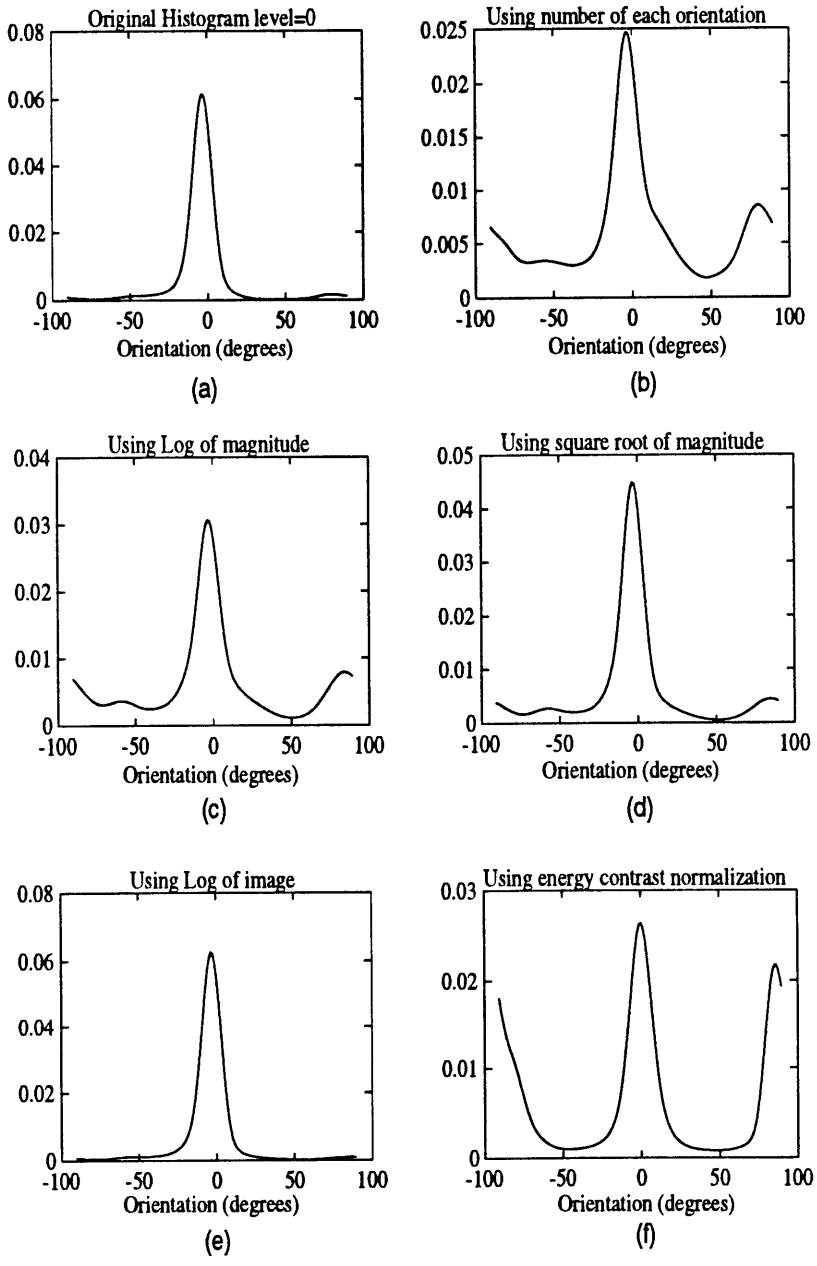


Figure 4-7: Orientation histograms for $D6$ using difference contrast measures

the horizontal orientation is more noticeable but it is still much smaller than the other peak. One big problem with performing the logarithm transformation is that also enhances the noise in the strength values which may lead to false peaks in the orientation histogram.

Another non-linear transformation that is also used to enhance the small values is taking the nth power of the strength values where $0 < n < 1$. Figure 4-7(d) shows the H_s histogram for D6 after the square root of the orientation strength was calculated. As can be seen the peak is more noticeable but still very small. Like the logarithm, taking the square root also enhances the noise in the strength values.

4.3.3 Contrast Enhancement of Images

Another way of making the orientation estimation less dependent on the contrast of the image is to enhance the contrast in the image before analyzing the orientation information. The contrast enhancement of the image enables the low contrast oriented structures to be more noticeable; however, it also boosts the contrast of the non-oriented structures which can result in false peaks in the orientation histogram.

One popular method of enhancing the contrast of an image is by performing histogram equalization. The grey levels of the image are transformed to obtain an approximate uniform density for the image. A uniform density implies an increase in the dynamic range of the pixels which can have a considerable effect on the appearance of an image [34]. Chaudhuri et. al. [14] used histogram equalization to improve the global contrast of the images before applying their directional finding method. The problem with global histogram equalization is that it usually fails to enhance details over small areas. For local enhancement, the histogram equalization technique can be done over neighborhoods in the image. The size of the neighborhoods of

course depends on the scales present in the image! Also histogram equalization can introduce contouring artifacts which may possibly lead to false alarms for orientation.

Another method to enhance the contrast is to perform a logarithm transformation over the image. It has been shown that for high mean luminance and for test stimuli of large area and long duration, psychophysical sensitivity follows Weber's Law:

$$\frac{\Delta L_T}{L_B} = k \tag{4.9}$$

where k is a constant and known as the threshold contrast and also as the Weber ratio [33]. According to (4.9) when Weber's law is obeyed, the visual system's criterion for contrast detection is that the stimulus contrast must exceed the fixed value, k, the threshold contrast. The differential of the logarithm of intensity of image I is:

$$\delta(\log I(x,y)) = \frac{\delta I(x,y)}{I(x,y)}. \tag{4.10}$$

With respect to Weber's law, (4.10) implies that changes in the logarithm of the intensity of the image which equal the Weber's ratio will be perceived to be of the same contrast over the region of intensities for the which the Weber fraction is constant. For this reason, it may be preferable to perform operations on the logarithm of the intensity of the image point, rather than the intensity itself [35].

Also, if each pixel value of image I can be regarded to be product of the illumination component, i_l and the reflectance component r_e :

$$I(x,y) = i_l(x,y)r_e(x,y) \tag{4.11}$$

then taking the logarithm transform of I separates these components:

$$\log I(x,y) = \log i_l(x,y) + \log r_e(x,y). \tag{4.12}$$

The difference in response across a uniformly illuminated step edge in the log-transformed image will then be given by

$$(\log i_l + \log r_{e2}) - (\log i_l + \log r_{e1}) = \log(\frac{r_{e2}}{r_{e1}}) \tag{4.13}$$

where r_{e2} and r_{e1} are the reflectances on the two sides of the edge. As can be seen, the difference is independent of the illuminance and as a result, the response to a given reflectance change will always be the same irrespective of the illumination [35].

Figure 4-7(e) shows the orientation histogram for D6 when the image was log-transformed. The peak for the horizontal orientation is even smaller than before. One problem is that since the horizontal lines and the background have high intensity values, apply a logarithm transform will make their difference even smaller since the transform compresses large values.

4.3.4 Contrast Normalization

Another way to make the orientation estimation more robust to contrast is to make the the filter outputs less dependent on the input image contrast by performing an energy contrast normalization. One way to do this is through local contrast normalization. The contrast normalization used in low-level perception of images can be modeled by dividing the energy of the filter outputs by the sum of energies corresponding to filters at all orientations [36] [33]. For example, Bergen in [9] did contrast normalization by taking a local average of the energies

of the four directional filters he ran on textures and divided the energy output (in his case he defined energy to be the squared filter output) of each filter by the sum of local energies of all four filters.

A contrast normalization scheme is implemented in the spirit of the contrast energy normalization methods used in [36] [9] and an algorithm used by Freeman for the "Gain Control" method in his PhD thesis [37]. For the contrast normalization, the outputs of each of the basis filters of the steerable pyramid can be normalized and used to find the normalized oriented energy $E_n(\theta)$. However, this contrast normalization is tricky since it can affect the steerability of the oriented filter. In the case of the steerable pyramid directional filters , since the basis filters have an orientation tuning proportional to $\cos^3(\theta)$, squaring the outputs of these filters (for energy) means that the resultant images have a finer orientation tuning proportional to $\cos^6(\theta)$. Therefore to interpolate the oriented energy $E(\theta)$ to any orientation θ requires 7 basis filters.

Seven equally spaced samples of $E(\theta_s)$ are taken at $\theta_s = 0^o$, 25.7^o, 51.43^o,77.14^o, 102.86^o, 128.57^o, 154.26^o. This is the same as taking the squared outputs of the steerable pyramid directional filter and its approximate Hilbert transform rotated at 7 equally spaced orientations. Therefore, our new basis filter outputs are the squared outputs of the steerable pyramid directional filters and their quadrature pair rotated at the above mentioned directions.

Each of the sampled oriented energies $E(\theta_s)$ are normalized in the following way at each pixel position (x, y) :

$$E_n(\theta_s)(x,y) = \frac{E(\theta_s)(x,y)}{c + E_s} \quad c > 0 \tag{4.14}$$

where $E_n(\theta_s)$ is the normalized energy, E_s is the sum of $E(\theta_s)$ in a local neighborhood and c is a constant. Constant c is used to make sure that when $E(\theta_s)$ is zero, the result will not be undefined. It is important to chosen a value for c which is not bigger than most of the values of $E(\theta_s)$; otherwise, $E_n(\theta_s)(x, y)$ will be smaller than it should be. The choice of the local neighborhood size for calculating E_s is not straightforward. One way to find a good size is to calculate (4.14) for different sized neighborhoods and see which size is reasonable for a number of different images.

The normalized energy $E_n(\theta)$ can be expressed in terms of these seven normalized energy outputs using the coefficients k_l (2.9) needed to steer a filter with the angular tuning $\cos^6(\theta)$. The estimated orientation θ_o and its strength S described in (2.47) and (2.48) can be found for each position (x, y). The coefficients C_2 and C_3 in (2.48) will be combinations of the seven normalized energy outputs.

The energy normalization outlined above was implemented for the first level of the steerable pyramid and future work is to extend it to other levels. The range of values for the energies of the directional steerable pyramid filters at level 0 were found and the constant c was chosen to be an order of magnitude smaller than the minimum value in this range. Freeman in his PhD thesis [37] tested his contrast normalization scheme on different neighborhood sizes and found that the neighborhood size corresponding to the blur of the third level of a Gaussian pyramid gave very good results. We used this neighborhood size for our implementation. For future work the effect of neighborhood size on the results of the orientation finding scheme will be studied.

A contrast normalized orientation histogram was calculated for level 0 of the steerable pyramid using the values of θ_o and S calculated from the normalized energy. Figure 4-7(f)

shows the contrast normalized orientation histogram for $D6$. Compared to histograms Figure 4-7(c)-(e) the peak boosted at 90^o is much higher than the other methods.

As mentioned in Section 3.1.1 the problems of contrast and scale are not independent. Figure 4-7(b) shows the orientation histogram for $D6$ obtained by summing over the number of pixels at each orientation rather than summing over the strengths of the orientations. This orientation histogram is not affected by the contrast in the original image since orientation is calculated using the ratio of the filter responses. As can be seen, the peak at 90^o in Figure 4-7(b) is higher in value than the corresponding peak in Figure 4-7(a). However, it is much smaller and broader than the peak at 0^o. It appears that the spacing of the lines in $D6$ can also be causing the filter responses to be small. Looking at the result of the contrast normalization, it may be used to boost filter responses affected both by the contrast of the image and the size of the filters. Interestingly, subjects gave a higher strength to the horizontal orientation than the vertical one.

In Chapter 6, the results of the contrast normalized orientation histograms and the non-normalized orientation histograms will be compared to the human visual data.

Chapter 5

Human Visual Experiment

The goal is to determine how well the orientation finding algorithm described in Chapter 3 estimates dominant orientations in textured images. For one thing it is important to determine whether the levels used in the steerable pyramid and the contrast normalization technique described in Chapter 4 are sufficient to estimate orientations at different scales and contrasts. Humans are very good at detecting directionalities at different scales and contrasts. One way of assessing the effectiveness of the orientation detecting algorithm is through a comparison with a human visual experiment.

If an orientation is strongly perceived by humans then the orientation finding algorithm should be expected to also detect such an orientation. If it does not then it should be modified until the results are in closer agreement with the humans. This of course can be an exhaustive search problem since there are many combination of parameters that can be changed such as filter size, contrast normalization, etc. For now, it is sufficient to choose a set of parameters initially and vary some of them, watching how they affect the agreement between the human and the filters so that the majority of the the results are similar to those obtained from the

human visual test.

5.1 Human Visual Experiment

5.1.1 Subjects

Forty subjects participated in the visual experiment. The majority of the subjects were MIT students and had no previous encounter with this research topic. Of the forty subjects, sixteen were female. They came from different areas of specialization and from various ethnic origins. The subject data are identified with subject numbers $Sub1 \ldots Sub40$ where the data $Sub1$ corresponds to the data of the first subject who went through the visual experiment.

5.1.2 Experimental Setup

The subjects were seated about forty centimeters from the computer monitor. This distance was sufficient to make sure that the subjects did not see any particular pixel in the image but still could detect fine details. The lighting in the room was turned off except for a background light. This insured that there would not be any false illumination on the images. The images of size 256 × 256 were positioned in the middle of monitor screen. To indicate the orientation perceived, the subject used a mouse to rotate a red bar centered in the middle of the image to a particular orientation as illustrated in Figure 5-1. To insure that the subjects were not influenced by the horizontal and vertical image boundaries, each image was multiplied with a disc of size 256 × 256 with a radius of 128 pixels.

Using a red bar in the center of the image to indicate the orientation can bias the subject to the orientations close to the center. Other methods to indicate the orientations were considered

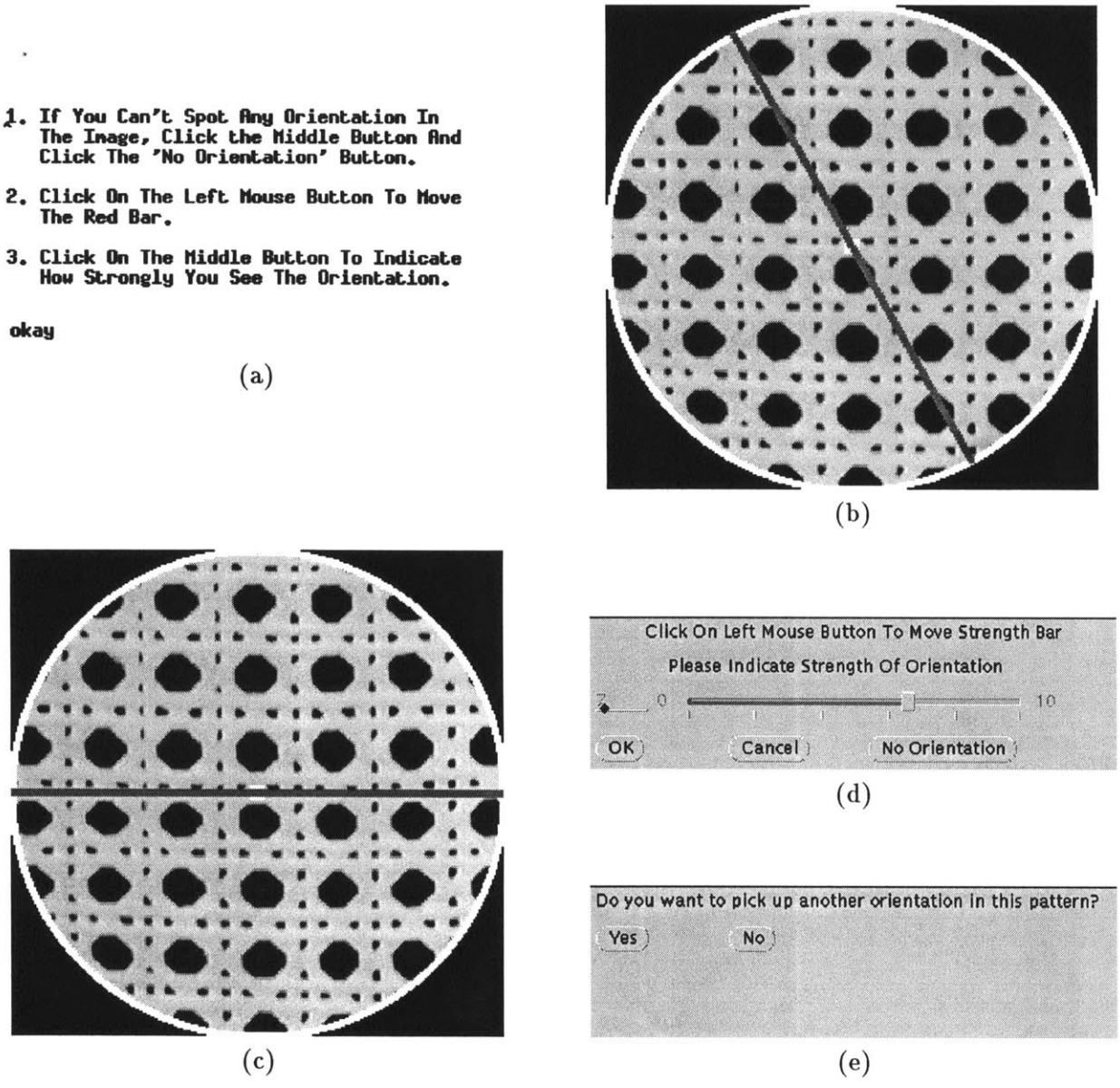


Figure 5-1: Human test setup: (a) The subject clicks mouse to indicate whether or not they see a dominant orientation. (b) If they see one, a red bar pops up for them to rotate. (c) One of the orientations selected. (d) Human subject specifies strength of orientation. (e) A panel pops up asking if subject sees another dominant orientation.

such as moving a bar freely around the image; however, this could encourage the subject to do more of a template matching than orientation detection.

To indicate how strongly the orientation was perceived, a menu bar popped up under the test image as shown in Figure 5-1. The subject used a slide bar to give a strength value between 0 and 10. The strength values could only be integer values. A finer scale could have been used; however, it is difficult for the subject to give very quantitative measure of their perception, i.e., whether an orientation should have a strength 6.7 instead of 6.9. After a strength was indicated, the subject was prompted whether they wanted to pick another orientation or move to the next pattern. In this way, the subjects should not have been biased toward picking any particular number of orientations.

5.1.3 Training Session

The training for this experiment is difficult because it is important not to bias the subject to a certain orientation, strength, or structure. To give a subject a notion of what is orientation and what is meant by a strength of orientation, two images were first shown: an ideal directional pattern and a random noise image. The first image was a sinusoidal grating oriented 90^o as shown in Figure 5-2(a). The subject was told that this particular image has a vertical orientation and was shown how to indicate this orientation using a mouse to rotate the red bar. Since the orientation of the grating is clearly perceivable, a maximum strength of 10 was assigned to it. The subject was shown how to indicate this strength using the slide bar. Next, a random noise image, Figure 5-2(b) , was shown. The subject was told that this image had no dominant orientations and was assigned a minimum strength of 0. For images with no dominant orientations, the subject was asked to choose the **No Orientation** option in the

strength menu. To make sure that the subject is not biased towards continuous lines, the training image, Figure 5-2(c), was shown with a broken line oriented near the corner. The broken line was drawn near the corner to also indicate that the orientations are not necessarily near the center. When the experiment was first conducted, a strength value of 10 was given to the training images. However, compared to the first training image, Figure 5-2(a), it is not obvious that they should be given the same strength. To avoid biasing the subject from picking 10 all the time, for the rest of the training images the subject was asked to choose their own strength considering the first two cases. The subject was also asked to practice using the mouse to indicate the orientation.

To show that there can be multiple orientations in an image, Figures 5-2(d) and 5-2(e) were shown. In Figure 5-2(d), there are two lines of different thicknesses at different orientations (not symmetrical). The subject was asked to indicate the orientations (in whichever order) and their associated strengths. Interestingly, the majority of subjects gave a slightly higher strength to the thicker line than the thinner. The next training image, Figure 5-2(e), has 3 dominant orientations. The subjects were asked again to indicate the three orientations. The majority gave the vertical line a much higher strength than the two diagonal lines. This image was a good example to illustrate that not all the orientations are equally visible.

Finally to illustrate that there can be cases where there are a lot of orientations in an image but the overall effect is that there is no dominant directionality, Figure 5-2(f) was shown. For this image, the subject was asked to only choose the orientation of the white bar (again displaced from the center) as the dominant orientation. This avoids the case where the subject chooses all the orientations of the line segments even though the overall effect is of no dominant

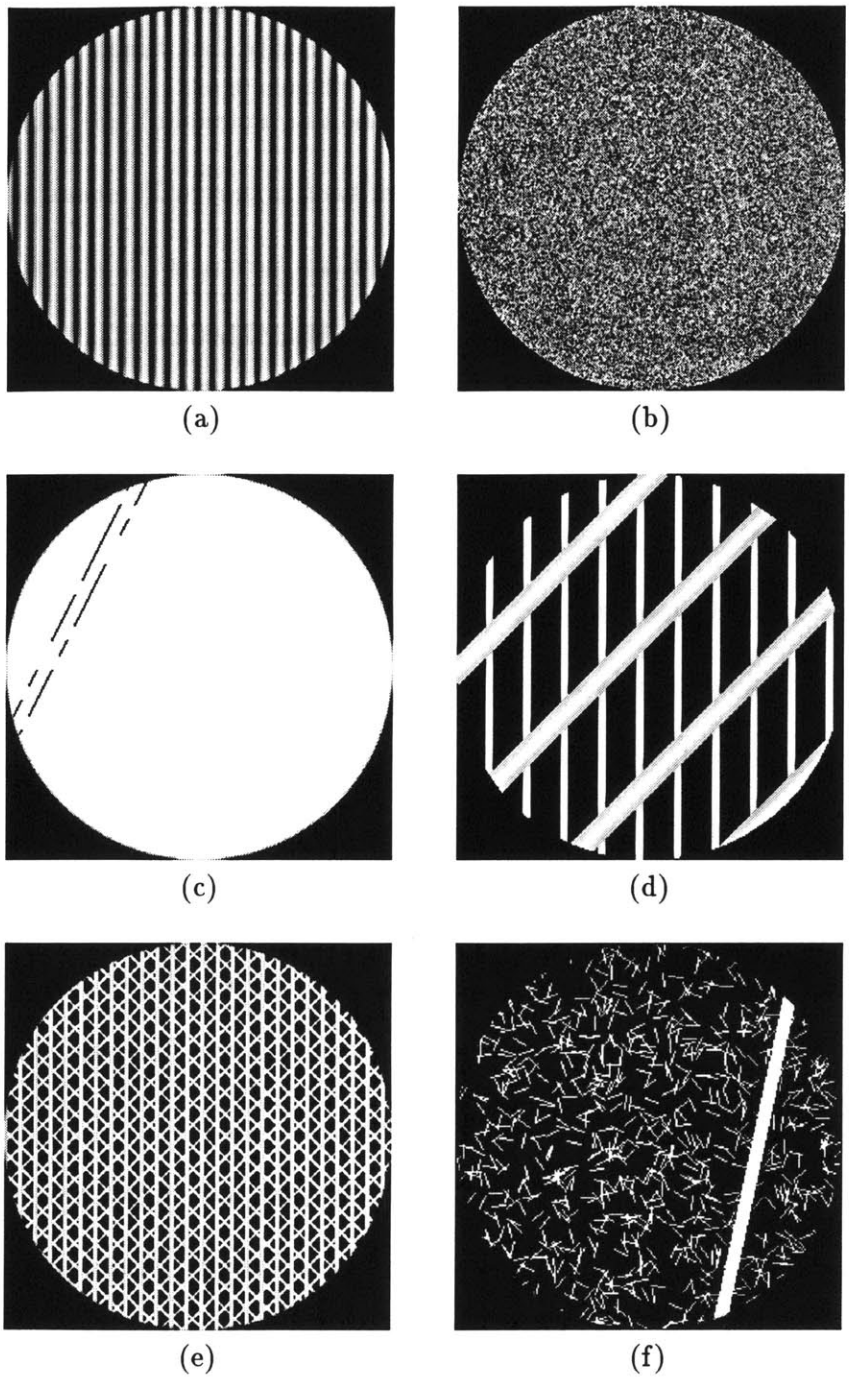


Figure 5-2: Traing images (a)-(f)

orientation.

The words **Dominant** and **Minimum** were repeated several times to make sure that the subject only picks the "minimum number of dominant orientations" that he/she thinks are dominant in an image. The instructions were shown in the upper left corner of the monitor screen. Near the top of the screen, the reminder **REMEMBER TO PICK THE MINIMUM NUMBER OF DOMINANT ORIENTATIONS** was also shown.

5.1.4 Test Images

Each subject was shown 37 test images. 30 of the images were from the Brodatz album cropped from the middle of the 512 × 512 digitized images. Since it would have taken too long for a subject to analyze 111 of the textured images, the images were divided into four sets so that each subject only have to analyze 30 of the images. We made sure that very similar images were not placed in the same set. Each of the test images except for nine test images (*test*1, *test*17, *test*23, *test*24, *test*36, *test*73, *test*77, *test*98 and *test*100) were analyzed by ten subjects. Each of the nine test images were analyzed by twenty subjects. The reason that these test images were analyzed by more subjects is that the 111 images could not be evenly divided into four groups.

Results are described in the comparison section in Chapter 6. The last seven test images shown in the Appendix A were visual teasers. Results on teasers are not going to be included in the thesis since the main emphasis is on analyzing real-life textures.

The Brodatz texture images are not all homogeneous or directional. Some of them have very complex patterns for which it is difficult to indicate a minimum number of dominant orientations. Also there are not many images with low contrast since the images were taken

in a very controlled environment with little noise and uniform illumination. Use of the 111 Brodatz textures makes this study considerably larger in texture variety than any other study to date. There is definitely a majority of horizontal and vertical oriented structures (as there is also in the environments where most people spend daylight hours.) Nonetheless, each test image was treated as one region, making the task more difficult but the results more realistic for extension to real scenes. Also Brodatz textures are used by many computer vision researchers for their analysis and they contain real-life textures for which the human visual system is well-adapted. The test images will be called $test1, test2, \ldots, test112$ corresponding to Brodatz textures $D1, D2, \ldots, D112$.

5.1.5 Recording of the Human Visual Data

The orientations and strengths picked by the subject for each test image are recorded in a file corresponding to that subject. The orientation and strength data for each test image are collected across these files. Information about number of orientations, their relative strengths, expected number of orientations for each test image and distribution of strength values chosen by each subject could be found from these files.

Several statistical properties of the experimental data were considered, and the following conclusions were made:

1. The subjects used the full range of strengths, $0 - 10$, with an approximately uniform distribution.

2. No single subject's use of the strength scale deviated sufficiently to be discarded.

One of the concerns for the experiment was that the subjects might not use the full range

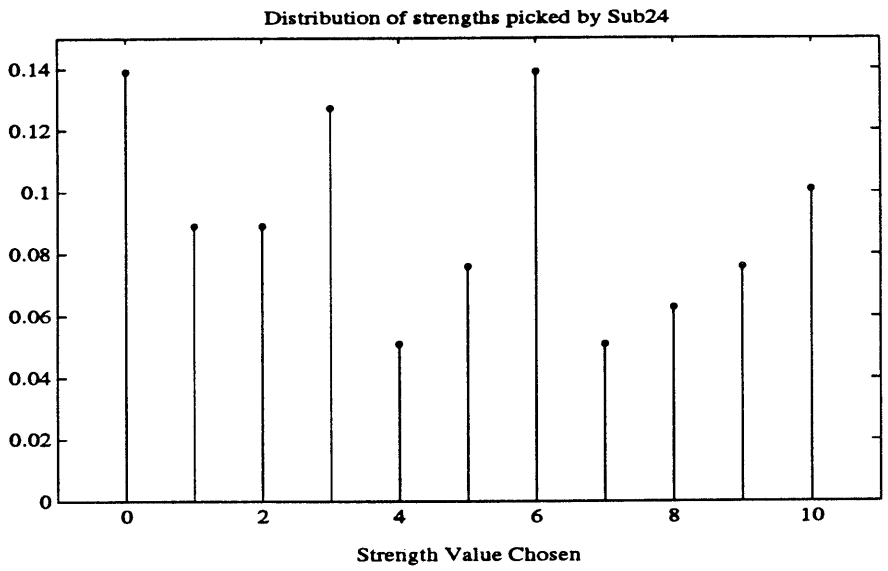


Figure 5-3: The distribution of strengths chosen by $Sub24$ over all test images

---

of orientation strength values between $0 - 10$. A histogram for each subject was calculated showing over the total number of test images shown to each subject how many times each strength from 0 to 10 was chosen. Figure 5-3 shows a sample histogram for $Sub24$. The vertical axis corresponds to the fraction of the total number of orientations chosen by $Sub24$ over all test images assigned a particular strength. As can be seen from the histogram $Sub24$ chose strengths from the full range. To get some measure of the range of strengths chosen by all subjects, a mean of all 40 histograms was calculated.

This mean histogram is shown in Figure 5-4 and is pretty flat across the strength values. To find the deviations of subjects' histogram from this mean histogram, a variance measure σ_s^2

was calculated for each subject:

$$\sigma_s^2(i) = \frac{1}{40}\sum_{j=0}^{10}(\text{Fraction}_i(j) - \text{AvgFraction(j)})^2 \tag{5.1}$$

where $\sigma_s^2(i)$ corresponds to the variance for $Subi$, $\text{Fraction}_i(j)$ is the histogram value for $Subi$ for strength j, and AvgFraction(j) is the mean histogram value for strength j. Figure 5-4 shows the variance values for all the subject. The horizontal axis is the subject numbers $1-40$. The variances are reasonably low and close to each other except for 5 subjects: $Sub1$, $Sub3$, $Sub5$, $Sub7$, and $Sub9$.

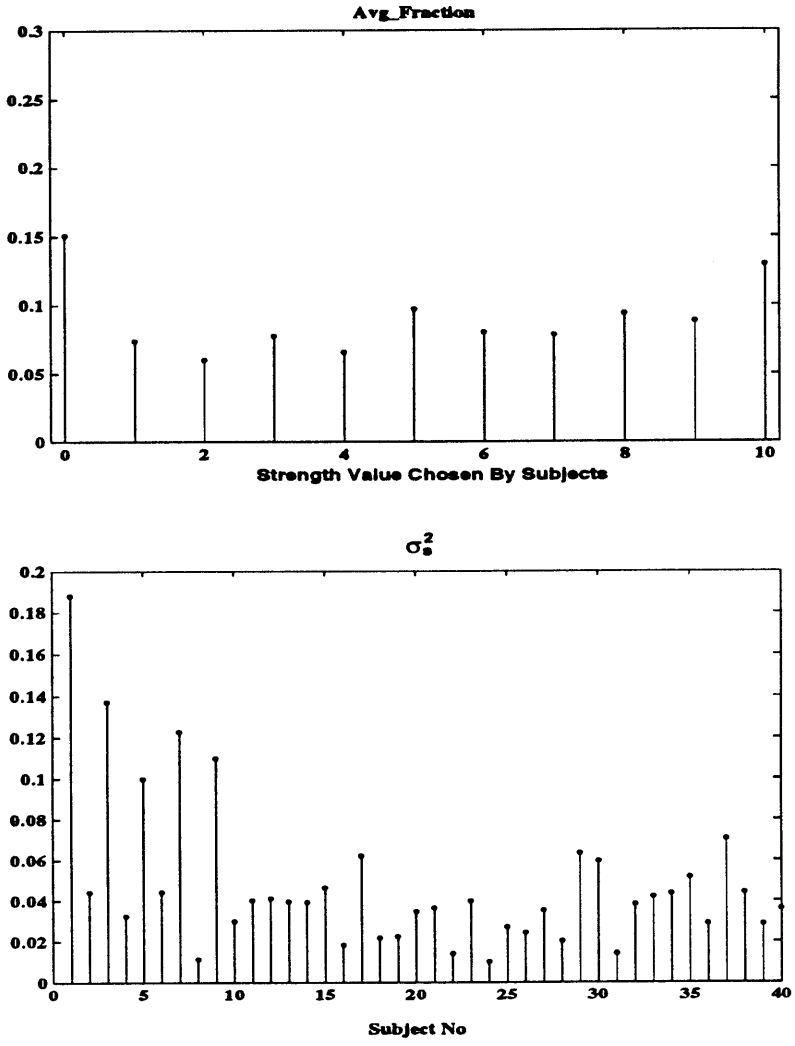


Figure 5-4: Statistics for the distribution of the strengths chosen by all subjects over all test images

Chapter 6

Comparison of Human and Computer Data

The main objective of the human visual experiment was to provide a goal for the orientation finding algorithm to achieve.

It is necessary then to put the human visual data in some compatible format so that it can be compared with the computer data. If there was only one subject, this comparison would be simple: the orientations chosen by the subject had to only be compared with the orientations corresponding to the peaks found by the peak finder described in Chapter 4. However, since there are several subjects there is a spread in the dominant orientations chosen. Also, moving the red bar to indicate orientations introduces errors since not all subjects are equally careful to exactly line up the bar with the orientation they see.

6.1 Organization of Human Visual Data

For the first attempt of comparing the data, the orientations chosen by subjects for each test image will be quantized to 10^o bins. Of course, this kind of quantization will result in a 10^o uncertainty for the dominant orientation. Considering that on test images such as Figure 5-2 (a) the humans' responses spread about 6^o, the 10^o quantization should not be a significant loss. When comparing the human picked orientation with the orientation estimated by the filters this uncertainty should be accounted for. The quantized orientation will be in the range -85^o to 85^o with 10^o intervals.

The strengths associated with all the chosen orientations which fall in a particular bin will be summed together. Three variables will be calculated for each test image : $\vec{\theta}_{h_I}$, $\vec{\gamma}_{h_I}$, and N_I. The vector $\vec{\theta}_{h_I}$ for test image I is defined in the following way:

$$\vec{\theta}_{h_I} = [\theta_{h_1} , \ldots , \theta_{h_{n_I}}] \tag{6.1}$$

where the θ_{h_j} is the jth quantized orientation chosen for test image i and n_I is the total number of quantized orientations chosen for test image i.

The vector $\vec{\gamma}_{h_I}$ for test image I is defined as:

$$\vec{\gamma}_{h_I} = [\gamma_{h_1} , \ldots , \gamma_{h_{n_I}}] \tag{6.2}$$

where γ_{h_j} corresponds to the strength associated with quantized orientation θ_{h_j} divided by the maximum strength that could have been given to θ_{h_j}. The maximum strength is 10 $\times$ the total number of subjects analyzing test image I. If all the subjects selected the quantized angle

θ_{h_j} with strength 10 then $\gamma_{h_j} = 1$; otherwise $\gamma_{h_j} < 1$. The elements of $\vec{\gamma}_{h_I}$ can be viewed as the relative strength of each quantized orientation with respect to the maximum strength that could have been given to that orientation.

Finally, variable N_I is defined to be the number of subjects who specified that the image had zero dominant orientations divided by the total number of subjects responding to that image.

6.2 Organization of Computer Data

It would be desirable for each test image to organize the histogram data similar to the human visual data as described in the previous section. Since we have the histogram data over the original scale and the four steerable pyramid levels, it is important to find a way to combine all this data. It would be desirable to find a threshold where if the salience measure γ described in Section 4.2.2 for peaks in the orientation histograms is below the threshold, then the value represents a non-dominant orientation for the majority of the test images. Since the histogram values over all levels of the steerable pyramid are normalized, they can be compared to each other. For the first step of the comparison, the salience measure γ of each peak was ranked from lowest to highest magnitudes across the levels of the pyramid. The first M highest salience measures and their associated orientations were compared with the human visual data.

The average size of vector $\vec{\theta}_{h_I}$ over all test images is 5. The average number of orientations picked by subjects over all the images was 1.3. The maximum number of orientations picked by a subject was 7. Only one subject chose 7 orientations. Therefore M was chosen to be equal to 5. Since some orientations exist over more than one pyramid level, they were not picked more than once when choosing the highest M salience measures. It was found by looking at

images where orientations existed in more than one level of the pyramid that an allowance of a 10^o spread insured that orientations were not picked more than once in the ranking. A 10^o spread means that in the ranking, an orientation is not included if it is is closer than 10^o to the orientations picked in the other steerable pyramid levels.

Two variables $\vec{\theta}_{c_I}$ and $\vec{\gamma}_{c_I}$ are defined for each test image I :

$$\begin{aligned} \vec{\theta}_{c_I} &= [\theta_{c_1}, \ldots, \theta_{c_M}] \\ \vec{\gamma}_{c_I} &= [\gamma_{c_1}, \ldots, \gamma_{c_M}] \end{aligned} \tag{6.3}$$

where θ_{c_j} corresponds the jth orientation whose salience measure denoted by γ_{c_j} was included in the top M measures. A summary of the variables used to represent human and computer data is shown in Table 6.1. The objective is then to compare the human data with the data from the computer algorithm, i.e., comparing $\vec{\theta}_{h_I}$ to $\vec{\theta}_{c_I}$ and $\vec{\gamma}_{h_I}$ to $\vec{\gamma}_{c_I}$ over each test image I.

6.3 Comparison Between Orientations Picked by Humans and Estimated by Filters

For the initial comparison the cases summarized in Table 6.2 were considered. The new vector $\vec{\theta}^M_{h_I}$ consists of the elements of $\vec{\theta}_{h_I}$ (orientations humans picked) for which matches were found in $\vec{\theta}_{c_I}$. For the present, this will be all orientations which are within 10° of those found by the filters. The vector $\vec{\theta}^R_{h_I}$ consists of all other elements of $\vec{\theta}_{h_I}$. Clearly it is desirable that $\vec{\theta}^M_{h_I} = \vec{\theta}_{h_I}$, i.e., that the computer found all the orientations important to the humans. The

Human response data:

$\vec{\theta}_{h_I}$	the vector of orientations chosen by the humans
$\vec{\gamma}_{h_I}$	the corresponding vector of strengths
N_I	a measure of how non-directional the image is perceived to be.

Computer algorithm output data:

$\vec{\theta}_{c_I}$	vector of M orientations chosen from all pyramid levels
$\vec{\gamma}_{c_I}$	the corresponding vector of salience measures

Table 6.1: Notation for recording human and computer outputs.

Human picked orientations		**Computer picked orientations**	
Matched to $\vec{\theta}_{c_I}$	*Rejected*	*Matched to* $\vec{\theta}_{h_I}$	*Rejected*
$\vec{\theta}^M_{h_I}$	$\vec{\theta}^R_{h_I}$	$(\vec{\theta}^M_{c_I})$	$(\vec{\theta}^R_{c_I})$
$(\vec{\gamma}^M_{h_I})$	$\vec{\gamma}^R_{h_I}$	$\vec{\gamma}^M_{c_I}$	$\vec{\gamma}^R_{c_I}$

Table 6.2: Notation for comparisons of human and computer outputs.

values in $\vec{\gamma}_{c_I}^R$, the salience measures of orientations which were found by the computer algorithm but did not correspond to ones humans found, should not be considered "good enough". The mean of $\vec{\gamma}_{c_I}^R$ over all the images was then used for the initial salience threshold, $\gamma_c = 0.085$ to decide if a peak is a dominant orientation. A few cases where $\vec{\gamma}_{h_I}^R$ is important (the humans found orientations that the computer could not match) are indicated in the "difficult cases" in Section 6.3.3. The other values above which are in parentheses are not necessary in the results reported here.

To evaluate the threshold γ_c, a measure of the "wrong" rejections caused by γ_c was considered:

$$r_c = \text{Number of Elements of } \vec{\gamma}_{c_I}^M < \gamma_c, \tag{6.4}$$

It is of course desirable that $r_c = 0$. A rejection measure can also be formed for the human data to see which human picked orientations with strengths bigger than threshold γ_h were not picked:

$$r_h = \text{Number of Elements of } \vec{\gamma}_{h_I}^R > \gamma_h, \tag{6.5}$$

where γ_h is a (typically very low) value used to ignore some stray low-strength orientations from the human data.

There is a need for threshold γ_h on the strength of the orientations perceived by the humans. Ideally we would want to match all the dominant orientations perceived by the humans. However, in some cases, the images were not very directional and the subjects gave very low

values to the orientation they perceived. It is not as important, therefore, to match these orientations as it is to match the strong dominant orientations. Choosing a reasonable threshold is a problem. Since γ_{h_j} values are relative measures of the strength of the orientation with respect to the maximum strength it can be given, then γ_h can represent a percentage below which the relative strength is too small. $\gamma_h = 0.15$ was chosen. In essence choosing this value for the threshold means that the strength of γ_{h_j} should be at least 15% of the maximum strength that can be assigned to it.

Measures r_c and r_h were found for all the test images. Figures 6-1-6-9 show the test images organized in groups with the same number of dominant orientations corresponding to the elements in $\vec{\gamma}_{c_I}^{M}$ greater than the threshold γ_c. The letter A underneath some of the images indicates that both $r_c = 0$ and $r_h = 0$ with $\gamma_h = 0.15$ (the number and position of the orientations were matched exactly). The results of this are shown in the first line of Table 6.3. A total of 59/111 of the images agree 100%.

For the agreements shown in Figures 6-1 and 6-2, except for three cases, more than 50% ($N_I > 0.5$) of the subjects indicated the textures to be non-directional. This means that the particular threshold γ_h was effective in separating the clearly dominant orientations from the weak orientations. For *test*74, 40% ($N_{74} = 0.4$) of the subjects indicated no dominant directions; for *test*2, 30% ($N_2 = 0.3$) of the subjects indicated no dominant orientations and for *test*40, 40% ($N_{40} = 0.4$) indicated no dominant orientations. It is interesting that the computer found no dominant orientations for *test*75 and *test*66 while the majority of subjects indicated dominant orientations for these textures, $N_{75} = 0.8$ and $N_{66} = 0.75$. It appears that there may be some gestalt grouping occurring among the objects in these images. Looking at the results in Figures 6-1-6-9 several interesting cases were found.

Number of dominant Orientations:	0	1	2	3	4
With fixed threshold, γ_c:	$\frac{23}{32} = 72\%$	$\frac{16}{43} = 37\%$	$\frac{17}{28} = 61\%$	$\frac{1}{4} = 25\%$	$\frac{2}{3} = 67\%$
With different γ_{c0} - γ_{c3}:	$\frac{32}{41} = 78\%$	$\frac{18}{44} = 41\%$	$\frac{15}{19} = 79\%$	$\frac{1}{4} = 25\%$	$\frac{2}{2} = 100\%$
Gain of A's (agreement) (+):	+9	+2	0	0	0
Loss of A's (−):	-1	-1	-2	0	0

Table 6.3: Comparison of results using one threshold and using level-dependent thresholds. Values in ratios are (number of patterns which agree)/(number of patterns computer found with that orientation). Agreement is measured between human study data and computer algorithm.

Case 1: Effect of γ_c on choosing peaks at different levels of pyramid For coarse contrasty images like *test*60 and *test*99 , there were prominent peaks in the orientation histograms in the third level of the steerable pyramid. Figure 6-10 shows the orientation histograms for *test*60 and *test*99 at the third level of the pyramid. The peaks at 70^o for *test*60 and the peaks at 10^o for *test*99 did correspond to the orientations picked by humans. However the strength values given to these orientations by humans were very small yet the salience measures for these peaks are much larger than γ_c.

Images like *test*52 and *test*17 illustrate the opposite difficulties in having one value for γ_c. The peaks corresponding to orientations 45^o and -45^o in *test*52 and *test*17 are given relatively large strength values by humans. However their salience measures are much smaller than γ_c. Figure 6-11 shows the orientation histogram for *test*17 at level 1 of the steerable pyramid and

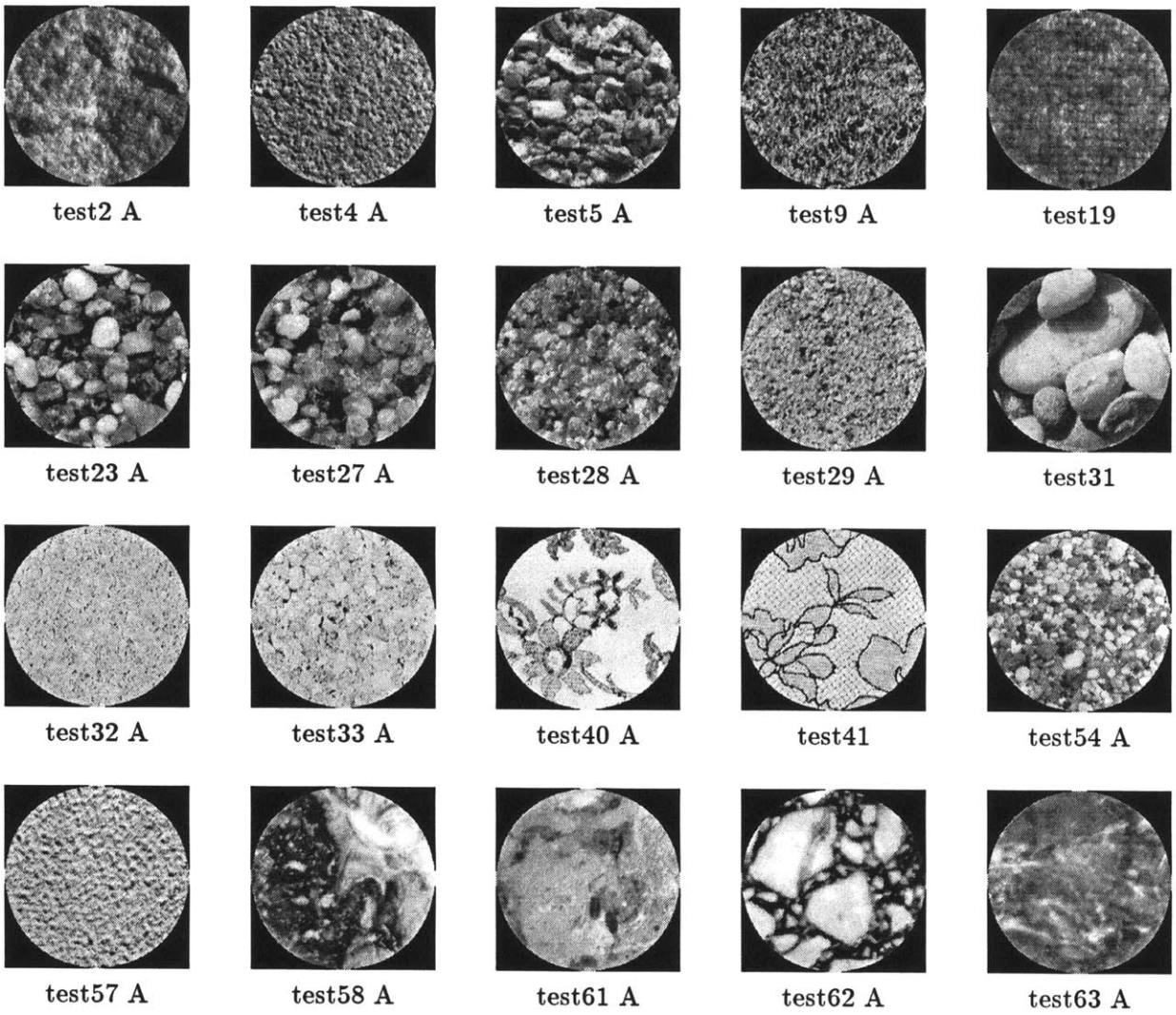


Figure 6-1: No dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data

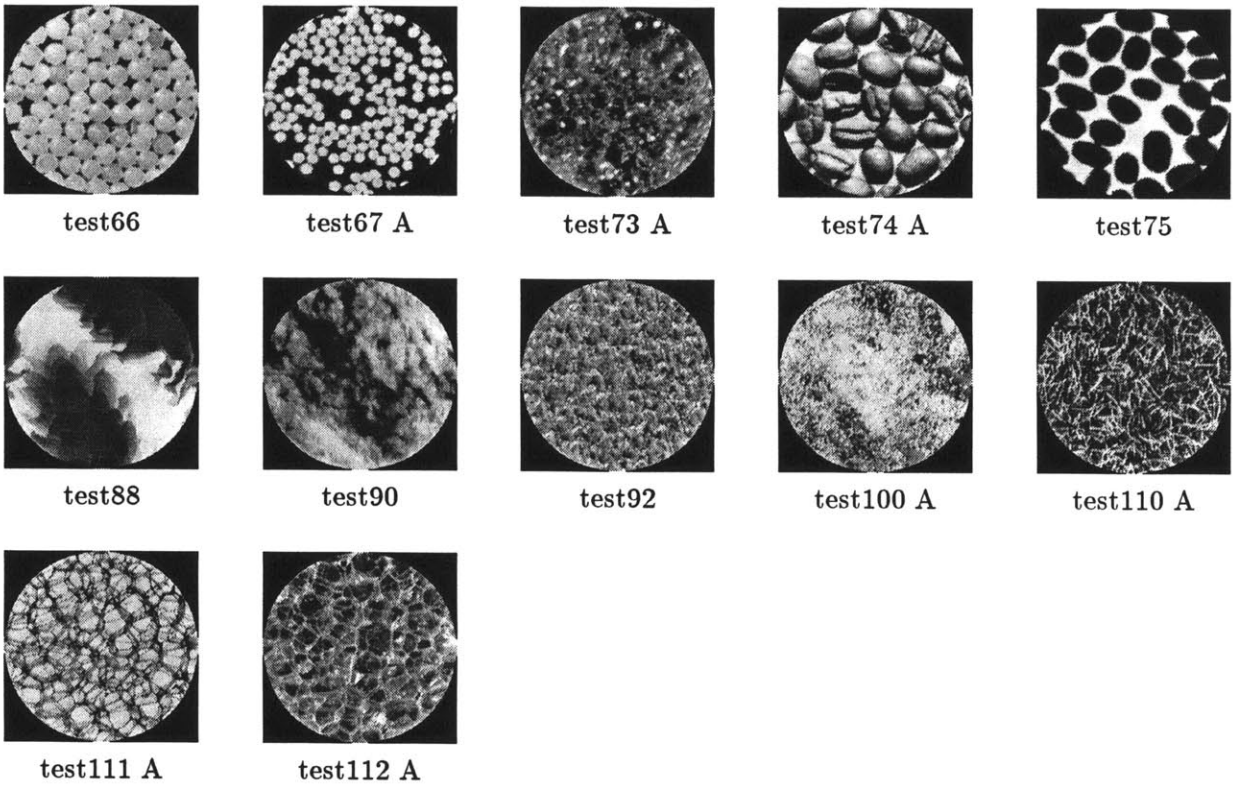


Figure 6-2: Figure 6-1 continued.

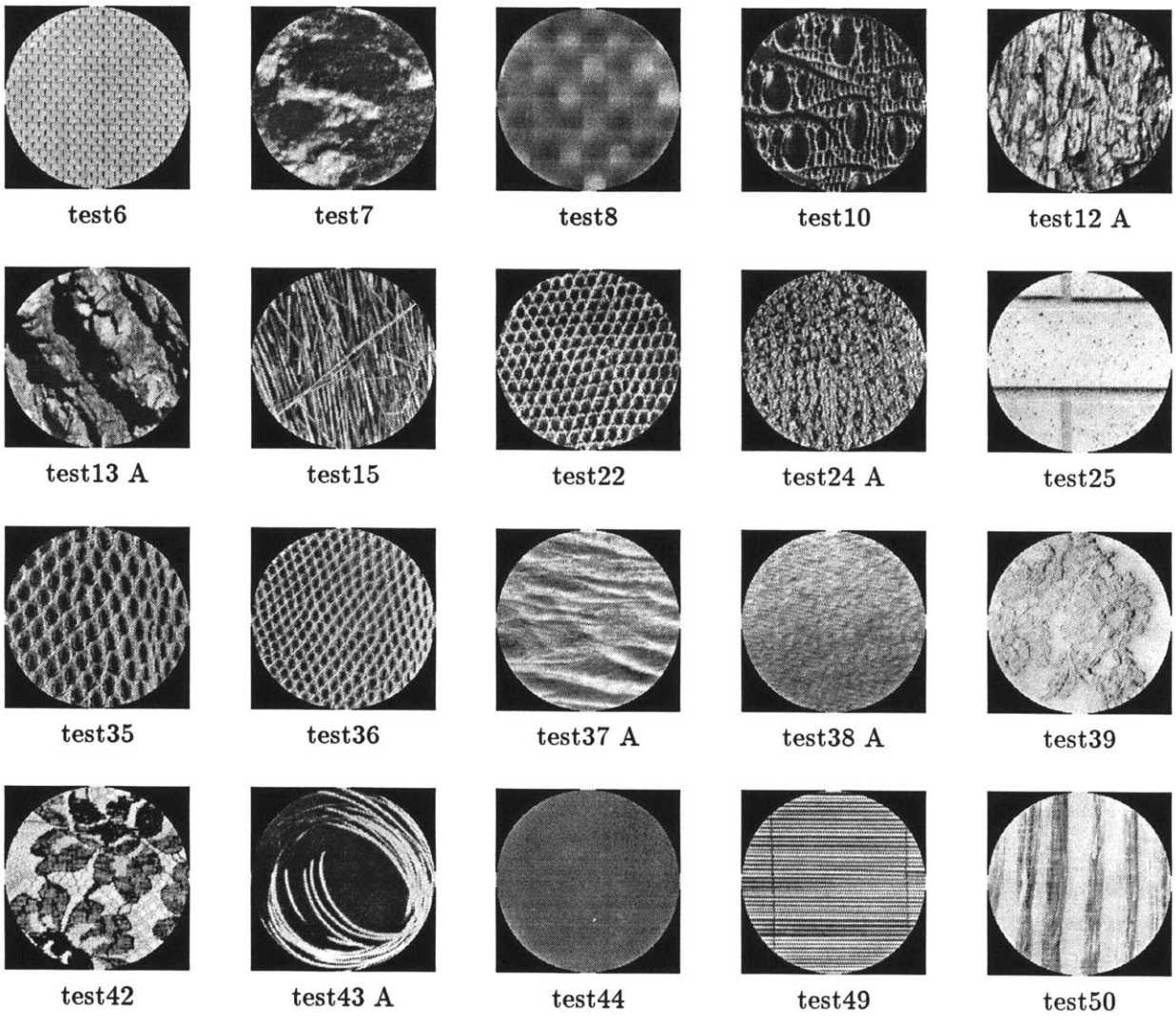


Figure 6-3: One dominant orientation was found by the computer in these. An "A" under an image indicates agreement with the human study data

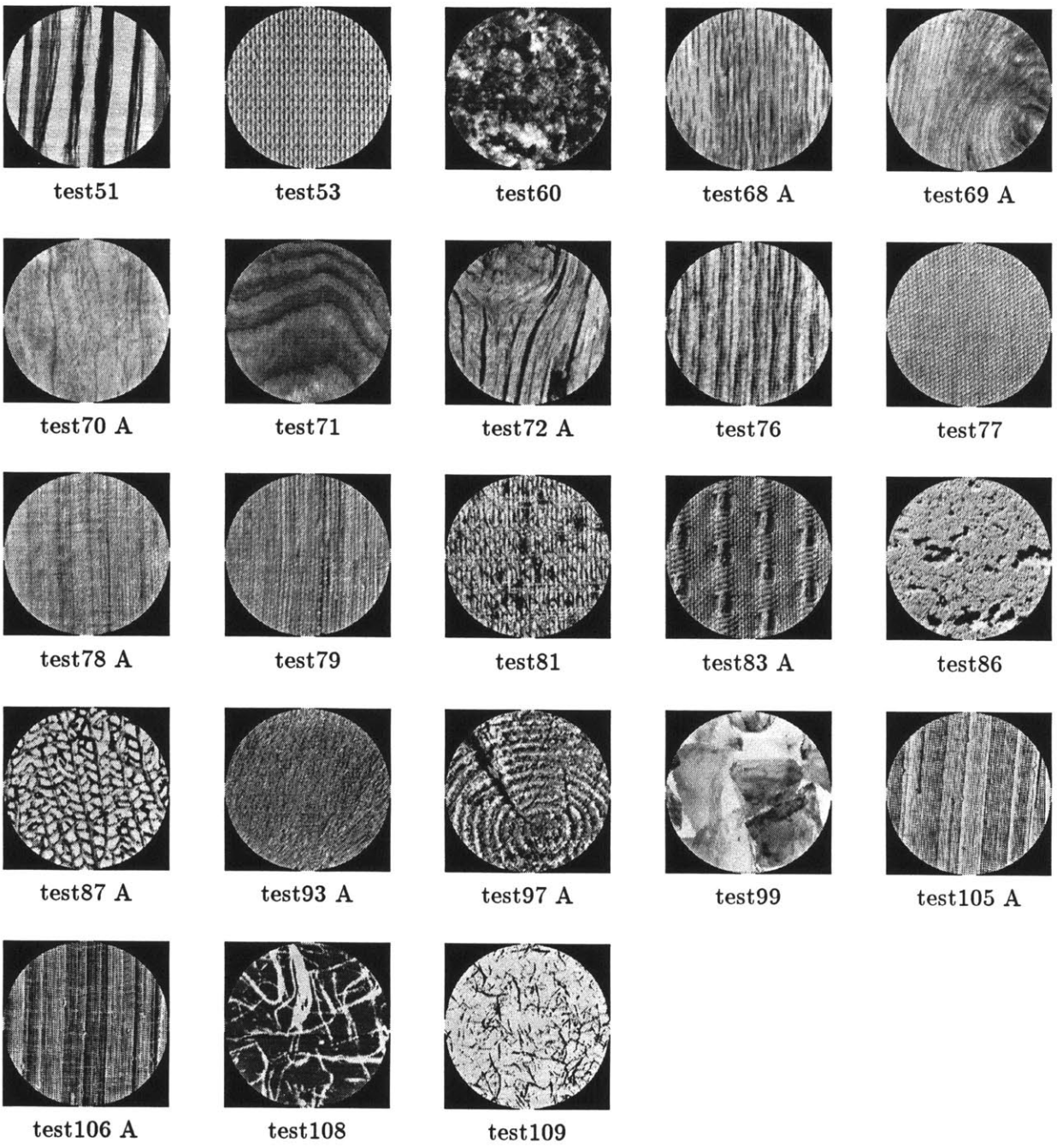


Figure 6-4: Figure 6-3 continued.

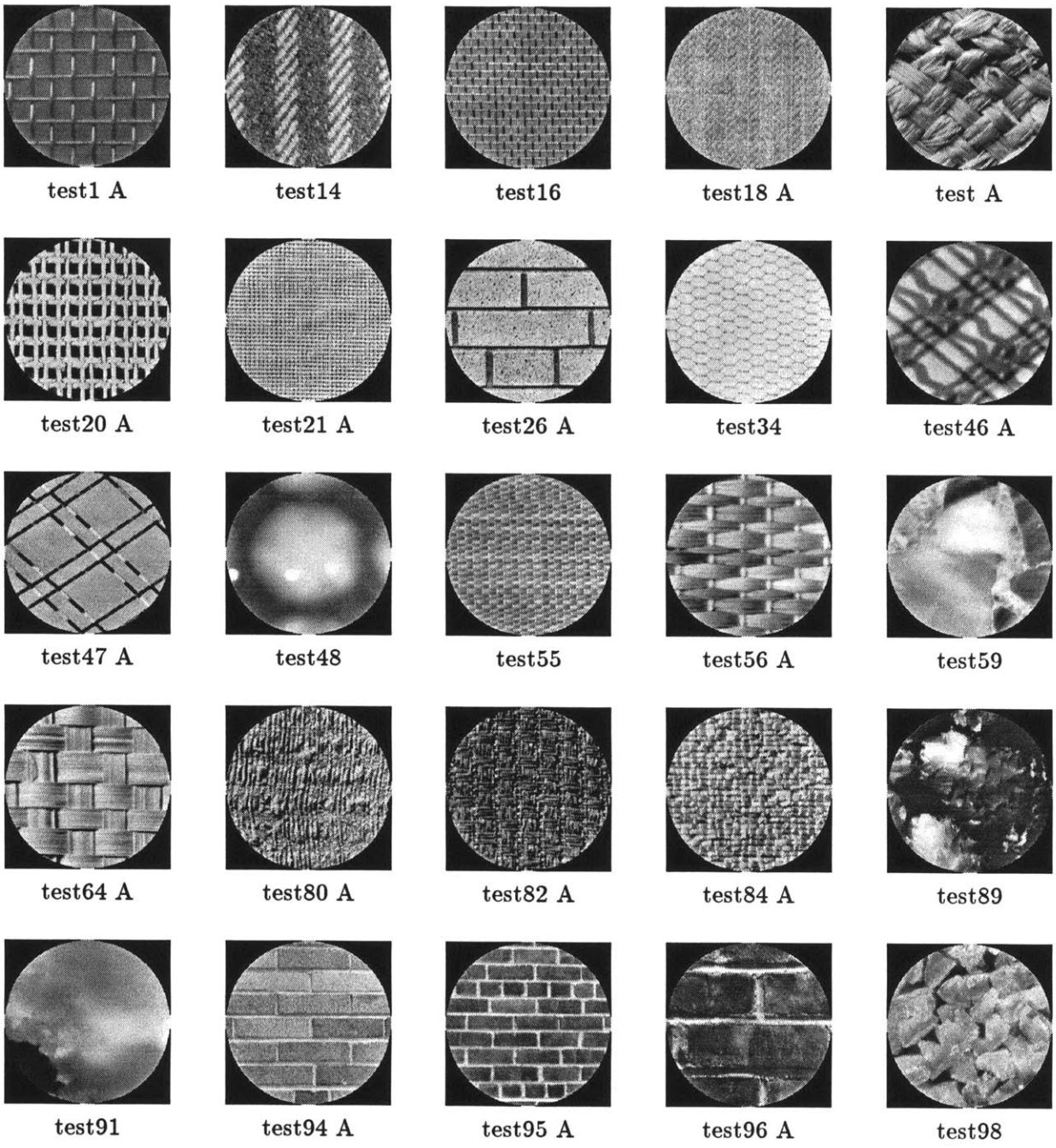
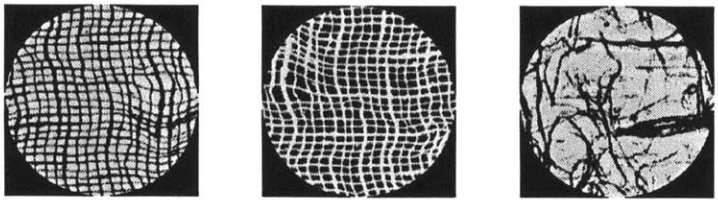
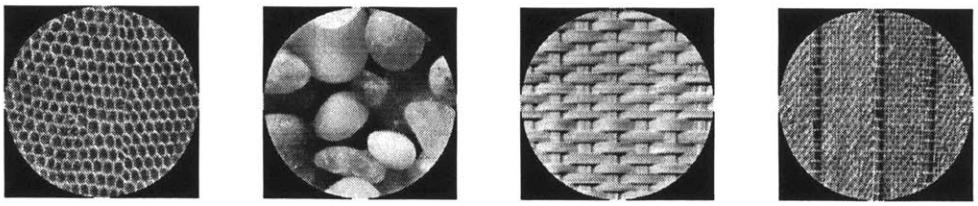


Figure 6-5: Two dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



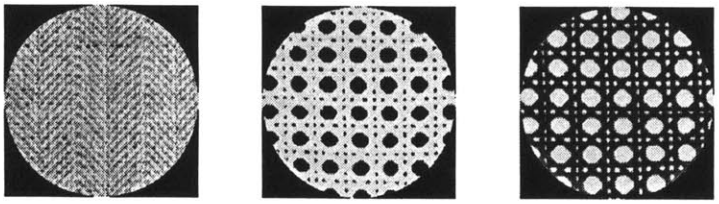
test103 A test104 A test107

Figure 6-6: Figure 6-5 continued



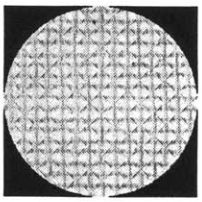
test3 A test30 test65 test85

Figure 6-7: Three dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



test17 test102 A test101 A

Figure 6-8: Four dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



test52

Figure 6-9: Five dominant orientations were found by the computer for this image.

for *test*52 at level 0 of the steerable pyramid. In the case of *test*17, the parallel lines along -45^o have a lower contrast than the lines along 45^o ; therefore the peak at 45^o is less prominent than the one at -45^o. In the case of *test*52, the spacing between the lines oriented at -45^o and 45^o and the horizontal line is very small so there is a greater probability that there will be interference between these lines when estimating the orientations; therefore, the salience measures for these orientations are much smaller than γ_c.

Several cases like the four described above imply that a fixed threshold value across all levels of the pyramid will not enable the human results to be closely matched. To avoid choosing the peaks at the highest level shown in cases like test60 and test99, a higher threshold than γ_c should be chosen for the peaks in the third and second level where the contrasty coarse elements will be prominent. In the first two levels, there is a problem with spacing of the structures as in the case of *test*52 and contrast as in the case of *test*17. Therefore, a lower threshold than γ_c should be chosen.

To find the new threshold for level 3, we looked at range of the salience measures for some of the coarse contrasty textures, then picked a value for which the majority of peaks in level 3 either not matching the orientations chosen by humans or being given small strengths by humans smaller than γ_h will not be chosen by the computer. A threshold for level 2 was picked in the same way. For level 0 and 1 thresholds were picked enabling the majority of peaks with contrast or spacing problems to be chosen. This threshold picking was repeated several iterations until the "best balance" was reached between results agreeing with human data and those that did not agree.

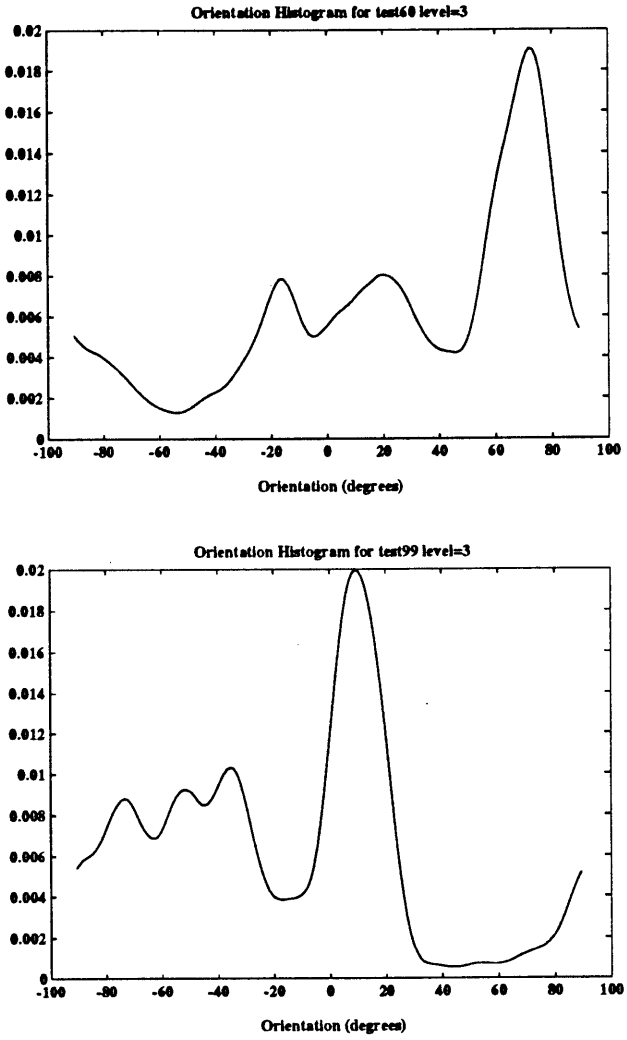


Figure 6-10: Orientation histograms of *test*60 and *test*99.

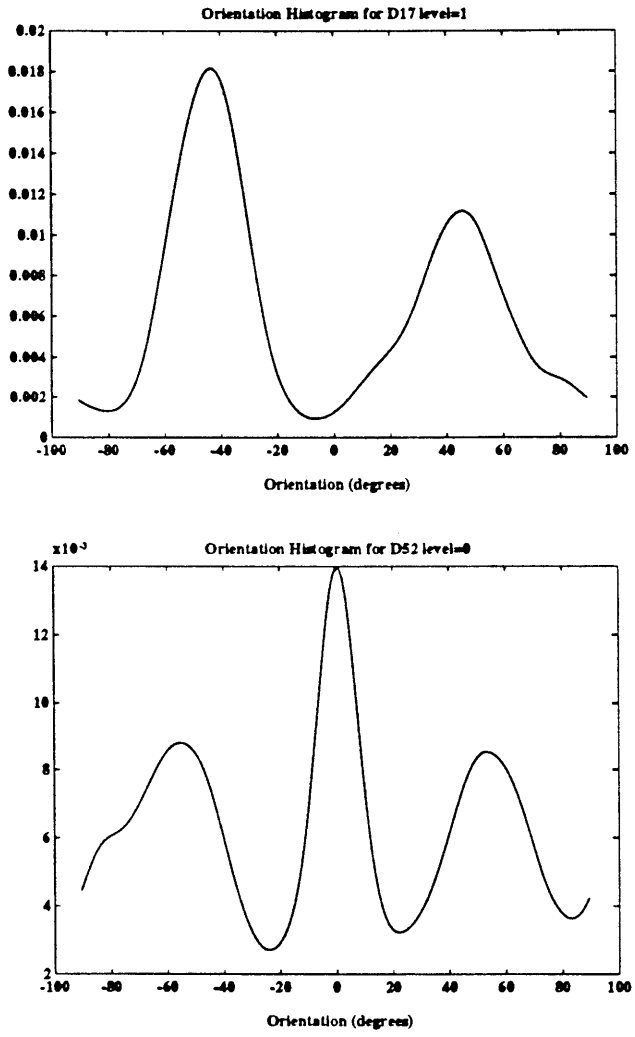


Figure 6-11: Orientation histograms of *test*17 and *test*52.

Case 2: Broad peaks with sharp slopes There were cases where the peak in the orientation histogram was broad but the slopes of the curve were sharp. Figure B-1 in Appendix B gives an example of this for the orientation histogram of *test*109 in the level 0 of the pyramid. The peak at 0^o has the same salience measure as the -45^o peak for *test*52 shown in Figure 6-11. Even though 70% of the subjects indicated that they did not see any dominant orientations in *test*109 and they gave the -45^o orientation for *test*52 a large strength. The salience measures for the two are the same since even though the peak for *test*109 is much broader than the -45^o peak for *test*52, it has a bigger value and a steeper slope. Very broad peaks are usually caused when there are structures at many orientations. These peaks in our experience did not coincide with the orientations perceived strongly by humans. To make sure that these peaks are not considered to be prominent, a weight w_{br} is incorporated in the calculation for γ (4.5) for peaks which are broader than $\Delta\theta_{br}$. Different values for w_{br} and $\Delta\theta_{br}$ were tried taking in the consideration the human data . The values 0.10 and 72^o were finally chosen for w_{br} and $\Delta\theta_{br}$ respectively. This weight was incorporated in the salience measure before choosing the thresholds for the levels of the steerable pyramid.

6.3.1 Results with the Different Thresholds Assigned to the Levels of the Steerable Pyramid

Figures 6-13 - 6-20 show the Brodatz textures grouped by the computer by their number of dominant orientations using different thresholds at different levels. The values for the thresholds $\gamma_{c0} - \gamma_{c3}$ for the pyramid $level = 0, 1, 2, 3$ are given in Table 6.4. The number of agreements in this case vs. the case of a fixed threshold are summarized in Table 6.3. Each column of the

tf_0	0.418
tf_1	0.075
tf_2	0.034
tf_3	0.0098

Table 6.4: Pyramid level-dependent threshold values.

---

third row indicates how many patterns agreed in row 2 that did not agree in row 1 (+), and how many patterns disagreed in row 2 that had agreed in row 1 (−). If the thresholds are overly conservative, one might find all the entries for (+) will be large and all for (−) would be zero. Intuitively, as the two start to balance, one would expect the thresholds are nearer their critical points. The following observations were made made when looking at the disagreements :

Observation1: There were many textures which were classified to have one dominant orientation even though people chose two orientations. These textures include: *test*18, *test*25, *test*50, *test*51, *test*53, *test*79, *test*80, *test*96. In the case of *test*79, the horizontal orientation detected by humans with relative strength 0.23 does not exist in the orientation histogram. For the other test images, the orientation chosen by humans has a corresponding peak in the orientation histogram but this peak is less than 20% of the other dominant peak. In the case of *test*50 and *test*51, humans picked the horizontal orientation with a relative strength of 0.22 and 0.27 respectively but the peak associated with this orientation is less than 1% of the value of the vertical orientation. It seems that people can detect the low contrast horizontal line much

better than the computer. Next section shows that incorporating the contrast normalized orientation histograms described in Section 4.3.4 in the peak detection can allow many of these small peaks to be picked up by the computer.

Observation2: The vertical orientation in *test*49 was chosen by humans with a relative strength of 0.32 and the orientation of one piece of grass in *test*15 was chosen with a relative strength of 0.58. In the orientation histogram for *test*15, the vertical orientation is less than 0.1% of the peak corresponding to the horizontal lines. In the orientation histogram of *test*15 the peak corresponding to the orientation of the piece of grass was less than 6% of the values of the peak corresponding to the direction of surrounding grass. Since both the piece of grass in *test*15 and the vertical lines in *test*49 cover only a small area, they can not be picked up by a filtering method over the entire area. However, they will be prominent if the orientation finding algorithm is performed in a smaller region.

Observation3: For the lizard skin patterns *test*22 and *test*35, there were three human picked orientations with relative strengths much bigger than γ_h. In the orientation histogram of *test*22 at level 0, there are three peaks but two of them are very broad and did not even look visually prominent as seen in Figure 6-12. For *test*36, there were two human picked orientations with relative strengths bigger than γ_h; however looking at the orientation histogram of *test*36 in Figure 6-12, the peak near -45° is very broad and does not look as prominent as the peak at 0°. If the threshold is set so that the peak -45° is chosen, then there will be a number of broad peaks in other patterns which will be picked corresponding to no dominant orientation.

One of the reasons that some of the peaks are not very prominent for textures (and perhaps the key reason) is due to their non-homogeneity. The humans may focus attention on a smaller

region which is homogeneous. Another problem can be that in a local isotropic region, these textures do not look very directional; however a much narrower and longer filter might be able to pick up the orientations.

Observation4: Very contrasty textures like *test*98, *test*107 and *test*108 cause the filter outputs to be big, leading to prominent peaks in the histograms even though 50% of the subjects chose *test*108 to be non-directional, 70% chose *test*98 to be non-directional and 80% chose *test*107 to be non-directional. Most of the high contrast coarse textures with no perceptually dominant orientation were in agreement with computer algorithm because of the higher thresholds at the last two levels of the pyramid. However, in the case of *test*98, *test*107 and *test*108 they had prominent peaks in the level 0 of the steerable pyramid where the threshold is chosen to be the smallest; hence the algorithm assigned them one dominant orientation.

Subjects found *test*42 which is made up of a complicated pattern having many different orientations somewhat directional (60 % chose at least one orientation for it) but gave very small relative strengths. However because of high contrast nature of the image, there is a prominent peak which does not correspond to any of the orientations chosen by the subjects. For textures like this because of the complicated pattern and the high contrast, the filter outputs may be large leading to peaks which are not perceptually prominent. Images like *test*42 are considered perhaps too directional for this low-level method.

For some of the disagreements like *test*35 different shaped filters might have to be used. However, it would be interesting to see whether incorporating the contrast normalized orientation histogram effects the results in any way.

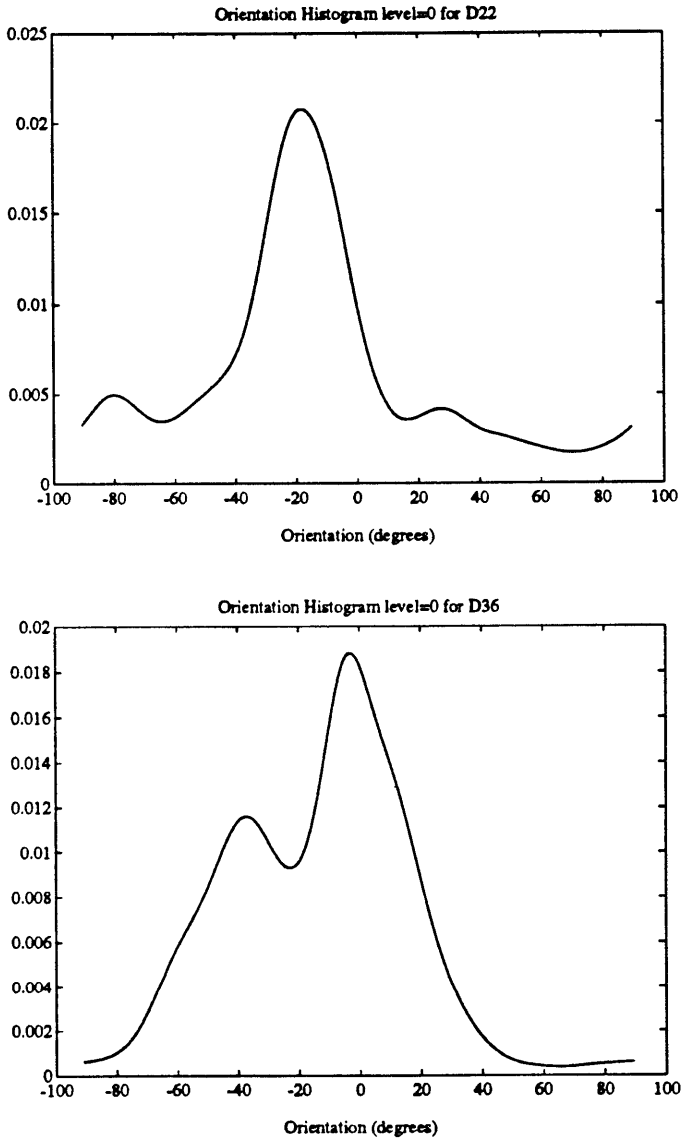


Figure 6-12: Orientation histograms for *test*22 and *test*36.

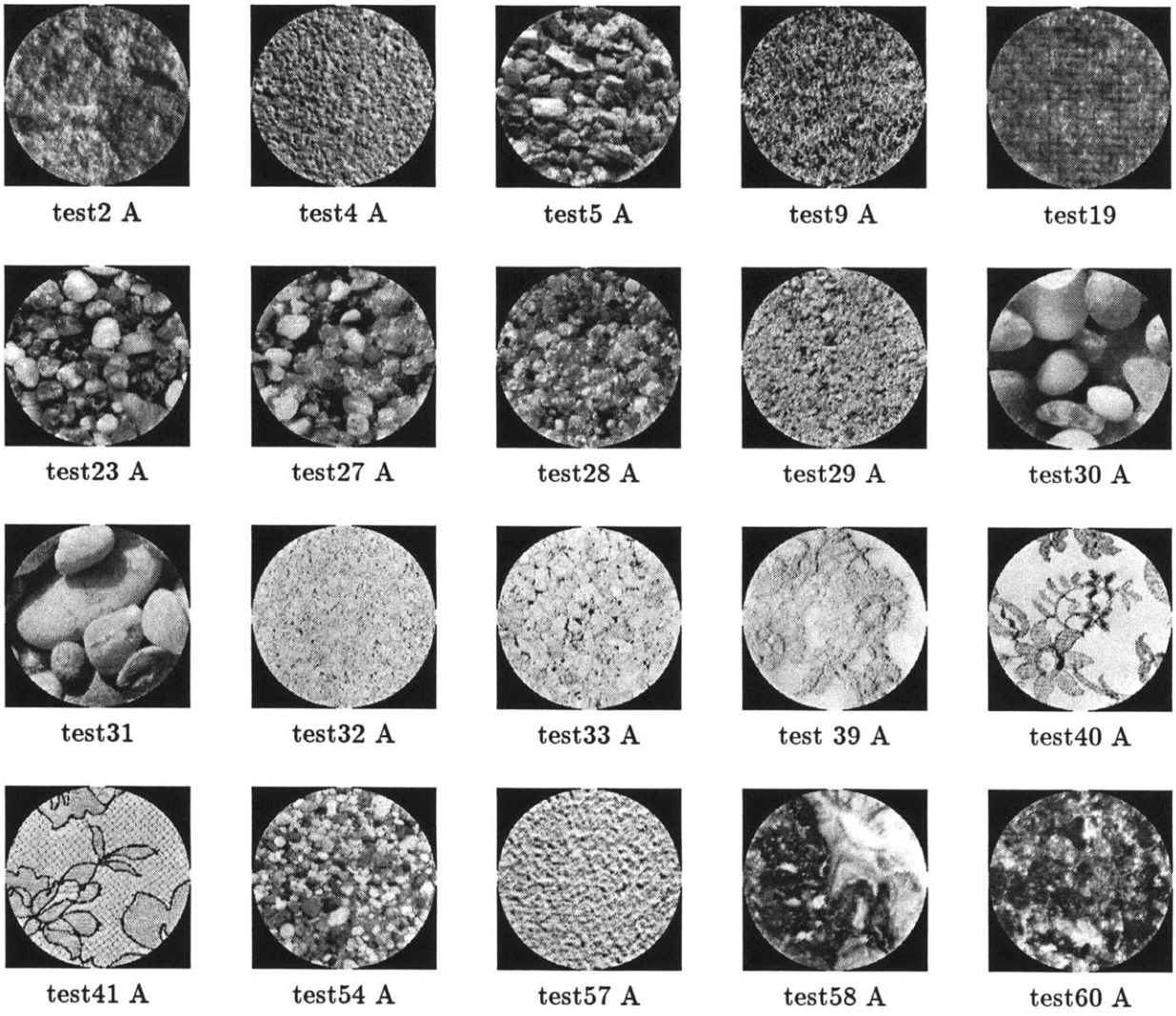


Figure 6-13: No dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data

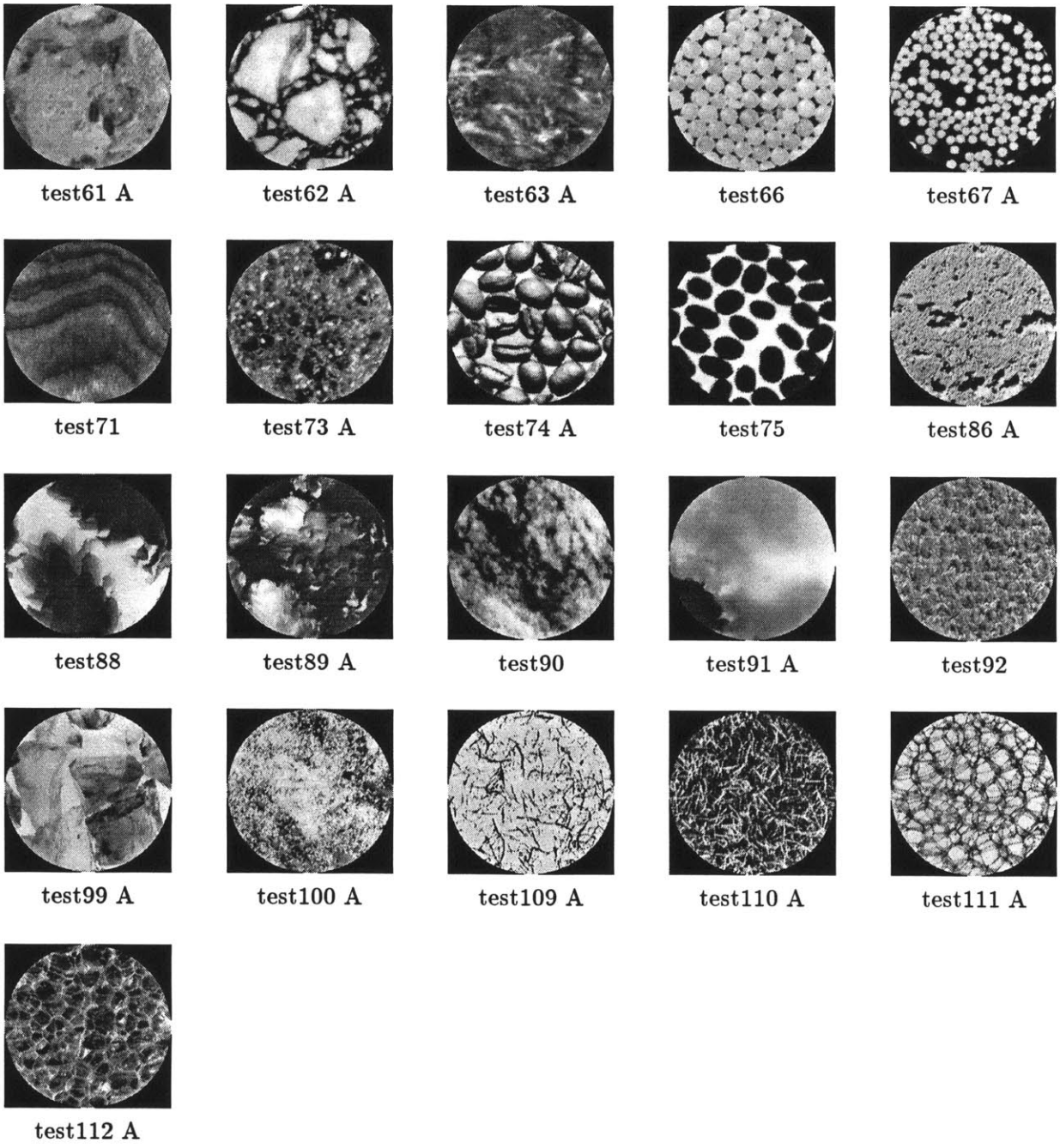


Figure 6-14: Figure 6-13 continued.

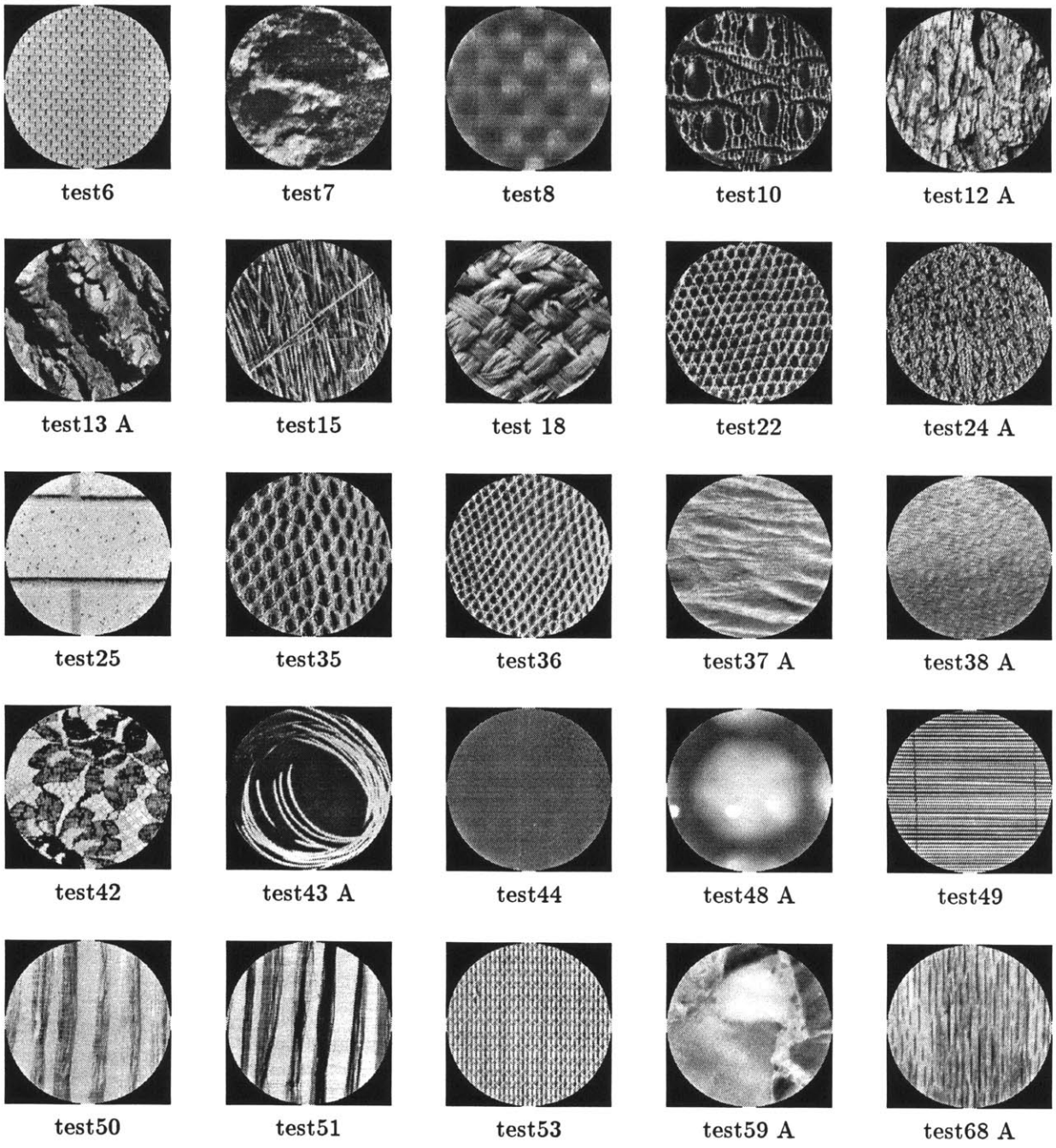


Figure 6-15: One dominant orientation was found by the computer in these. An "A" under an image indicates agreement with the human study data

Number of dominant orientations:	0	1	2	3	4
With different $\gamma_{c0} - \gamma_{c3}$ No contrast normalization	$\frac{32}{41} = 78\%$	$\frac{18}{44} = 41\%$	$\frac{15}{19} = 79\%$	$\frac{1}{4} = 25\%$	$\frac{2}{2} = 100\%$
With different $\gamma_{c0} - \gamma_{c3}$ Contrast Normalized	$\frac{32}{41} = 78\%$	$\frac{18}{36} = 50\%$	$\frac{21}{27} = 78\%$	$\frac{1}{4} = 25\%$	$\frac{2}{2} = 100\%$
Gain of A's (agreement) (+):	0	0	+6	0	0
Loss of A's (−):	0	0	0	0	0

Table 6.5: Comparison of results using different thresholds γ_{c0} and γ_{c1} without contrast normalization and without contrast normalization. Values in ratios are (number of patterns which agree)/(number of patterns computer found with that orientation). Agreement is measured between human study data and computer algorithm.

6.3.2 Results from Incorporating the Contrast Normalized Histogram

A summary of the improved results obtained using the multiple thresholds above, and also compensating for contrast, are given in Table 6.6.

For initial analysis, only the contrast normalized histograms of level 0 were incorporated. Using the contrast normalized histogram is tricky since it greatly boosts low contrast structures which may or may not be very directional; therefore it should be used selectively. The rule we used was: *At least two peaks must correspond between the original histogram at level 0 and the contrast normalized histogram and one of the peaks must have a salience measure above the threshold and the other peak a salience measure below the threshold.* If the condition is true and the ratio of the peaks values is less than 20% in the original histogram and more than 80% in the contrast normalized histogram then the weak peak is chosen as a dominant orientation.

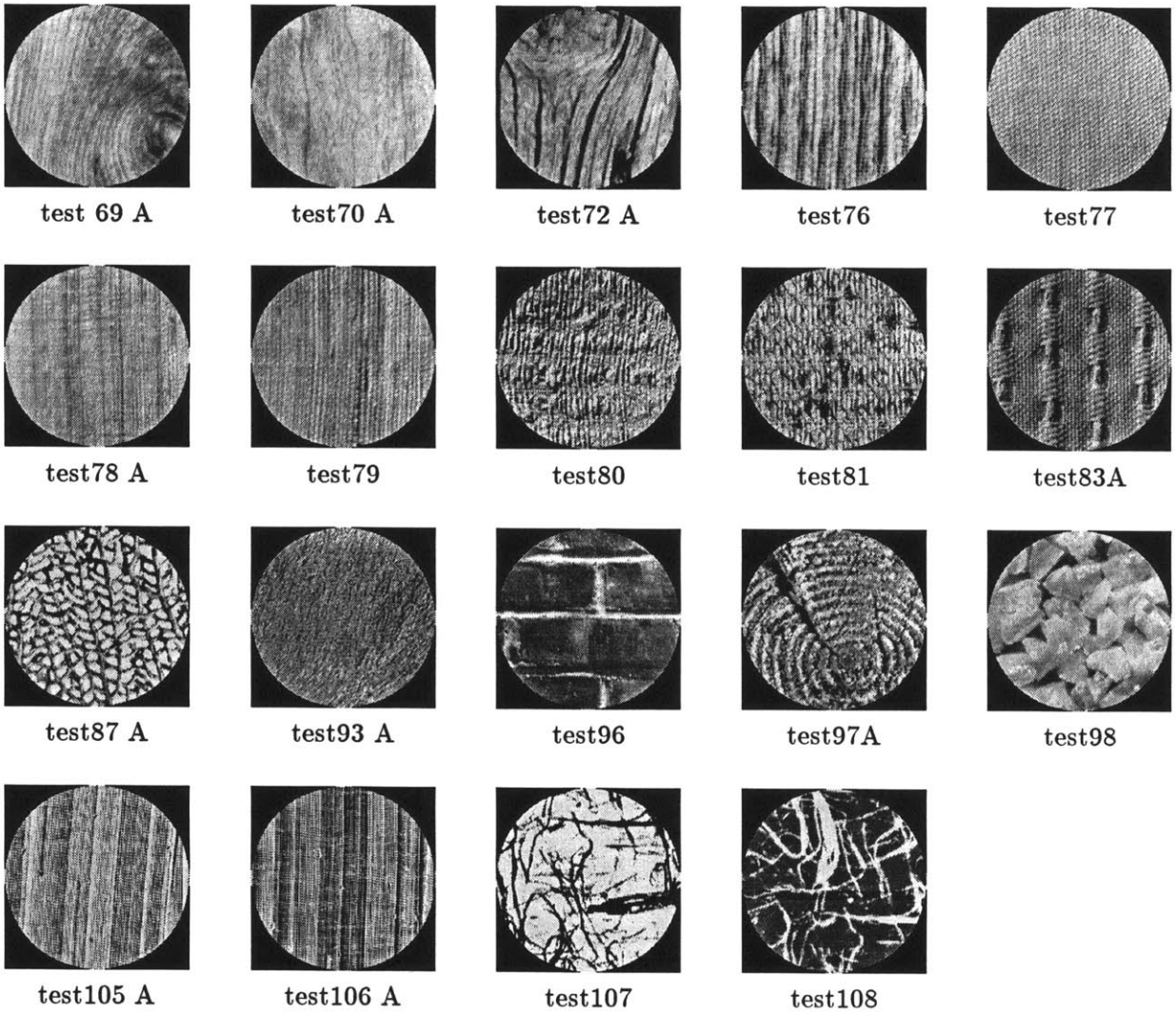


Figure 6-16: Figure 6-15 continued.

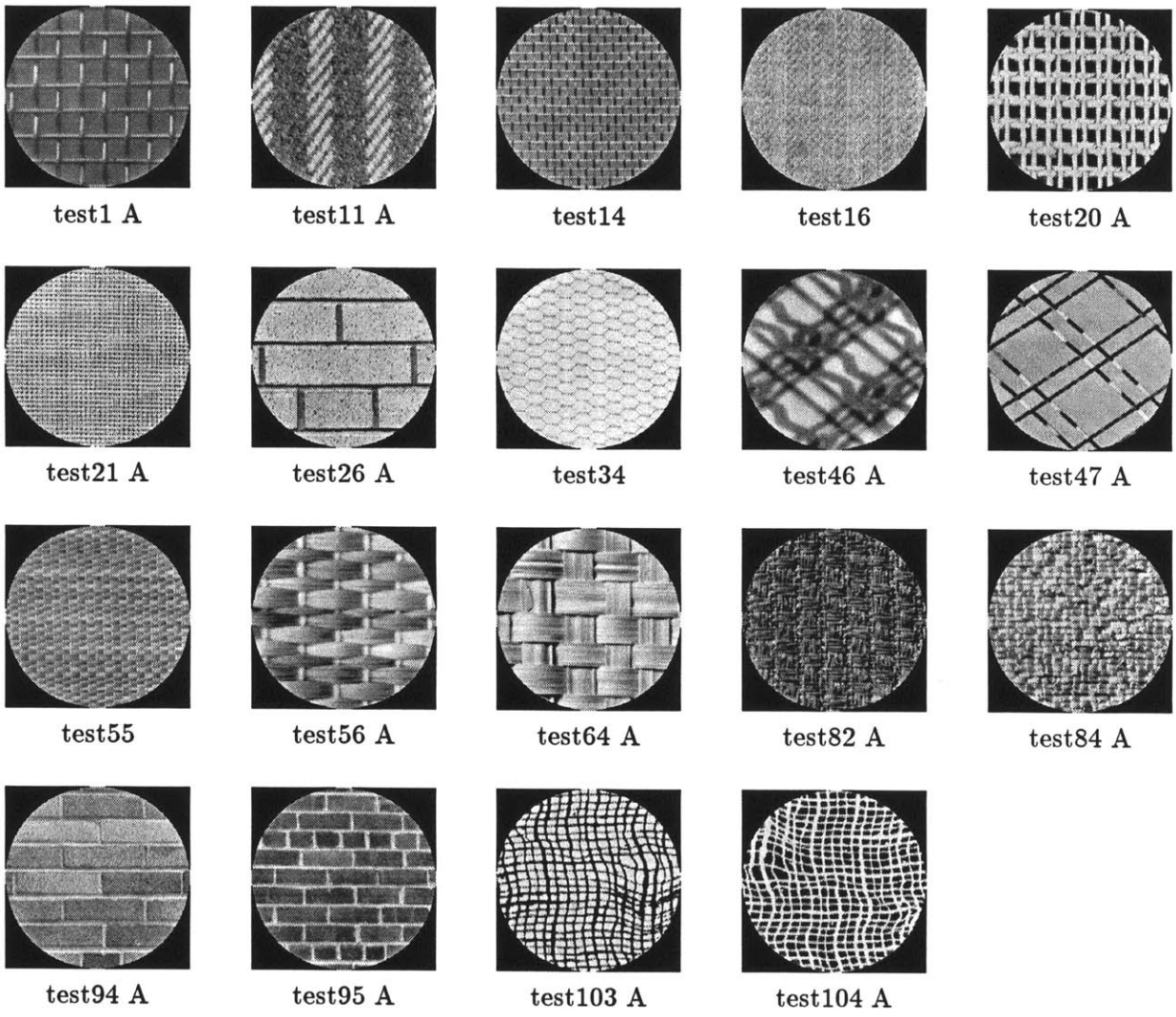
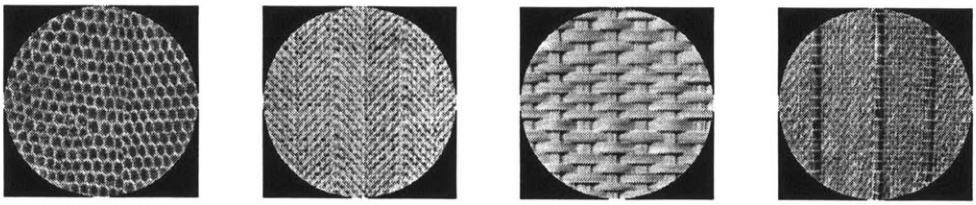
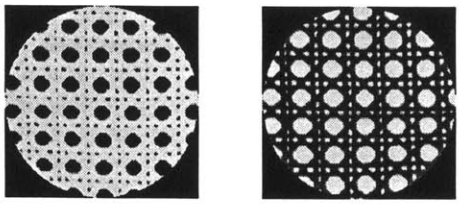


Figure 6-17: Two dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



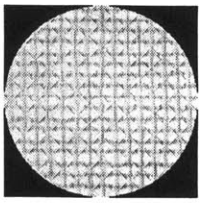
test3 A test17 test65 test85

Figure 6-18: Three dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



test102 A test101 A

Figure 6-19: Four dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data



test52

Figure 6-20: Five dominant orientations were found by the computer for this image.

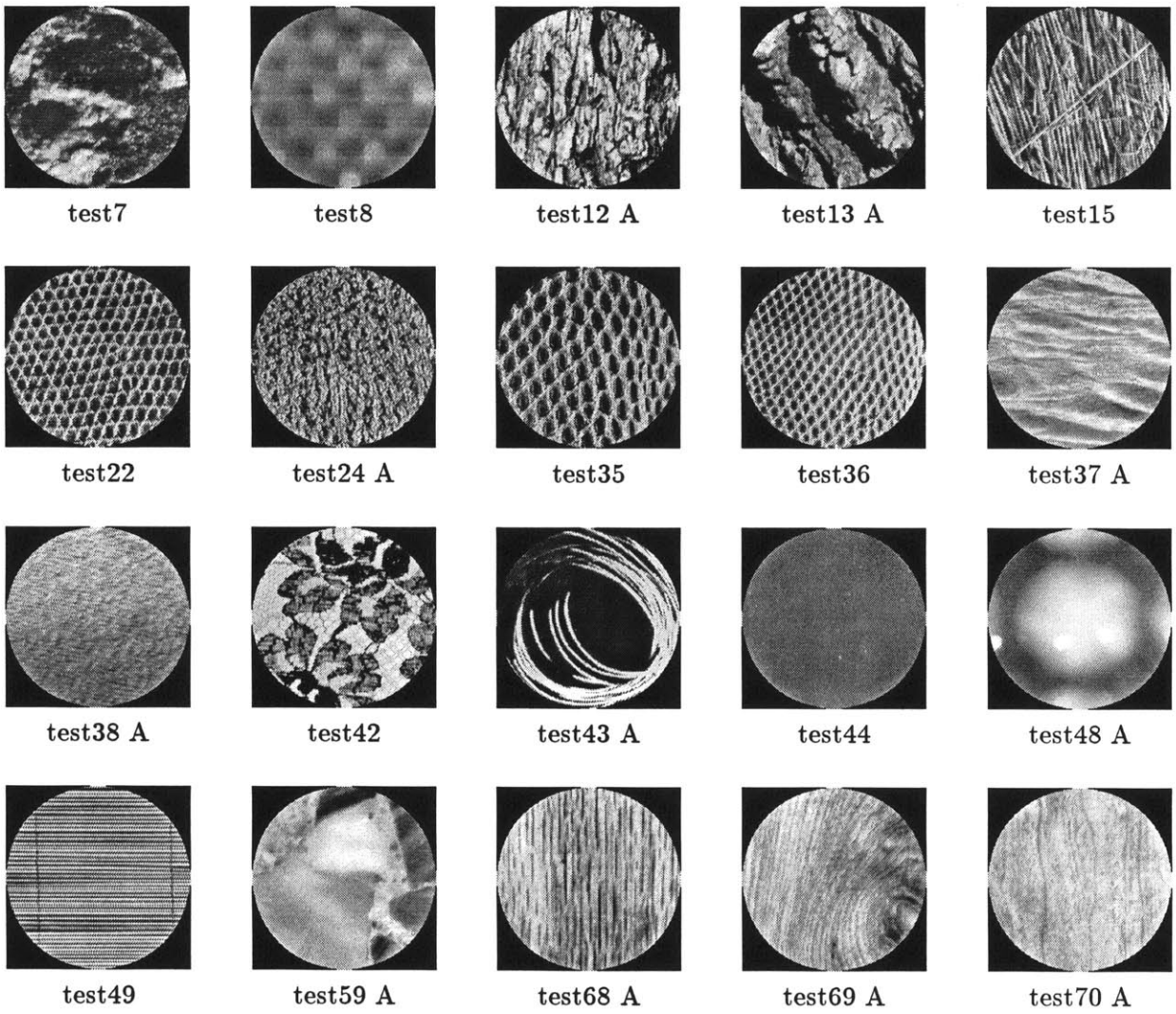
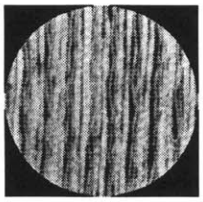
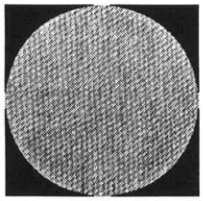
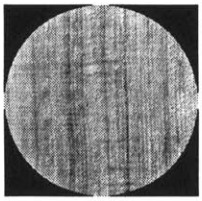


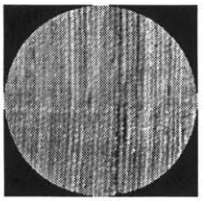
Figure 6-21: One dominant orientation was found by the computer in these. An "A" under an image indicates agreement with the human study data

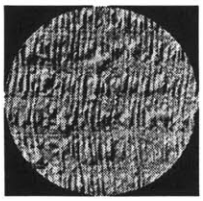

test72 A

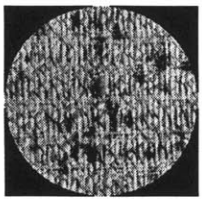

test76

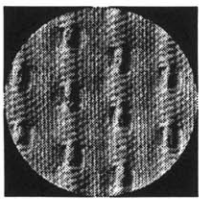

test77

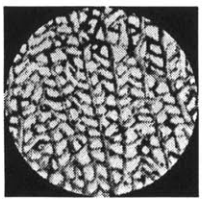

test78 A

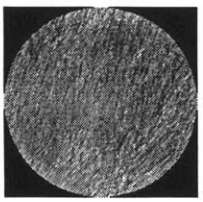

test79


test 80

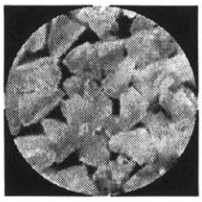

test81

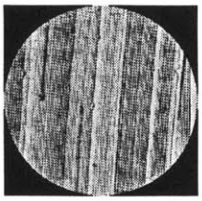

test83 A

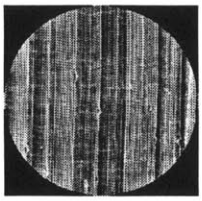

test87 A


test93 A


test97 A


test98


test105A


tes106 A


test107

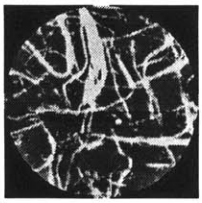

test108

Figure 6-22: Figure 6-21 continued.

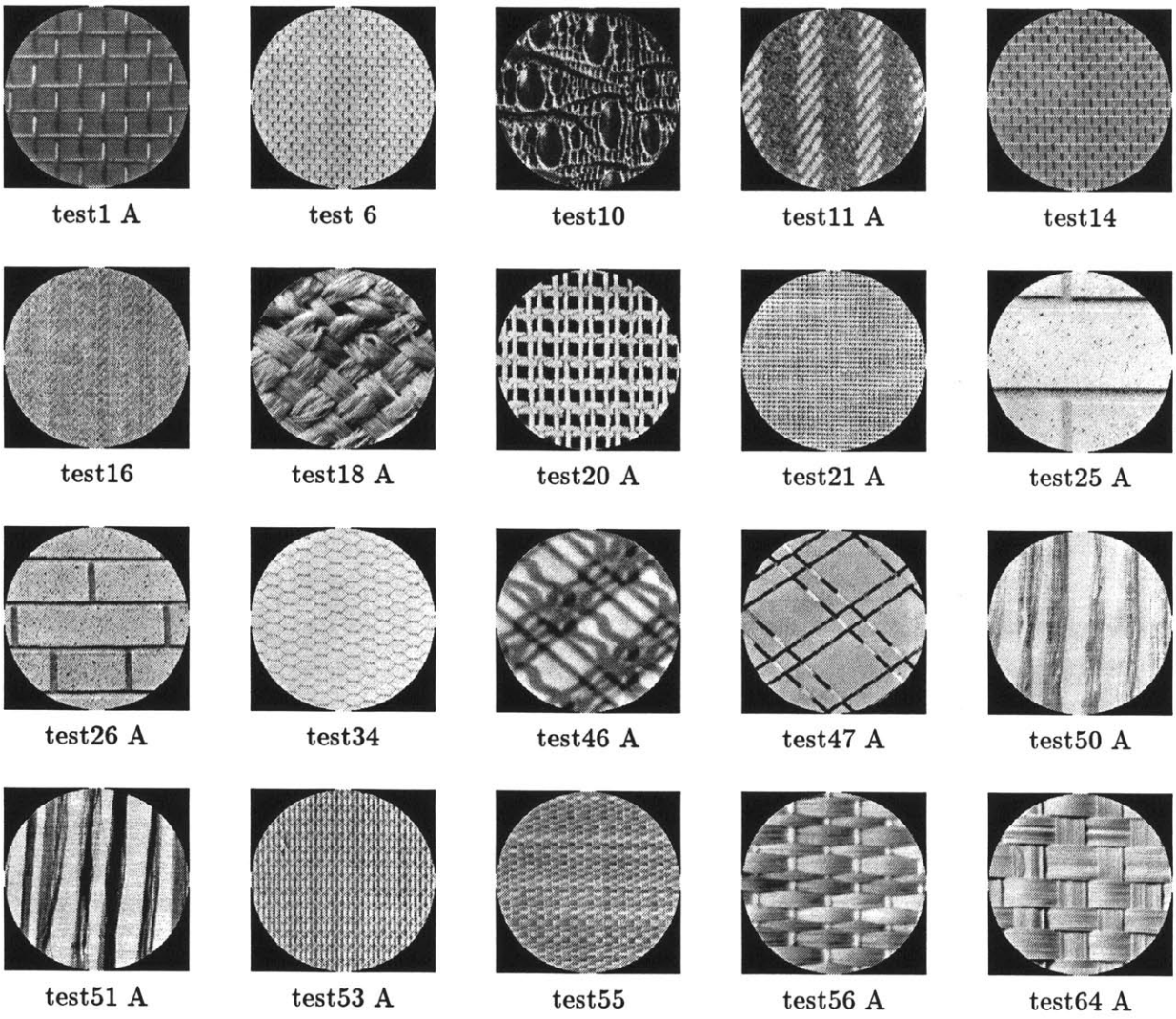


Figure 6-23: Two dominant orientations were found by the computer in these. An "A" under an image indicates agreement with the human study data

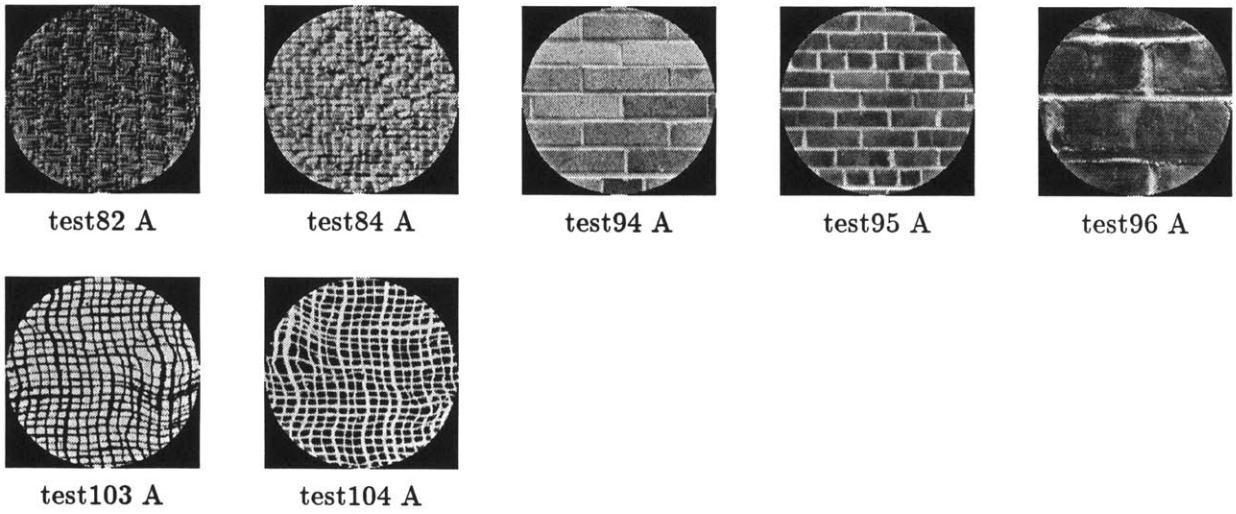


Figure 6-24: Figure 6-23 continued.

	Agree in at least one dominant orientation and its position	Agree in biggest dominant orientation and its position	Agree in total number of dominant orientations	Agree in all dominant orientations and their positions
No Contrast Normalization	95	86	70	68
Level 0 Contrast Normalized	95	86	76	74

Table 6.6: Table summarizing current results

This condition was formulated by examining all the Brodatz textures.

Incorporating the contrast normalization on all images only changed the computer groupings of the test images with one and two dominant orientations shown in Figures 6-21, 6-22, 6-23 and 6-24. The number of agreements in this case vs. the case of a fixed threshold are summarized in Table 6.5. The following textures were changed to "A"'s by changing the contrast: *test*18, *test*25, *test*50, *test*52, *test*53, and *test*96. In terms of the categories for the different numbers of orientations, the following conclusions were made: the net number of agreements for the "two orientation" case increased, number of disagreements for the "one orientation" case decreased and the rest of the cases remained the same. Overall, the use of contrast and the above condition only increased agreement between the computer and human results.

The contrast normalization helped boost textures like *test*51 and *test*52 where there were two orientations, one corresponding to a prominent peak in the histogram and the other corresponding to a very small peak. It is interesting that the contrast normalization boosted the small peaks which are perceptually noticeable to the subjects.

There are several weaknesses with using the present way of incorporating the contrast normalized histogram:

1. If there is only one low-contrast orientation in the texture then it will not be picked up using this method.

2. So far only the relative values of the contrast normalized histogram have been considered and not the shapes of its peaks. It could happen that there are two peaks close in value but one of the peaks does not look very prominent (too broad). In this case the peak is chosen to be prominent even though perceptually it might not be.

3. Using the contrast normalized histogram for higher levels of the steerable pyramid may cause problems since these levels are more effected by high contrast coarse textures. For these cases, the contrast normalization may need to be adapted to the level just as the salience thresholds were found to work better if they were level-dependent.

6.3.3 Cases Where There was no Match between the Human and Computer Data

This section investigates the $111 - 95 = 16$ textures that did not agree in any dominant orientations. These are shown in Figures 6-25 and 6-26. Beside each image in the figures the following information is given:

1. How many orientations were chosen by human?
2. How many orientations were chosen by computer?
3. What were the orientations chosen by humans (the relative strength indicated in brackets)?
4. Did any orientations found by filters match with the human picked orientations? If there is a match then the salience measure γ of the peak corresponding to the orientation is shown in brackets.
5. Speculation about the problem.

Although a set of size 16 is small considering the large variety of data in the original 111, and the fact that the results here are only for the initial round of optimization, there are still some significant problems raised by this set of images. One of these problems is the filter shape and

size. In images such as *test*19 and *test*77, the orientation information is quite fine, and should be detected at level 0 of the pyramid, the finest level. However, in neither case do the filters detect a significant peak at this level. The probable cause here is that the orientation tuning on the filters is not fine enough, i.e., the filter shape should be more narrow, perhaps achievable by using higher Gaussian derivatives.

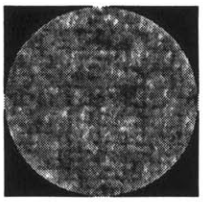
Several of these patterns also contain some evidence for higher-level processes than merely orientation detection. In particular there is evidence for some amount of grouping of objects (*test*66, *test*75), segmentation of objects (*test*41, *test*88, *test*90) and perhaps also of object identification with shadow removal (*test*31). The latter phenomenon may also be caused by some semantic interpretation, although this is a difficult cause to verify.

The salience measure for *test*7 in level 3 of the pyramid is quite high (1.5094) because *test*7 is very coarse and contrasty. Raising the threshold γ_{c3} to this salience measure means that some orientations may no longer be picked up such as the 31° orientation in *test*13 which was given a high strength by humans. The problems with images *test*42, *test*98, *test*107 and *test*108 were discussed above in **Observation4**.

The curves in image *test*71 are visually prominent: humans picked the orientations of the two edges of the curves. However, their contribution to the orientation histogram is small since they do not cover a large area and they are not very high contrast. The problem with image *test*92 can be due to low contrast. Incorporating the contrast normalization for the higher levels of the pyramid may allow the perceived orientations in *test*71 and *test*92 to be picked up.

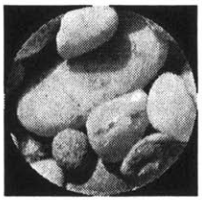
As discussed in Section 4.1.1, multiplying the image with a disc may hide certain structures in the corners of the original image. This problem occurs for *test*44. There are high contrast lines in the lower right hand corner of the square image which were not visible to the subjects.

Many of the problems discussed appear to be fixable by improvements to the contrast handling of the algorithm. These improvements, as well as subsequent optimization of the agreement between the computer and human results are part of our continuing research.



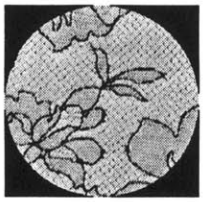
Humans: 2
Computer: 0
Hθ = −5° (0.19), 85° (0.52)
Cθ: level0 (0.0051), level0 (0.0003)
Speculation: filters

test19



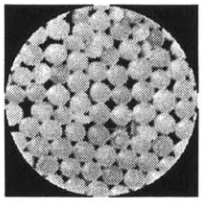
Humans: 1
Computer: 0
Hθ = −65° (0.43)
Cθ: level1 (0.00231)
Speculation: semantic

test31



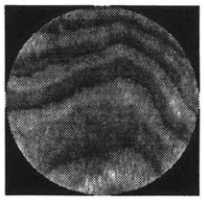
Humans: 1
Computer: 0
Hθ = −55° (0.24)
Cθ: level1 (0.0018)
Speculation: grouping

test41



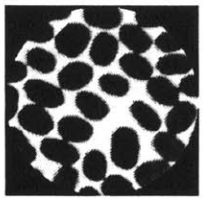
Humans: 2
Computer: 0
Hθ = −55° (0.21), 75° (0.27)
Cθ: level1 (0.0123), level3 (0.0017)
Speculation: grouping

test66



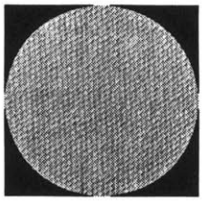
Humans: 2
Computer: 0
Hθ = −75° (0.42), 35° (0.21)
Cθ: level2 (0.034), (0.0143)
Speculation: contrast

test71



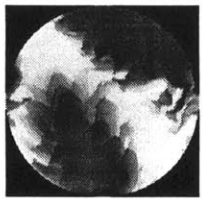
Humans: 1
Computer: 0
Hθ = −45° (0.22)
Cθ: level2 (0.0031)
Speculation: grouping

test75



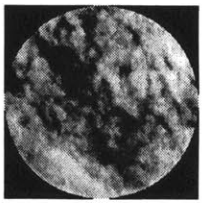
Humans: 2
Computer: 1
Hθ = −65° (0.51), −5° (0.445)
Cθ: level2 (0.0001), level3 (0.0452)
Speculation: filters

test77



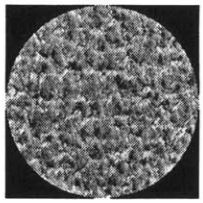
Humans: 1
Computer: 0
Hθ = 35° (0.29)
Cθ: level2 (0.0011)
Speculation: grouping

test88



Humans: 1
Computer: 0
Hθ = 25° (0.28)
Cθ: level2 (0.0460)
Speculation: filters

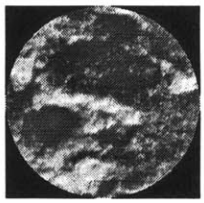
test90



Humans: 1
Computer: 0
Hθ = 85° (0.27)
Cθ: level2 (0.0129)
Speculation: contrast

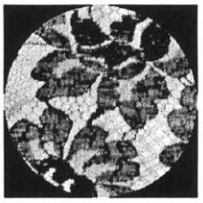
test92

Figure 6-25: "Difficult" images where the computer and human did not agree on any of the dominant orientations. Here the humans picked more orientations than the computer.



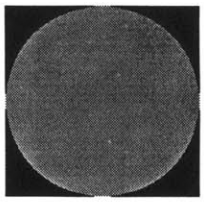
test7

Humans: 0
Computer: 1
Cθ = 77°, level=3 (1.5094)
Hθ: (0.10)
Speculation: contrast



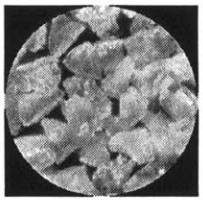
test42

Humans: 0
Computer: 1
Cθ = 83°, level=0 (0.0108)
Hθ: no match
Speculation: contrast



test44

Humans: 0
Computer: 1
Cθ = −45°, level=0 (9.6)
Hθ: no match
Speculation: Missing information in corner

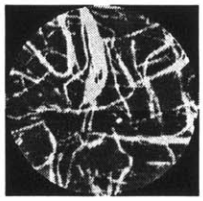


test98

Humans: 0
Computer: 1
Cθ = −53°, level=0 (0.0107)
Hθ: (0.02)
Speculation: contrast

test107

Humans: 0
Computer: 1
Cθ = −83°, level=2 (0.2003)
Hθ: (0.09)
Speculation: contrast



test108

Humans: 0
Computer: 1
Cθ = 11°, level=0 (0.0156)
Hθ: (0.09)
Speculation: contrast

Figure 6-26: More "difficult" images where the computer and human did not agree on any of the dominant orientations. Here the computer picked more orientations than the humans.

Chapter 7

Conclusion And Future Goals

The results in Chapter 6 are very encouraging. Using different thresholds for the four levels of the steerable pyramid is sufficient to detect all the perceptually dominant orientations of 68 out of 111 textured images without even using contrast normalization. These orientations were the same chosen by 40 human subjects. Using contrast normalization brings this number up to 74. The algorithm is able to detect at least one perceptually dominant orientation chosen by the 40 human subjects in 95 out of the 111 textured images. These numbers are impressive considering that

1. There were 111 different natural textures (vs. a typical small 10-20 size set, and vs. common use of synthetic images).

2. Some of the textures were not homogeneous.

3. The semantic meaning of the textures was not removed.

Nevertheless, there is still a lot of room for improvement. More research is warranted in the following areas.

7.1 Picking Better Thresholds

There are some cases where a small adjustment in the thresholds will enable the computer to pick out a relevant peak. As the difficult cases are studied more and their salience measures are known, the thresholds might be able to be changed slightly enabling these cases to be accepted without causing further disagreements. Also there were some disagreements for which the human data threshold t_h was relatively small meaning that the subjects did not find those orientations to be very strong. Perhaps a lower t_h will reduce the number of disagreements without causing very strongly perceivable orientations to be ignored by the computer.

7.2 Filters of Different Shapes

There were cases like *test*22 and *test*36 where a narrower filter could have picked out the orientations chosen by humans. Malik and Perona in [10] use bar-like filters to pick out orientations. It would be interesting to see whether filters like this can be more effective in picking out finer orientations without losing the ones the steerable filters have found. Higher orders of the derivatives of Gaussians with their finer angular tuning may also be explored for picking out orientations of closely spaced structures like in *test*52 for which the current algorithm picked 5 orientations and humans picked four. Unfortunately, the more filters are used for orientation picking, the harder it is to incorporate all their outputs for decision making.

7.3 Contrast Normalization

The results of using the contrast normalization in chapter 6 are very interesting. It appears that using the contrast normalization enables the computer to pick out orientations due to low

contrast structures like *test51* whose contribution in the orientation histogram is very small. These orientations are perceived by humans although they are given smaller strengths than more contrasty or bigger structures. Presently it is not known when the contrast normalization can cause false decisions to be made since the requirements for using the contrast normalization technique were very strict in chapter 6. More research needs to be directed in this area to find the limitations of this technique and also determine how it may effect the steerability of the filters.

7.4 Testing the Algorithm on Different Test Images

It is important to know whether the choice of the thresholds is restricted to the test images used in this thesis or will it be "universal" in its effectiveness for other images. This algorithm should then be used on a different database and the effectiveness of the thresholds be confirmed.

Appendix A

Teaser Images

Figure A-1(a) shows horizontal lines with a vertical break between them. It would be interesting to see whether most subjects perceived the break's vertical orientation and also to see whether the filters would detect this orientation. Figure A-1(b) is the same as Figure A-1(a) except that the distance between the breaks is closer. The goal here is to see whether the width of the break changes the strength of the perception of the vertically oriented line, and whether the current sized filters would pick this break at such a small width. Figures A-1(c) and A-1(d) are similar to Figures A-1(a) and A-1(b) except that the lines are no longer horizontal and the break is no longer vertical. In Figure A-1(e) there can be a perception of horizontal orientation even though there is no horizontal lines. Again it is interesting to see whether people and filters pick up this horizontal perception. Figure A-1(f) illustrates orientations at two different scales. Figure A-2 is the famous Zölliner illusion. The filters detect those lines to be parallel. The illusion might not work since the red bar aligns on top one of the lines which reduces the effect. Also since the perceived orientations are slightly different, with the inaccuracies which has to

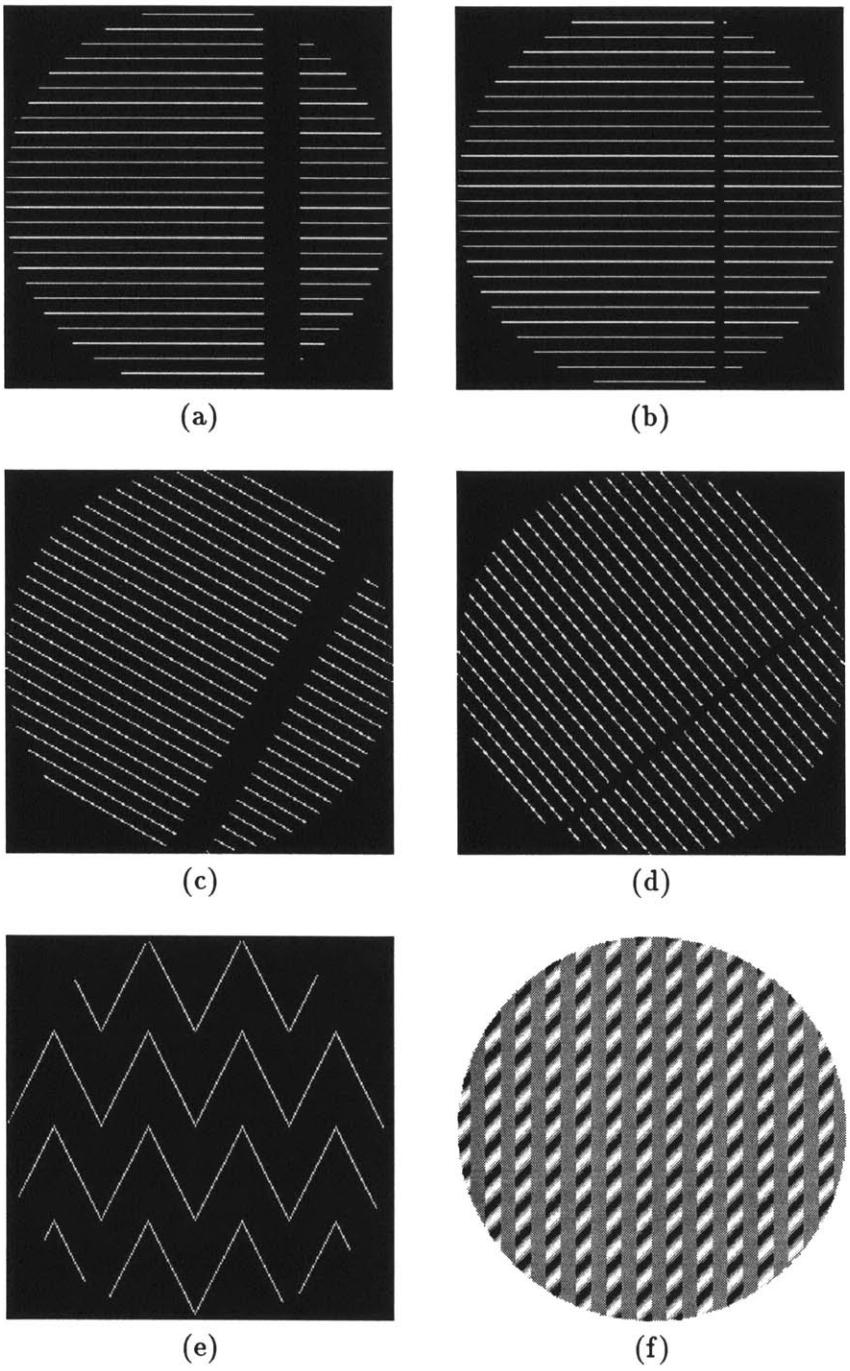


Figure A-1: Teaser test images shown to subjects

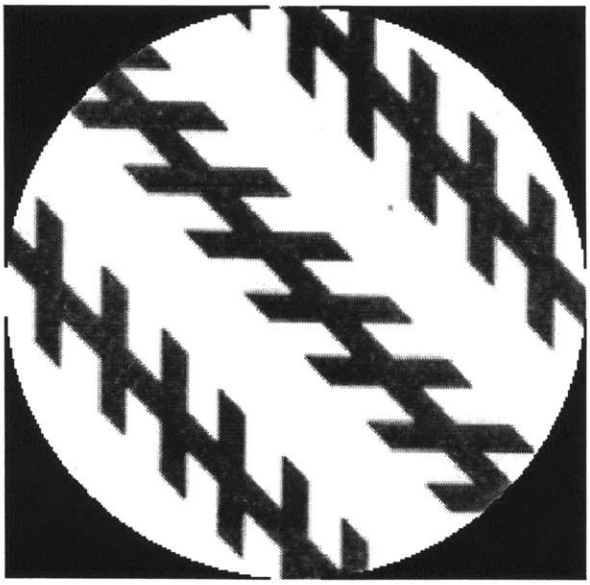


Figure A-2: Teaser test image known as the **Zölliner illusion** shown to subjects.

---

be allowed to the aligning of red bar, they might not be detected. In future work, the results of the orientation algorithm on these teaser images will be compared with the human results.

Appendix B

Histogram Peak Finding: Case 2

As described in Chapter 4, the maximum and minimum points in d_H corresponding to the inflection points of H_s are used for the calculation of V_{dist} and H_{dist} in (4.5). It would be very simple if for each pair of maximum and minimum points there was a zero crossing signifying the presence of a peak. However, this is not the case all the time. When the slope in the H_s changes concave to convex or vice versa, there will be a maximum and a minimum point in d_H. However, these points may not have a zero crossing between them. How should these points be incorporated?

Initially, these points were ignored and only maximum and minimum pairs in d_H where the maximum was bigger than zero and the minimum below zero were used. For very sharp and narrow peaks, like P_1 in Figure 4-4, using the maximum and minimum on either side of a zero-crossing gave a high salience measure. A high salience measure is desirable for sharp and narrow peaks.

However, for broad peaks like P_1 in Figure B-1 corresponding to $D109$ (look at $test109$ in Figure 6-14) , the distance between the maximum and minimum on either side of the zero-

crossing does not indicate the broadness of the peak (70% of human subjects tested on $D109$ considered it to be non-directional). In Figure B-1, using the distance W_{d2} for H_{dist} gives a smaller salience measure for P_1 than W_{d1}. W_{d1} is the H_{dist} if the maximum and minimum pair ($max2$, $min2$) were chosen and W_{d2} is the H_{dist} if the maximum and minimum pair ($max1$, $min3$) were chosen.

Which maximum and minimum pair in d_H should then be used to calculate the V_{dist} and H_{dist} for the salience measure? The criteria for choosing the maximum and minimum pair at locations (max_D, min_D) in d_H were found by looking at the different types of peaks in the orientation histograms. The pair best characterizing the shape of the peaks were chosen. The results of the human visual test were used to verify the criteria developed for picking maximum and minimum at locations max_D and min_D.

The following steps were followed in the algorithm to find the desired maximum and minimum points for the calculation of the salience measure of the peaks in H_s. Note that the steps are described in pseudo-code.

1. Find all the maximum and minimum points in d_H.

2. Find the most negative minimum and denote its location min_Z. Location min_Z is in degrees.

3. Find the first positive maximum to the right of location min_Z. The location of this maximum will be denoted max_D.

4. Check the minimum to the right of location max_D and denote its location min_D. If $d_H(min_D) < 0$ then a zero-crossing is found. Label this zero-crossing as a peak. Go to step 6. Else, go to step 5.

5. Check the next maximum and minimum pair in d_H. Denote their locations (max_N, min_N). Consider the following case.

   Case A If $d_H(max_D) < 0.4\ d_H(max_N)$ then choose maximum at max_N instead of current max_D : $max_D = max_N$.

   If $d_H(min_N) < 0$ then $min_D = min_N$. Go to step 6. Else repeat step 5.

6. Check the next maximum and minimum pair. Denote their locations (max_N, min_N). If $d_H(max_N) > 0$ then go to step 7. Else consider the following case:

   Case B If $|d_H(min_N) - d_H(min_D)| \leq 0.10\ d_H(min_D)$ then $min_D = min_N$.

   If Case B is not true then consider Case C:

   Case C If $|d_H(min_N)| > 2.5\ |d_H(min_D)|$ then $min_D = min_N$. If Case B and Case C not true then consider Case D:

   Case D If $d_H(min_N) > 0.4\ d_H(min_D)$ and $d_H(min_N) < 0.2\ d_H(min_Z)$. The term $0.2\ d_H(min_Z)$ insures that the minimum is not too close to zero.

   If a minimum satisfies the above cases then it will be considered instead of the minimum at the current min_D.

   Repeat step 6.

7. $H_{dist} = max_D - min_D$

   $V_{dist} = d_H(max_D) - d_H(min_D)$

   $max_D = max_N$

   Goto step 4.

For an illustration, one iteration of steps 1-7 was done for the orientation histogram in Figure B-1. The positions min_Z, max_D, min_D, max_N and min_N corresponding to the maximums and minimums in d_H checked in the first iteration are shown in Figure B-1.

Example for Case A: In Figure B-2, the maximum and minimum points in d_H are checked following the steps mentioned above. The maximum $max1$ is reached and its position assigned to max_D. Step 5 is followed with the position of $max2$ corresponding to max_N. Since $max2$ satisfies Case A, it will be chosen over $max1$ resulting in a smaller H_{dist} and a bigger V_{dist} for P_3. Therefore, the salience measure for P_1 will be larger using $max1$ instead of $max2$. A higher salience measure is verified by the fact the subjects gave a high strength to the orientation corresponding to P_1 (look at $test13$ in Figure 6-15).

Example for Case B: In Figure B-3, the maximum and minimum points in d_H are checked following the steps mentioned above. The minimum $min1$ is reached and it position assigned to min_D. Following step 6, since the maximum to the right of $min1$ is negative, $min2$ is considered for the three cases. The value of $min2$ satisfies Case B and $min2$ is chosen over $min1$. Choosing $min2$ over $min1$ results in a bigger H_{dist} and a smaller V_{dist} for P_1. Therefore, the salience measure will be smaller for P_1. A smaller salience measure for P_1 is verified by the fact that 80% of the subjects considered $D73$ to be non-directional (look at $test73$ in Figure 6-14).

Example for Case C: In Figure B-4, the maximum and minimum points in d_H are checked following the steps mentioned above. The minimum $min1$ is reached and its position assigned to min_D. Following step 6, since the maximum to the right of $min1$ is still negative, $min2$ will be considered for the three cases. The value of $min2$ satisfies Case C and $min2$ is chosen over

$min1$. Choosing $min2$ over $min1$ gives a better characterization of the sharpness and the width of the P_2.

Example for Case D: In Figure B-5, the maximum and minimum points in d_H were checked following the steps mentioned above. The minimum $min1$ is reached and its position assigned to min_D. Following step 6, since the maximum before $min2$ is still negative, $min2$ is considered for the three cases. The value of $min2$ satisfies Case D and $min2$ is chosen over $min1$. Choosing $min2$ over $min1$ results in a bigger H_{dist} and a smaller V_{dist}. Therefore, the salience measure for P_1 will be smaller. A smaller salience measure is verified by the fact that 55% of subjects consider $D100$ to be non-directional (look at $test100$ in Figure 6-14).

For broad peaks, it would have been better if the value of min_N for Case D did not have to be smaller than 0.2 $d_H(min_Z)$. For example in Figure B-1, if the value of $min4$ did not have to be smaller than 0.2 times the value of $min3$ (position of $min3$ equals min_Z) then it would be chosen over $min3$. Choosing $min4$ would result in a bigger H_{dist} and P_1 would then have a smaller salience measure . However, for peaks whose salience value should be big (subjects chose the orientations of these peaks with high strength), choosing a minimum close to zero, results in a high H_{dist} value. The salience measure will then be smaller for these peaks. It was found that an effective way to ignore broad peaks without reducing the salience measure of "desired" peaks was to incorporate a weighting factor w_{br} discussed in Section 6.3.

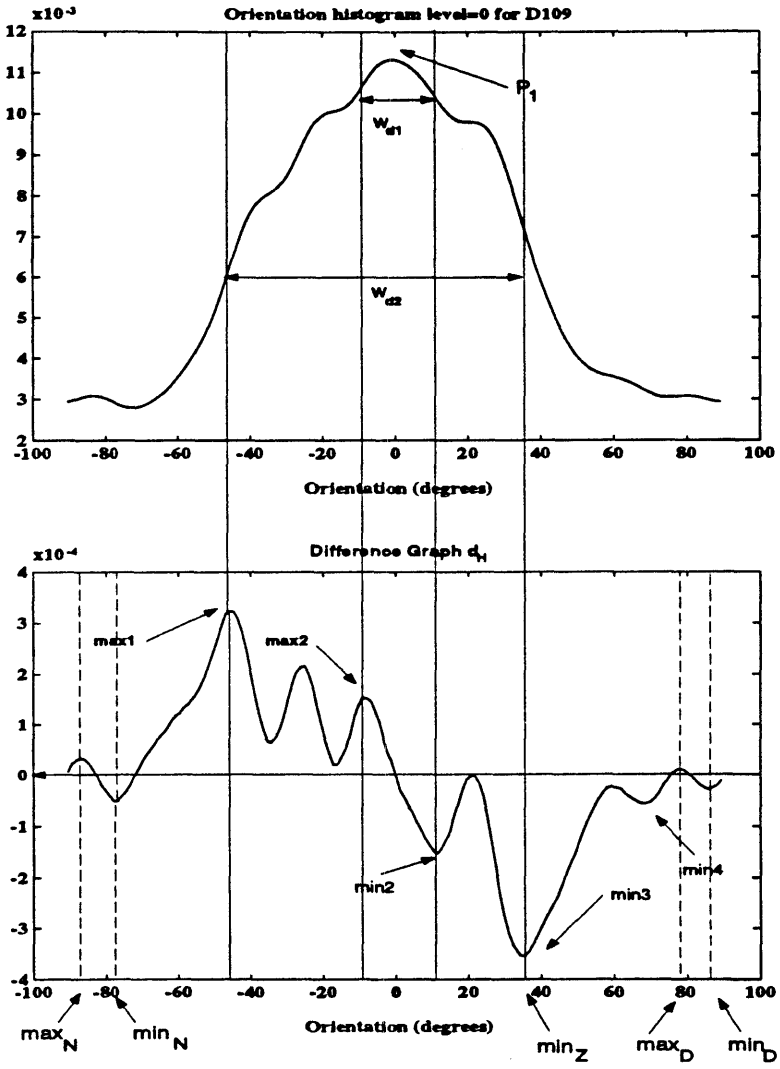


Figure B-1: Orientation histogram and the corresponding graph of d_H for $D109$

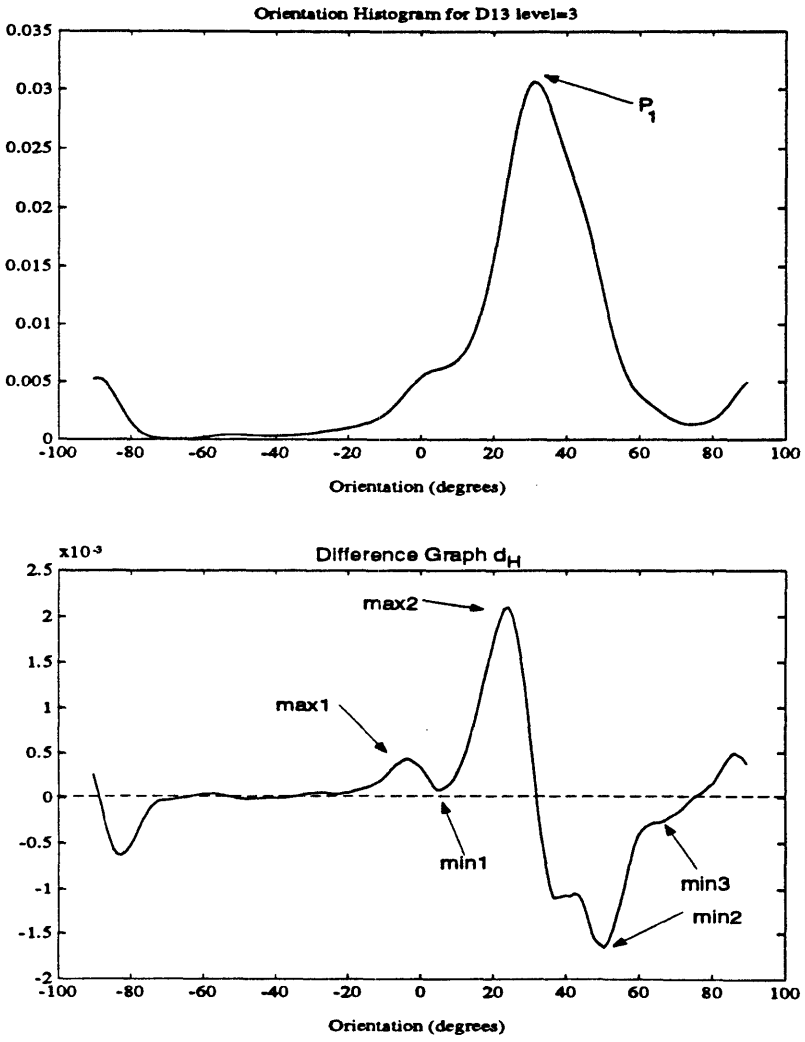


Figure B-2: Orientation histogram and the corresponding graph of d_H for $D13$

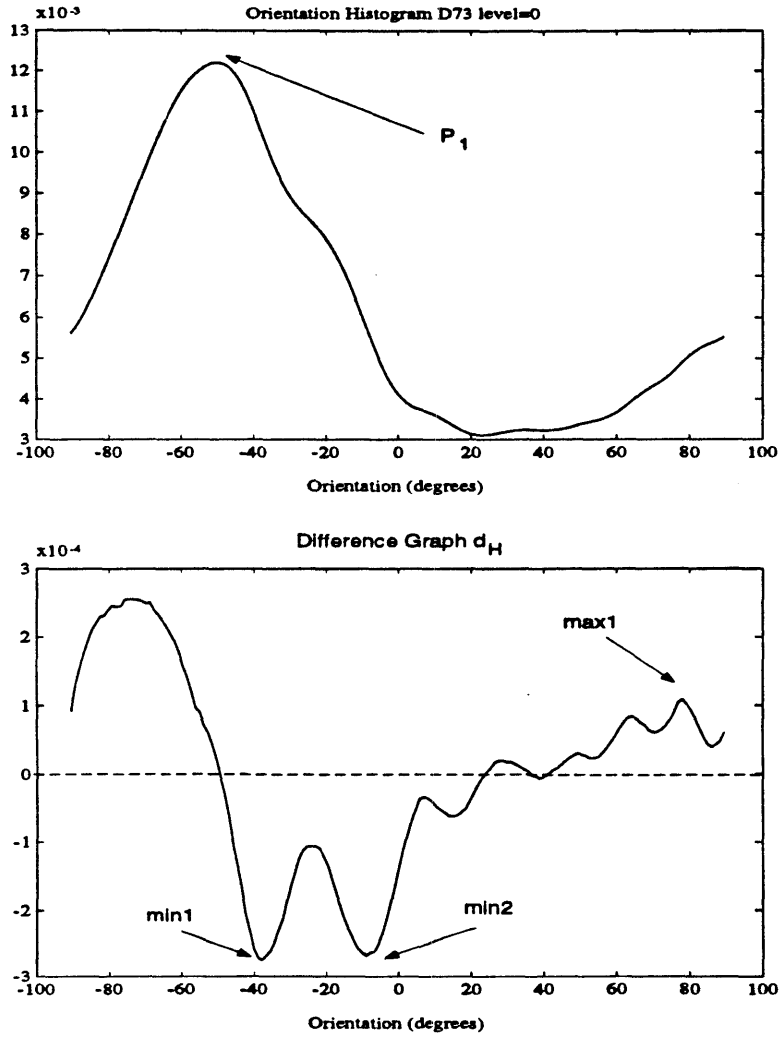


Figure B-3: Orientation histogram and the corresponding graph of d_H for $D73$

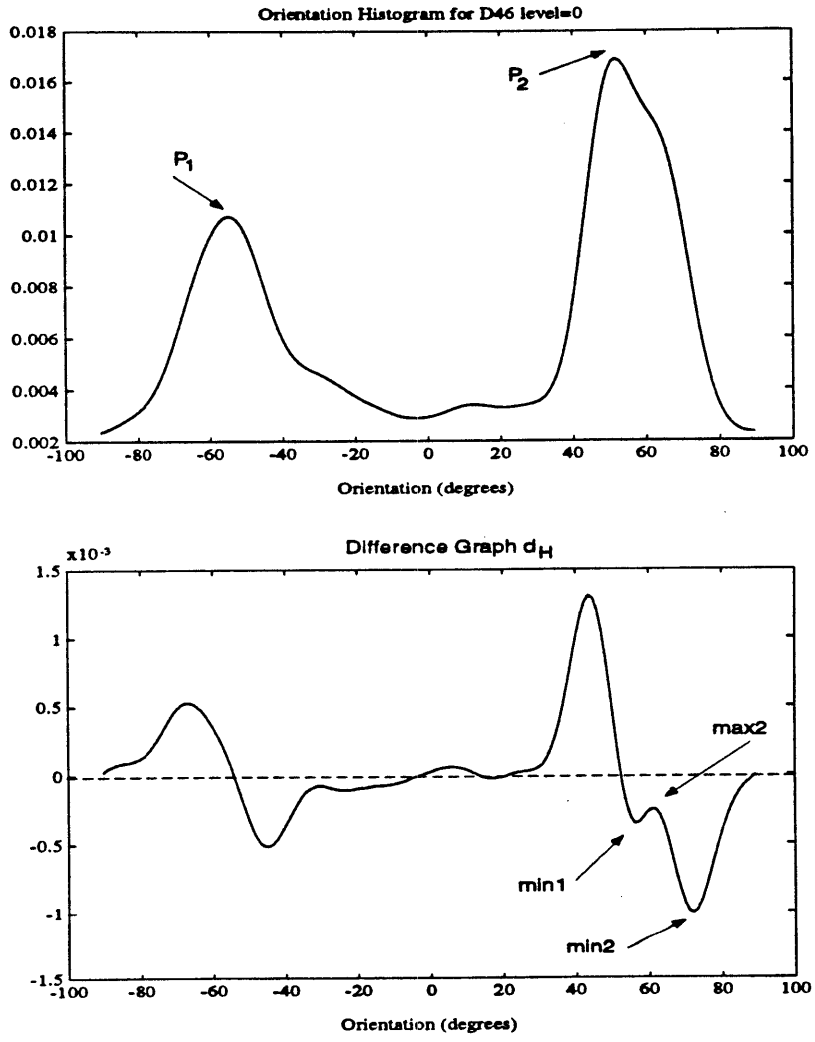


Figure B-4: Orientation histogram and the corresponding graph of d_H for $D46$

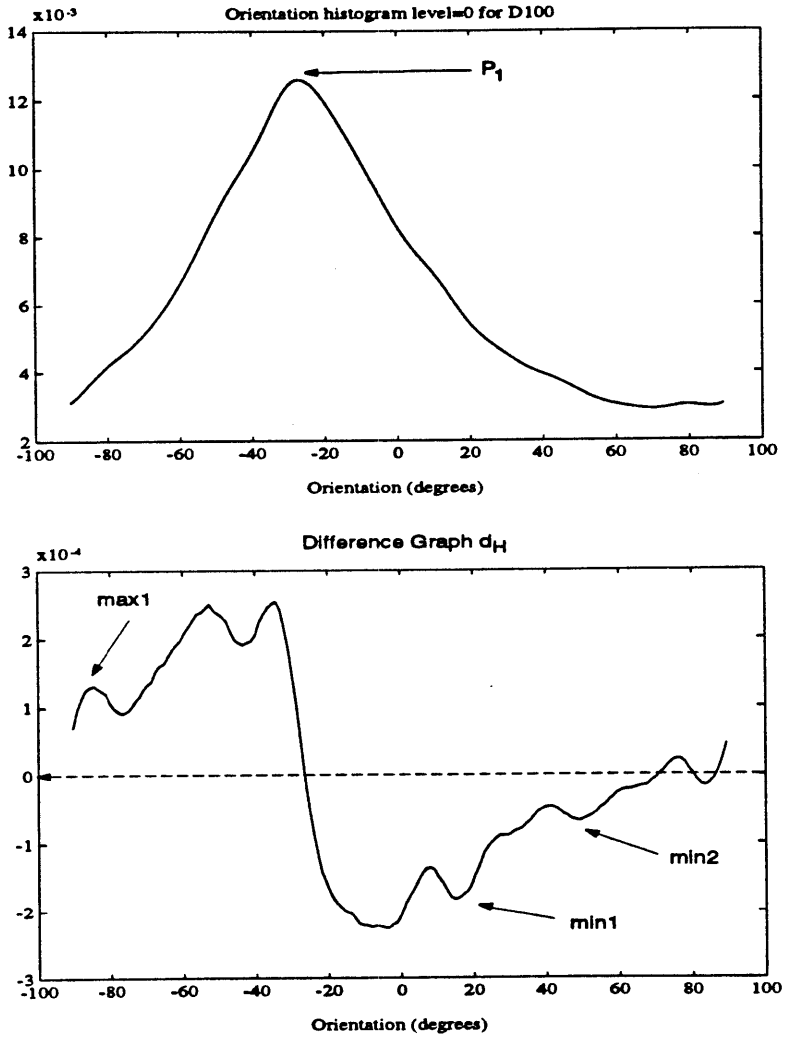


Figure B-5: Orientation histogram and the corresponding graph of d_H for $D100$

Bibliography

[1] A. R. Rao and J. Lohse, "Identifying high level features of texture perception," Tech. Rep. RC17629, IBM Laboratory, February 1992.

[2] H. Tamura, S. Mori, and T. Yamawaki, "Visual discrimination of stochastic texture fields," *IEEE Trans. Sys., Man and Cyber.*, vol. SMC-8, no. 11, pp. 796–804, 1978.

[3] G. C. Phillips and H. R. Wilson, "Orientation bandwidths of spatial mechanisms measured by masking," *Journal of Optical Society of America A*, vol. 1, pp. 226–232, 1984.

[4] H. D. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiology*, vol. 195, pp. 215–243, 1968.

[5] M. A. Webster and R. L. D. Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *Journal of Optical Society of America A*, vol. 2, no. 7, pp. 1124–1132, 1985.

[6] R. N. Shepard and L. A. Cooper, *Mental Images and their Transformations.* Cambridge, MA: The M.I.T. Press, 1982.

[7] J. Aloimonos and D. Shulman, *Integration of visual modules : an extension of the Marr paradigm.* New York: Academic Press, 1989.

[8] Y. Choe and R. L. Kashyap, "3-d shape from a shaded and textural surface image," vol. PAMI-13, no. 9, pp. 907–919, 1991.

[9] J. R. Bergen and M. S. Landy, "Computational modeling of visual texture segregation," in *Computational Models of Visual Processing* (M. S. Landy and J. A. Movshon, eds.), (Cambridge, MA), pp. 253–271, MIT Press, 1991.

[10] J. Malik and P. Perona, "Preattentive texture discrimination with early visual mechanisms," *Journal of Optical Society of America*, vol. 7, pp. 923–932, 1990.

[11] P. Brodatz, *Textures: A Photographic Album for Artists and Designers.* New York: Dover, 1966.

[12] R. Bajcsy, "Computer description of textured surfaces," *Int. Joint. Conf. Artificial Intelligence*, pp. 572–578, 1973.

[13] M. Kass and A. Witkin, "Analyzing oriented patterns," in *Readings in Computer Vision* (M. A. Fischler and M. Kaufman, eds.), pp. 268–276, M. Kaufman, 1987.

[14] S. Chaudhuri, H. Nguyen, R. M. Rangayyan, S. Walsh, and C. B. Frank, "A Fourier domain directional filtering method for analysis of collagen alignment in ligaments," *IEEE Transactions on Biomedical Engineering*, vol. 34, no. 7, pp. 509–517, 1987.

[15] R. A. Young, "The Gaussian derivative model for machine vision: Visual cortex simulation," Tech. Rep. GMR-5323, General Motors Research Laboratories, 1986.

[16] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[17] P. Perona, "Deformal kernels for early vision," Tech. Rep. MIT-LIDS-P-2039, MIT Laboratory for Information and Decision Sciences, MIT, Cambridge MA, October 1991.

[18] R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation and Noise.* Boston, MA: Houghton Mifflin Company, 1990.

[19] H. Knutsson and G. H. Granland, "Texture analysis using two-dimensional quadrature filters," in *IEEE Workshop on Comp. Arch. Patt. Anal. Im. Dat. Base Man.*, (Pasadena, CA), pp. 206–213, 1983.

[20] J. Bigün and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proc. 1st Int. Conf. Comp. Vis.*, (London, England), pp. 433–438, 1987.

[21] R. Rao and B. G. Schunck, "Computing oriented texture fields," *CVGIP Graphical Models and Image Processing*, vol. 53, no. 2, pp. 157–185, 1991.

[22] B. Jähne, *Digital Image Processing.* New York: Springer-Verlag, 1991.

[23] M. T. Andersson, *Controllable Multidimensional Filters and Models in Low Level Computer Vision.* PhD thesis, Linköping University, 1992. Dissertation No. 282 ISBN 91-7870-981-4.

[24] A. Rosenfeld and A. C. Kak, *Digital Picture Processing - Vols. 1 and 2.* New York: Academic, 2 ed., 1982.

[25] J. Bigün and J. M. H. du Buf, "Geometric image primitives by complex moments in Gabor space and application to texture segmentation," in *IEEE Computer Society on Computer Vision and Pattern Recognition*, June 1992.

[26] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition Journal*, vol. 24, pp. 1167–1186, 1991.

[27] T. Chang and C. C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," Tech. Rep. USC-SIPI198, University of Southern California, Los Angeles, CA, February 1992.

[28] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[29] E. H. Adelson, E. Simoncelli, and W. T. Freeman, "Pyramids and multiscale representations," in *European Conference on Visual Perception*, (Paris), 1990.

[30] J. Bigün, "Frequency and orientation sensitive texture measures using linear symmetry," *Signal Processing*, vol. 29, pp. 1–16, 1992.

[31] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[32] G. A. Wright and M. E. Jernigan, "Texture discriminants from spatial frequency channels," in *IEEE International Conference SMC*, pp. 519–524, 1986.

[33] R. Shapley, "Retinal regulations of visual contrast," *Optics and Photonics*, vol. 2, no. 8, pp. 17–23, 1990.

[34] R. C. Gonzalez and P. Wintz, *Digital Image Processing.* Addison-Wesley Publishing Company, 2 ed., 1987.

[35] T. G. Stockham, "Image processing in the context of a visual model," *Proceedings of the IEEE*, vol. 60, no. 7, pp. 828–842, 1972.

[36] D. J. Heeger, "Nonlinear model of neural responses in cat visual cortex," in *Computational Models of Visual Processing* (M. S. Landy and J. A. Movshon, eds.), (Cambridge, MA), MIT Press, 1991.

[37] W. T. Freeman, *Steerable Filters and Local Analysis of Image Structure*. PhD thesis, Media Arts and Sciences, MIT, 1992.