Spatial Vision, Vol. 8, No. 2, pp. 221-253 (1994)

Finding perceptually dominant orientations in natural textures

ROSALIND W. PICARD and MONIKA GORKANI
Perceptual Computing Section, MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA

Received for publication 22 September 1993

Abstract—An algorithm for detecting orientation in texture is developed and compared with results of humans detecting orientation in the same textures. The algorithm is based on the steerable filters of Freeman and Adelson (*IEEE Trans. PAMI* **13**, 891–906, 1991), orientation-selective filters derived from derivatives of Gaussians. The filters are applied over multiple scales and their outputs non-linearly contrast-normalized. The data for humans were 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 grey-level images of natural textures taken from the Brodatz album, a standard collection used in computer vision and image processing. Results show that the computer and humans 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 humans and computer. Disagreements are analyzed and possible causes are discussed. Some apparent limitations in the current filter shapes and sizes are illustrated, as well as some (surprisingly small) effects believed to be caused by semantic recognition and gestalt grouping.

1. INTRODUCTION

Orientation is one of the most perceptually significant components in texture recognition (Tamura *et al.*, 1978; Rao and Lohse, 1992) and in visual attention (Treisman and Gelade, 1980; Wolfe *et al.*, 1989). Psychophysical evidence exists that humans use orientation as a cue for discriminating textures (Tamura *et al.*, 1978; Phillips and Wilson, 1984; Julesz, 1991; Rao and Lohse, 1992), and physiological experiments suggest the existence of orientation selective mechanisms in the human visual system (Hubel and Wiesel, 1968; Webster and De Valois, 1985). Local orientation information has also been argued to play a critical role in curve detection (Zucker, 1985).

Extraction of orientation over a large scale can be used for rotating images to align them before beginning closer comparison, a process possibly done by humans during pattern recognition (Shepard and Cooper, 1982). An observer can also obtain shape and perspective information from texture (Aloimonos and Shulman, 1989; Choe and Kashyap, 1991) and consequently orientation may play a key part in this process. Orientation at a finer level is also a fundamental component of texture and pattern—wood grain, sand ripples, parquet floors, bookshelves, and parking lots; all these contain dominant orientation information which can be related to their formation or function.

Not only are orientation detection cells at work in the low levels of the visual cortex and in a possibly higher level process of pattern rotation for alignment, but orientation can also be considered a semantic feature of image data. 'Semantic' here refers to the human use of orientation in natural language when describing visual information. Semantic features are important in rapidly growing new applications such as searching for images by their visual context (Picard and Kabir, 1993).

1.1. Applications of orientation

Imagine in a few years when every home has 'video on demand' and all art collections, movies, patent libraries, photo albums, books, journals, and other visual collections are accessible online. There will be exabytes of information, making it impossible to find a particular image or video clip without spending inordinate amounts of time looking for it. One can currently search through text for keywords, but not through video for keyframes. In the next several years it will become tremendously important to have automated tools that can search for visual information whether or not it has a description attached. A likely scenario is one where a user shows the computer a pre-existing pattern, and requests all other patterns 'like' this one.

In such a scenario, the search for a similar image should agree with the human's notion of similarity. To succeed, the computer must know which features attract the human's attention, and how to combine these features for locating perceptually similar patterns. The goal is to get the computer to identify textures the person would identify if he or she had time to look through them all. As orientation is one of the most significant features for human attention and texture matching, it is important that algorithms which recognize orientation be developed.

The results of this study on orientation apply directly to the image search problem. Furthermore, in the future when search tools become more semantic and there are pre-stored object descriptions to be matched, then one will also be able to associate to keywords such as 'brick wall' the feature, 'usually has two dominant orientations'. The orientation detection algorithm can then do a much faster search for the corresponding visual data than it could do without this information. Even if a portion of an image has no dominant orientations, that is still important information to use in speeding up comparisons. For example, two primary categories of textures have long been recognized: structural and statistical (Haralick, 1979). Orientation is a salient feature that can be used to decide which of these (or other) categories a texture is close to, i.e. the statistical is likely to not have more than one dominant orientation. Given such information, one may select a model more suitable to recognizing that class of pattern. Detailed model features can then be extracted relative to the dominant orientations to provide invariant measures within inhomogeneous data.

1.2. Orientation over scale

How *much* pattern recognition can be achieved using only orientation information is an open question. A key difficulty is that it appears to be important to gather the information over multiple scales (Bergen and Adelson, 1988). Also, orientation information is complicated by its interactions with effects such as contrast (Heeger, 1991), similarity grouping and gestalt effects (Hamey, 1992), and prior knowledge present when a pattern is recognized (Richards, personal communication). In cases such as similarity grouping, it may be useful to detect orientation after first running the data through some (possbily nonlinear) transformation. Two stages of linear direction filtering, separated by a nonlinearity such as rectification, appear to help in a number of cases (Graham *et al.*, 1992, 1993).

This research focuses on finding orientations from image intensity values using one stage of linear filtering, applied over four different scales, and followed by nonlinear contrast normalization and decision-making. The output is the number of dominant orientations, their angles, and their strengths.

1.3. Brodatz image test data

The textures used throughout this research come from the Brodatz Album (Brodatz, 1966). These textures are the *de facto* standard used by researchers in computer vision and pattern recognition. For both the human and computer analysis done in this research, data are taken from a 256 × 256 section cropped from the center of a 512 × 512 8-bit grey-level image in the digitized Brodatz Album. This is repeated for 111 different images, yielding test images named $D1$, $D2, \ldots, D112$.[1] These square images were used as input to the computer orientation-finding algorithm. The images used in the human test and shown in the figures of this paper were these images multiplied by a disk and named Test1, . . ., Test112 (details in Section 4.2).

Use of the 111 Brodatz textures makes this study considerably larger in texture variety than any other study known to the authors. Nonetheless, this data set still has limitations. Almost all the images are 'frontal plane', i.e. they are not subject to any perspective distortions. Several images are inhomogeneous or have complex patterns for which it is difficult or ambiguous to identify dominant orientations. There is a majority of horizontal and vertical orientation (as there is also in the environments where most people spend time with their eyes open). Each test image was treated as one region, making the task more difficult but the results more realistic for each extension to real scenes.

1.4. Overview of paper

In this paper an algorithm is developed for detecting orientation in texture, and a study is done to determine the orientations found by humans in the same textures. The algorithm is described in Section 3 and the human study in Section 4. A careful comparison of the two is given in Section 5.

It is important to clarify that, while we are curious how the human visual system achieves recognition, the primary goal here is to develop an algorithm that imitates the human system's output (dominant orientations) for a given visual input (texture). Since what happens between the input and output is still largely unknown for the human, and since there may be more than one way to get the outputs from the inputs, this paper makes no argument that the proposed algorithm is a model of human computation. Nevertheless, it is the authors' aim to tune the model so its performance is as closely matched as possible to the human data.

No existing algorithm for pattern recognition has claimed to recognize perceptually similar images in the general case. However, orientation is likely to e a critical part of such an algorithm in the future. Hence, this study focuses on: 1) obtaining one set of measures from humans for how they characterize orienta- ion; and (2) developing an algorithm to achieve the same results.

. BACKGROUND: METHODS FOR FINDING ORIENTATION

.1. *Orientation detection in texture*

Researchers in texture recognition and discrimination have dealt with orientation n many ways. Most methods, including Gabor filters, wavelets, and other ubband representations, incorporate orientation information by using a small set f filters at pre-specified angles and scales (Malik and Perona, 1990; Bergen and andy, 1991; Cohen and You 1991; Jain and Farrokhnia, 1991). These epresentations may change appreciably when a pattern is rotated slightly. In nany applications it is desirable to explicitly extract the dominant orientation, nd then proceed with modeling or recognition relative to that orientation.

To extract orientation explicitly, researchers have explored methods using local lerivatives (Kass and Witkin, 1987; Rao and Schunck, 1991), moments in the patial and Fourier domains (Rosenfeld and Kak, 1982; Bigün and Granlund, 1987), and the Fourier spectrum directly (Bajcsy, 1973; Chaudhuri *et al.*, 1987).

.2. *Methods based on Gaussian derivative filters*

t is known that the frequency and spatial bandwidths of Gaussian derivatives natch well with the receptive fields of the primate striate cortex (Webster and De Valois, 1985; Young, 1986). Also, the Gaussian derivatives provide good imultaneous localization in the spatial and frequency domains (Young, 1986).

Another reason for using the derivatives of a Gaussian is that they can be steered' to any orientation by a linear combination of basis filters (Freeman and Adelson, 1991). For an *n*th Gaussian derivative, only $n + 1$ basis filters are needed to compute any orientation. Although it is understood that humans use nany more than $n + 1$ filters in parallel, the steerable filters achieve similar letection with considerably less computation.

A tool called the steerable pyramid was developed by Freeman and Adelson (1991) to estimate local orientation at multiple scales using steerable filters.

.3. *Steerable pyramid*

The bottom level of the steerable pyramid (level 0) is the original image, and each nigher level is obtained by filtering and subsampling the previous level. At each evel, steerable filters are used to estimate orientations.

The directional filter at a given level can be 'steered' to any orientation using our basis filters. Each basis filter is directional with angular tuning equal to that of a third derivative of a Gaussian. The radial tuning is not the same as a third lerivative of a Gaussian, but has been adjusted to provide a flat frequency response for the pyramid. These filters and examples of their use are shown in Freeman and Adelson (1991) and Simoncelli *et al.* (1992).

To make the orientation estimation independent of the input phase, i.e. so the response to an oriented step edge is the same as the response to an oriented line,

the Hilbert transform of the directional filter is computed (Ziemer and Tranter, 1990). An approximate Hilbert transform is found which is steerable, and the basis filters corresponding to the Hilbert transform are also applied to each level.

In summary, at each of the four levels, and at each pixel in a level, the filters output a single dominant orientation and its strength. Details of how orientation and strength are obtained from the filters are reviewed in Appendix A.

2.4. *Adaptability in scale*

A primary reason for using steerable filters is that they provide information at all orientations for a relatively small amount of computation. Correspondingly, it is desirable to find a set of 'scalable' filters which give information for orientations at all scales, without having to have a filter at every scale.

One major problem with such adaptability in scale is that the condition for it is in direct conflict with the Nyquist theorem. To avoid aliasing caused by sub-sampling, a filter should have a limited bandwidth in the frequency domain. But a filter with limited frequency bandwidth will have infinite extent in the spatial domain. To get adaptation in scale requires a compact region of support in the spatial domain. It is impossible to satisfy both of these conditions (Simoncelli *et al.*, 1992). One solution is to maintain full resolution in one of these parameters. Another solution is to design an approximate adaptive scale representation as described by Perona (1991) and Simoncelli *et al.* (1992) where a certain amount of joint aliasing is introduced. Additionally, adaptation in scale combined with adaptation in orientation can be cumbersome.

To reduce computational cost one can apply the filters on only a small set of discrete scales. This means that only a limited resolution in the scale space can be obtained. However, Andersson (1992) argues that for most low-level events such as line and edge elements, a limited resolution in scale is less severe than a limited resolution in orientation; these elements are present over several scales but only exist within a well-defined orientation. This implies that exact adaptation in orientation is more important than exact adaptation in scale except for the case of sine gratings which exist only at particular frequencies.

It would be nice if a small set of scales could be found to be sufficient for orientation analysis. Wright and Jernigan (1986) show that if filters are polar separable in the frequency domain then along the radial frequency direction, six overlapping Gaussian-shaped filters are effective for coding texture information. Along the angular direction φ, they indicated their results were not conclusive. It appeared to them that at least seven Gaussian-shaped filters with orientations uniformly spaced along the range ($-90° < \varphi < 90°$) are required to code white noise along this dimension irrespective of the radial spatial-frequency content of the image. The textures that they tested were all given by polar separable Gaussian random fields differing primarily in their local power spectra. However, their study is a significant first attempt to find the relevant scales needed to characterize textures.

For the orientation analysis in this research, the Freeman and Adelson steerable pyramid was used to analyze orientations at four different scales. Since the image sizes are 256×256, only four levels of the pyramid, *level* $= 0, 1, 2, 3$, were used with the subsampled image at level 3 being 32×32. Because of the kernel size of the filters, 17 × 17, any level higher than 3 will not give accurate information.

3. COMPUTER METHOD FOR ORIENTATION DETECTION

The method used here is closely related to looking at the first-order statistics of orientations. Hence, it is in the same spirit of Julesz's argument that first-order statistics of textons are important for visual texture discrimination, where orientation is one feature of a texton (Julesz, 1991). The computer orientation finding procedure is summarized as follows.

(1) At each level (0–3) in the pyramid:
   (a) at each pixel, find the dominant orientation and compute its strength, S;
   (b) accumulate S into an orientation histogram and smooth the histogram;
   (c) analyze the orientation histogram to assess the number, angle, and salience of dominant orientations; and
   (d) compensate for contrast (current version of algorithm only applies the compensation at pyramid level 0).

(2) Combine information from the different scales to decide the total number of dominant orientations and their angles.

3.1. *Finding orientation and strength*

The dominant orientation θ_0 and its strength S are found at each pixel in the region of interest using the steerable filters as outlined in the previous section. This is repeated at each level of the pyramid, and then the strengths at each orientation are accumulated into a histogram for each level. Thus, the output of this stage is four histograms for each texture.

3.2. *Orientation histogram smoothing and analysis*

After the orientation and its strength are found at each pixel, these values are accumulated into a histogram, H. The horizontal axis of the histogram indicates the angle from −90 to +90 deg and the vertical axis indicates the total strength of sites at each angle. In practice it is necessary to quantize the horizontal axis. Here the horizontal axis is divided into $b = 158$ bins giving angular quantization of 1.14 deg. As the angular bandwidth of the filter is much larger than 1.14 deg, this particular choice of $b = 158$ is sufficient.

More precisely the histogram entries can be written as:

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

where $N_\theta(k)$ is the sum of the strengths associated with all points having an angle in the interval: $-90° + 180°\, k/b \le \theta < -90° + 180°\,(k+1)/b$. The normalization of $N_\theta(k)$ ensures each histogram sums to 1, i.e. it can be thought of as a probability mass function. Thus, the histogram reflects the first order statistics of the orientations, weighted by their strengths. Also, the normalization facilitates comparison of the histograms over the four pyramid levels.

This definition of H is similar to the orientation histogram defined by Tamura *et al.* (1978) and Rao and Schunck (1991). Tamura *et al.* set $b = 16$ and imposed a minimum threshold on the strengths before accumulating them. They followed this with a strict algorithm that permits at most two peaks. Rao and Schunck 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 general method for finding peaks is given in Section 3.2.2.

To find the dominant orientations in a local region in the image, the orientation histogram for that region is analyzed. This involves smoothing the histogram to reduce the noise and finding the prominent peaks associated with the dominant orientations.

3.2.1. *Histogram smoothing.* The orientation histograms are noisy; therefore, some form of smoothing is necessary for finding the prominent peaks. As there is no optimal way to predict precisely how much smoothing is needed, different sizes and standard deviations of a 1D Gaussian filter were run on all 111 test images until a filter was found that smoothed the data yet kept the visually prominent peaks. 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.

The 1D Gaussian filter, $g(x)$, has the following form:

$$g(x) = \frac{(1.20)}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right), \tag{2}$$

where $\sigma^2 = 16$ and $x = 0, \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$ to give eleven points (convolving twice with this function is equivalent to convolving once with a scaled Gaussian having $\sigma^2 = 32$). The scale factor in front ensures that the coefficients of the filter sum to one. The twice smoothed histogram H will be designated H_s.[2]

3.2.2. *Finding the prominent peaks.* Part of this research includes studying how much of the salient orientation information can be retrieved from the histogram 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, as well as some similarity criterion to tell when the fit is good. Also, Gaussians are of infinite duration so they would need to be truncated, and more importantly, they are symmetric, implying they are best used when the underlying 'ideal' peaks are symmetric. A heuristic study of the 111 histograms showed that not all peaks of interest are symmetric like Gaussians. A Gaussian fitting method will tend to fit two or more Gaussians to the asymmetric peaks, which will not typically correspond to the number of different dominant orientations perceived. The method used below is therefore based on a peak being a local maxima surrounded by local minima, with no assumptions of peak shape.

To find the prominent peaks, the local extrema of the smoothed orientation histogram are detected. The local extrema can be found by approximating derivatives of H_s with first-order differences:

$$\Delta H(k) = H_s(k+1) - H_s(k), \quad 0 \le k \le b-2,$$
$$\Delta H(b-1) = H_s(0) - H_s(b-1), \tag{3}$$

where Eqn (3) is true since H_s is periodic, i.e. $H_s(b) = H_s(0)$. The zero crossings of ΔH indicate the local extrema of H_s.

For a strongly oriented pattern made up of line segments at a particular direction, whose width and spacing correspond to the size of the directional filter and its frequency tuning, there will be one peak in the orientation histogram. The peak will be prominent, i.e. narrow and large in height. If the pattern is not strongly oriented, or if the spacing and size of the structures do not correspond with the parameters of the directional filter, or if there are structures at many orientations, then the peak will be wider and its magnitude correspondingly smaller.

A 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 1 shows the orientation histogram calculated for Test3 and the corresponding graph of ΔH. As can be seen, the zero crossings correspond to the local extrema, and the inflection points correspond to the steepest part of the slope (positive or negative) on either side of a local extremum. A positive inflection point before a negative inflection point indicates that the zero crossing corresponds to a local maximum in H_s.

The vertical difference, d_v, between the inflection points gives a measure of the magnitude and steepness of the peak in H_s. The horizontal distance between the inflection points, d_h, gives the narrowness of the peak. These distances are marked on Fig. 1. Let $H_s(\theta_p)$ be the height of the histogram at the peak being considered. The following is proposed as a measure of the 'salience' of a peak:

$$\gamma = H_s(\theta_p)\,\frac{d_v}{d_h}\,w_m w_b, \tag{4}$$

where the weighting functions w_b and w_m are motivated below. Even though d_v is dependent on the magnitude of the peak, to make sure that γ is much smaller for small valued peaks than for large valued peaks, the peak magnitude $H_s(\theta_p)$ is included in the calculation of γ, shown in Eqn (4).

Motivation for w_m. There are a number of orientation histograms where the value of a peak in the histogram is close to the value of one or both of its neighboring minima. Figure 2 illustrates one such histogram. In these cases, the ratio d_v/d_h falsely signifies 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 Fig. 2, the peak in the histogram denoted with * will have a large d_v/d_h because it is sharp on one side but it is clearly not a prominent peak (as can be seen in its image, Test30, in Fig. 6).

The proposed weighting function w_m fixes these cases. 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 is expressed as:

$$w_m = 1 - \frac{MIN_{val}}{MAX_{val}}. \tag{5}$$

Motivation for w_b. There were cases where the peak in the orientation histogram was broad but the slopes of the curve were sharp. The salience measures for such peaks were high even though the orientations associated with these were not chosen by the subjects to be prominent. Very broad peaks are usually caused when there are structures at many orientations. To make sure that these peaks are not considered to be prominent (to agree with human perception), a multiplicative weight w_b is included in the salience measure for peaks broader than some trained value w_t. After considering the data for humans on all the images, values $w_b = 0.10$ and $w_t = 72$ deg were selected.

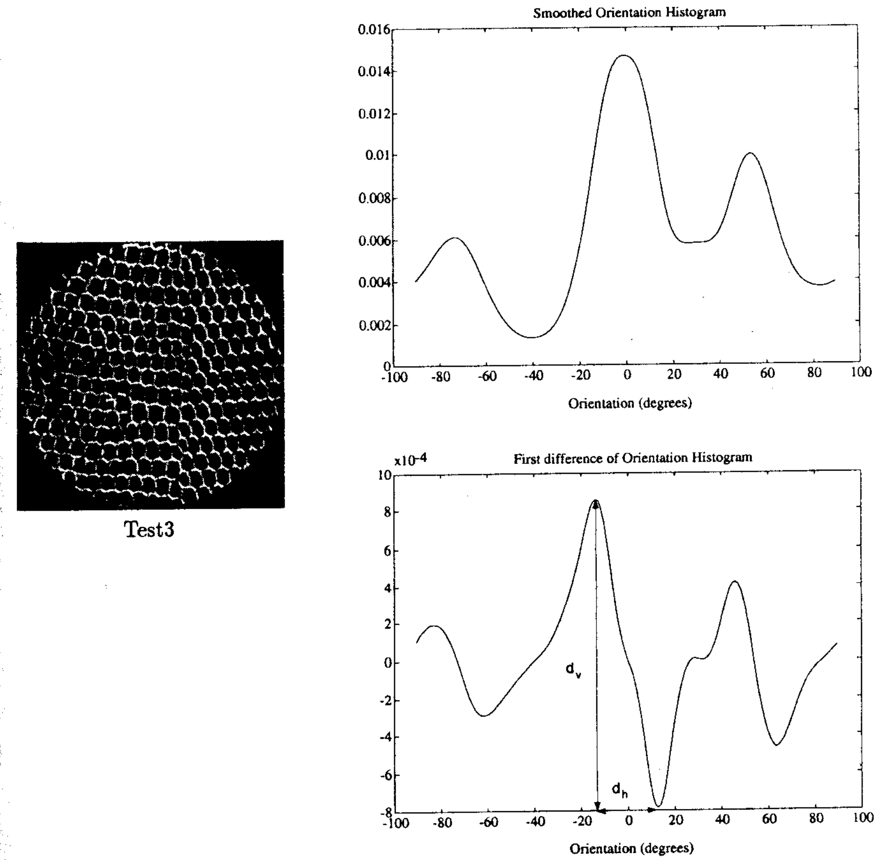


Figure 1. Brodatz image D3 used in tests, its smoothed orientation histogram H_s, and its difference histogram, ΔH. Two of the values used in the salience measure are marked on ΔH.

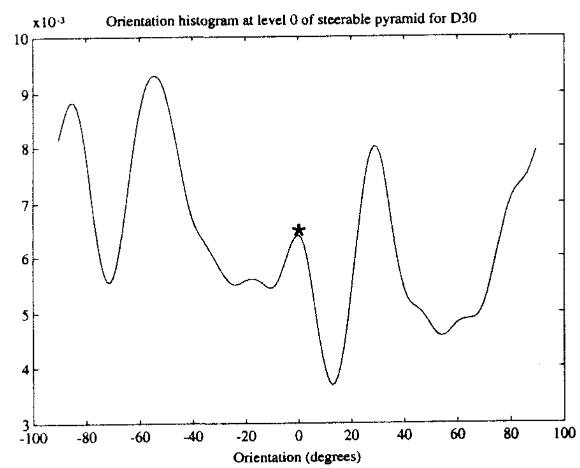


Figure 2. Histogram that motivates choice of w_m weighting function in salience measure γ.

The peak-finding method proposed in this paper is plainly heuristic and therefore has 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 four levels of resolution for each, the peaks found by the algorithm in the 444 histograms corresponded well with the peaks found by the humans. Details and illustrations of the peak-picking method are given in Gorkani (1993). This measure is computed for all candidate peaks.

Note that it is ultimately desirable that a salience measure of a peak corresponds to the strengths perceived by humans. This can be considered a long-range goal of this research, but will not be used to constrain the particular salience measure here. The primary use of γ in this research is to reduce all the data in a peak down to a decision of whether or not the peak corresponds to a perceptually dominant orientation.

3.3. *Contrast compensation*

One problem with orientation analysis using directional filter outputs is that the output energy of the filters increases with increasing image contrast. A non-oriented pattern with high contrast can give a response similar to an oriented pattern with small contrast. One could remove contrast effects completely, but since a high-contrast oriented structure is more strongly perceivable than a low-contrast oriented structure, removing contrast information could lead to results that disagree with the human perception of similarity. Several nonlinear transformation methods used to enhance the low-contrast oriented structures have been explored on the Brodatz textures (Gorkani, 1993).

Since the estimation of local orientation θ_0 is a ratio of responses from the steerable filters (Eqn (A3)), 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 (Eqn (A4)) is dependent on the contrast of the image.

Note that if the filters are large compared to the structure at the orientation they are detecting, or if the area of the structure is small compared to the region over which the filters are applied, then the filter response may be small compared with the human perception of the orientation. Even though the primary issue in these cases is one of relative size (and shape), and not one of contrast, the contrast normalization may still have the effect of boosting the filter response.

One way to model the contrast normalization used in low-level image perception is to divide the energy of the filter outputs by the sum of energies corresponding to filters at all orientations in a local neighborhood (Shapley, 1990; Heeger, 1991). For example, Bergen and Landy (1991) normalized for contrast by dividing each of the energy outputs of four directional filters by a local average of all their energy outputs.

For steerable filters this contrast normalization is difficult since it can affect the steerability of the oriented filter. The directional filters in the steerable pyramid have an orientation tuning proportional to $\cos^3(\theta)$. Squaring the outputs (for energy) of these filters means that the resultant images have a finer orientation tuning approximately proportional to $\cos^6(\theta)$. Because of the presence of the finer $\cos^6(\theta)$ component, one now needs seven basis filters instead of four, to interpolate the oriented energy $E(\theta)$ to any orientation.

The contrast normalization method used here begins with seven equally spaced samples of $E(\theta)$ at $\theta_s = 180°s/7$, $s = 0, 1, \ldots, 6$. This is the same as taking the energy of the output of a directional filter and its approximate Hilbert transform rotated at seven equally spaced orientations.

The seven sampled oriented energies $E(\theta_s)$, $s = 0, 1, \ldots, 6$ are normalized in the following way at each pixel position (x, y):

$$E_n(\theta_s)(x, y) = \frac{E(\theta_s)(x, y)}{c + \sum_{(j,k)\in\mathcal{N}(x,y)} E(\theta_s)(j, k)}, \tag{6}$$

where $E_n(\theta_s)$ is the normalized energy, $\sum_{\mathcal{N}(x,y)} E(\theta_s)$ is the sum of $E(\theta_s)$ in an isotropic local neighborhood about (x, y), and $c > 0$ is a constant used to prevent division by zero if $E(\theta_s) = 0$. It is important to choose 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 best choice of the size of the local neighborhood $|\mathcal{N}(x, y)|$ for calculating E_s is also an open problem. One way to find a good size is to calculate Eqn (6) for different sized neighborhoods, and see which size gives results closest to those for humans.

The normalized energy $E_n(\theta_s)$ can be expressed in terms of the seven normalized energy outputs using the coefficients needed to steer a filter with angular tuning $\cos^6(\theta)$. The estimated orientation θ_0 and its strength S described can be found for each position (x, y). In this case, C_2 and C_3 in Eqn (A4) will be combinations of the seven normalized energy outputs. These provide new strength values used to form a contrast-enhanced orientation histogram, $H_c(\theta)$.

The energy normalization described here was implemented for the first level only of the pyramid, level $= 0$. Future work is planned to extend this to the other levels. The range of values for the energies of the directional steerable pyramid filters were found for this level and the constant c was chosen to be an order of magnitude smaller than the minimum value in this range ($c = 1$). The neighborhood size was chosen empirically to correspond to the blur of the third level of a Gaussian pyramid, i.e. a 31 × 31 region centered around the current pixel.

For Test6 shown in Fig. 3 subjects gave high strengths to both the horizontal and vertical orientations. However, as can be seen in Fig. 3, the peak at 90 deg in the histogram of orientation strengths $H_s(\theta)$ has a tiny value. In the $H_n(\theta)$ histogram, which was calculated by summing the number of sites having θ as dominant orientation, the peak at 90 deg is more prominent but still smaller than the peak at 0 deg. Only in the contrast-normalized histogram $H_c(\theta)$, formed using the normalized energy $E_n(\theta)$ at level 0, is the peak at 90 deg close in value to the peak at 0 deg.

The contrast-normalized histogram greatly boosts low-contrast structures which may or may not be directional; therefore, it must only be used selectively. Based on matching human and computer results, it was found that the results improved on six of the test images and were not diminished on any images if the following (nonlinear) decision was applied first:

Contrast-normalization condition: At least two peaks must correspond between the original histogram at level 0 and the contrast normalized histogram, and in the original histogram one of the peaks must have a salience measure above the threshold and the other peak a salience measure below the threshold.

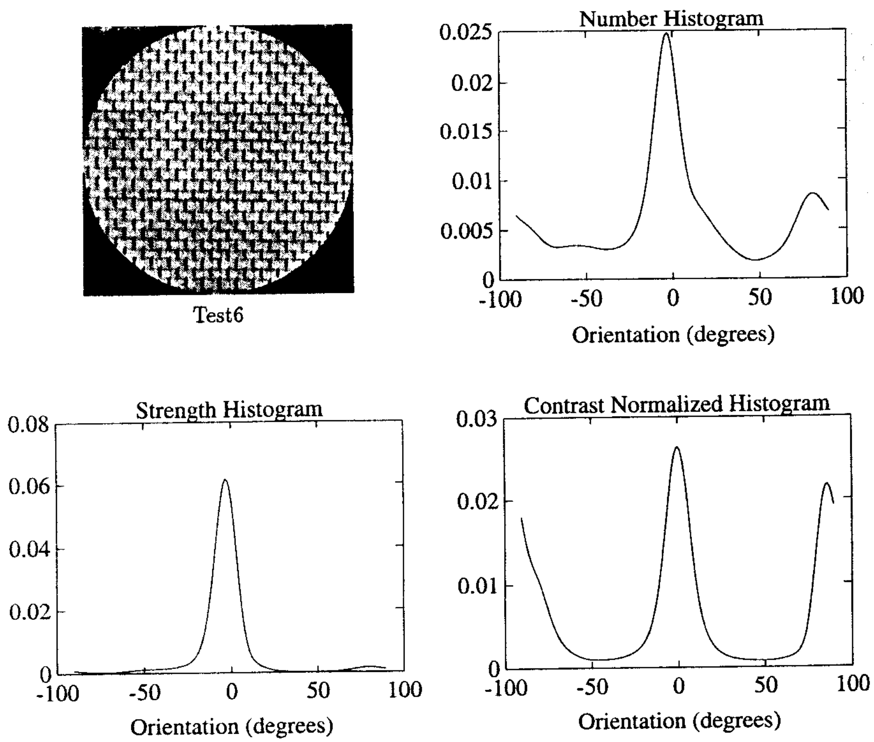


Figure 3. Clockwise from upper left: Test image D6; H_n, histogram of number of sites at each orientation; H_s, histogram of orientations weighted by strengths; H_c, contrast compensated histogram.

If this condition is true and the ratio of the peaks' heights 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. This condition was run on all the 111 Brodatz textures. It was found only to improve the agreement between the computer algorithm and human responses; see Section 5.6 for these results.

Consider Test6 shown in Fig. 3. Since the ratio of the peaks at 0 and 90 deg in the contrast-normalized histogram is more than 80% and in the strength histogram is less than 20%, then the orientation 90 deg is chosen to be dominant.

3.4. *Combining orientation information from different scales*

In the next section a study will be described where humans label orientations they perceive to be dominant. The results for humans thus produce one histogram for each image. The method described above produces four histograms for each image, one at each level of the pyramid. Consequently, it is necessary to combine the four histograms for comparison to the one for humans. The method for combining information from different scales involves making decisions about what is important at each stage. Part of the results of this work include determining these criteria. The criteria were determined iteratively, by picking decisions, comparing them to the results for humans, then refining them so they are closer. The exact process found to combine information and give results closest to those for humans is detailed in Section 5.5.

4. HUMAN ORIENTATION DETECTION STUDY

It is necessary to have some form of 'ground truth' with which to evaluate how well the above orientation-finding algorithm works. In this section we describe a study with humans undertaken to help provide information for evaluating the algorithm. In this study, subjects were asked to designate the dominant orientations they perceived in a set of test images, and how strongly they perceived each orientation.

4.1. *Subjects*

Forty subjects from a variety of ethnic origins, ages, and academic backgrounds participated in the visual experiment. The majority of the subjects were MIT undergraduate and graduate students. None of the subjects were researchers in computer or human vision, and none had previous experience with psychophysical visual experiments. The subjects were given an ice-cream gift certificate for their participation. Of the forty subjects, sixteen were female. Subjects had normal or corrected-to-normal visual acuity.

4.2. *Experimental setup*

The subjects were seated 36 cm from a 16-in Sony trinitron monitor and the displayed image was 7 cm × 7 cm. These values were sufficient to make sure that they did not see any particular pixel in the image but still could detect fine details. The display resolution was 35 × 35 dots per cm. The lighting in the room was dim (main lights turned off except for a background lamp). This ensured that there would not be any false illumination on the images. The images of size 256 × 256 were positioned in the middle of the monitor screen. The mean luminance of the screen was 23 cd m^{-2}. The images were displayed without gamma correction. To ensure that the subjects were not influenced by the horizontal and vertical image boundaries, each image was multiplied with a disk of size 256 × 256 with a radius of 128 pixels. A red bar centered in the middle of an image would pop up in a random orientation if the user indicated that he or she saw a dominant orientation in the image. The user was shown how to rotate this bar during the training procedure. Figure 4 shows the test sequence.

When the subject was ready to indicate orientation strength, a menu bar popped up under the test image (see Fig. 4). The subject used a slider to give an integer strength value between 0 and 10. Use of a finer scale was considered but decided against as it is difficult for a subject to distinguish quantitative strength at finer levels, e.g. 6.7 instead of 6.9. After a strength was indicated, the subject was prompted as to whether he or she 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.

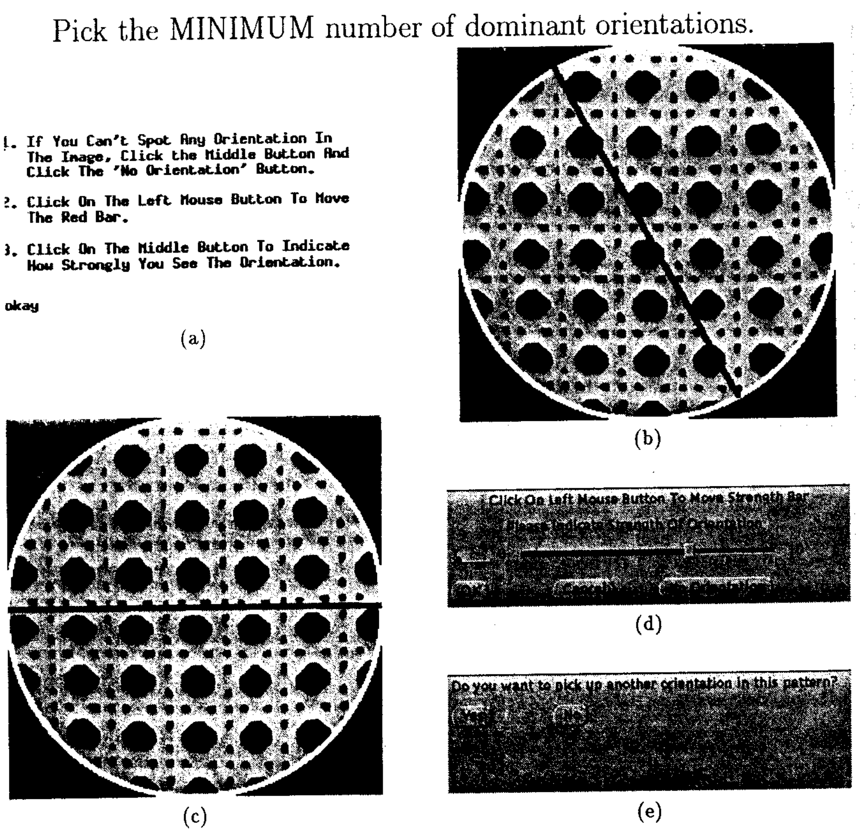


Figure 4. Human test setup: (a) The subject clicks mouse to indicate whether or not they see a dominant orientation. (b) If they see one, a red bar pops up for them to rotate. (c) One of the orientations selected. (d) Human subject specifies strength of orientation. (e) A panel pops up asking if subject sees another dominant orientation.

It is possible that using a red bar in the center of the image could bias the subject toward choosing the orientations dominant at the center. To try to minimize such effects, the bar was given the same length as the image diameter. Other methods to locate orientations were also considered such as spinning a bar around in an adjacent image, or allowing a bar to both translate and rotate in the test image. However, it was believed that the former would be less accurate in the presence of multiple orientations, and the latter might encourage the subject to template match rather than specify dominant orientations.

The choice of using 'real images' as opposed to synthetic images in the human study part of this research is unusual and introduces several complications. For example, a human shown an image with a small section of brick wall may say the dominant orientations are horizontal and vertical even if the image does not show enough bricks to reveal any periodicity. Although we did not reveal the content or name of the images to the subjects, many of them can still be recognized visually and therefore the possibility that some orientations were identified by semantic association cannot be ruled out. Of course a very important problem in computer vision is also to determine how and when semantics should be incorporated into the 'low-level' signal processing. By looking at where the computer and human results disagree and trying to understand the causes of disagreement, the importance of any semantic interaction should be revealed. The results of this analysis are in Section 5.7.

Recall that the use of real images is also to begin identifying which interactions in the real world contribute to the 'quick' recognition that things look 'similar' or 'different'. It is the role of dominant orientation in this capacity that we are trying to measure.

4.3. *Training session*

Before commencing the test, each subject went through a training session where a precise set of directions was read. The training for this experiment is difficult because it is important not to bias the subject toward a certain orientation, strength or structure. To communicate to a subject a minimal notion of 'orientation' and 'strength', two images were first shown: an ideal directional pattern and a random noise image. The first image was a sinusoidal grating oriented at 90 deg, shown in Fig. 5(a). The subject was told that this particular image has one dominant vertical orientation of strength 10, and was shown how to indicate this orientation and strength to the test system. Next, a uniform noise image, Fig. 5(b), was shown. The subject was told that this image has no dominant orientations and has 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.

The subject was also shown example images to illustrate these cases:

- dominant orientations do not have to correspond to the presence of continuous lines and they do not have to pass through the center (Fig. 5(c));
- multiple dominant orientations may be present (Fig. 5(d) and (e)); and
- there can be many orientations present that do not correspond to a dominant orientation (Fig. 5(f)).

Since different subjects might perceive the strengths of these cases differently,

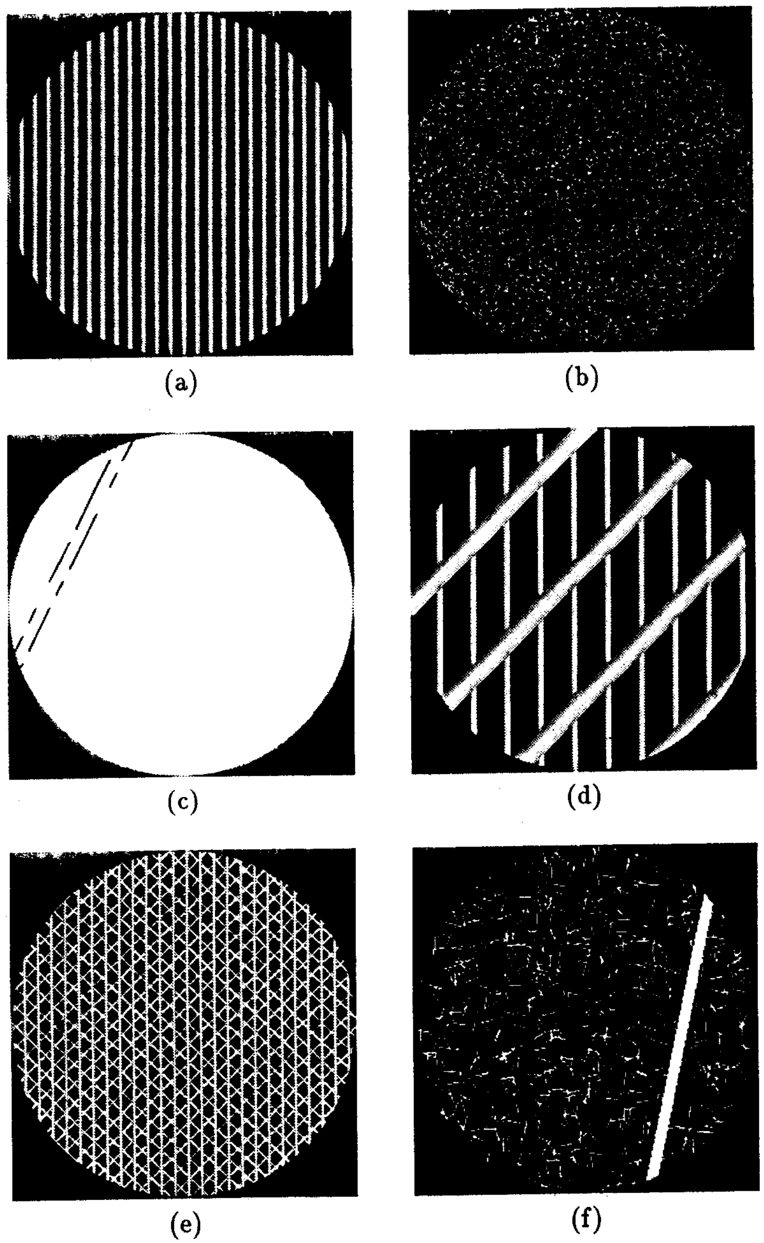


Figure 5. The top two images were used to train both strength and orientation: (a) Dominant vertical orientation with strength ten. (b) No dominant orientation. The remaining images were used only to indicate number of dominant orientations.

they were not told the strengths for these images. Instead, they were asked to pick strengths based on considering the first two examples. These images were not considered in our test results but were used only to ensure subjects had no difficulties manipulating the test system.

The words *dominant* and *minimum* were repeated several times during the training for reinforcement. The instructions were always visible in the upper left corner of the screen and a reminder was kept at the top of the screen: *Remember to pick the MINIMUM number of dominant orientations.*

4.4. *Test images*

Since it would have taken too long for a subject to analyze 111 images, the images were divided into four sets so that each subject only had to analyze thirty of the Brodatz images. Seven synthetic 'teaser' patterns from psychophysical experiments were also included for a total of thirty-seven test images for each subject.[3] Most subjects spent about 45 min on the test, including training time.

4.5. *Recording of the human visual data*

The orientations and strengths picked by each subject for each test image were recorded. Information about the number of orientations, their relative strengths, the expected number of orientations for each test image, and the distribution of strength values chosen by each subject were found. Several statistical properties of the experimental data were considered, and the following conditions were verified (Gorkani, 1993): (1) the subjects used the full range of strengths, 0–10, with an approximately uniform distribution; and (2) of the forty subjects, four had variances greater than 10% from the average range of strengths chosen. These four variances were between 10 and 19%. There was no compelling reason to remove any of these subjects from the test data; consequently, the data from all forty subjects was used in the comparisons reported here.

5. COMPARISON OF HUMAN AND COMPUTER ORIENTATION DETECTION

5.1. *A difficult optimization problem*

One can think of this research as a huge nonlinear optimization problem. There are two nonlinear systems, human and computer, each dependent on a large number of variables. The focus of this research is to solve for the computer variables so that the outputs of the two systems are as close as possible. In particular, the key variables described above can be informally summarized as follows. Details are described below.

Variables for human study data, $\mathscr{H}$. (1) Accept or reject decision on a subject's test data (for removal of outliers). (2) Histogram angular bin quantization. (3) Histogram 'peak salience' threshold, γ_h.

Variables for computer algorithm, $\mathscr{C}$. (1) Filter shape, bandwidth, frequency centers. (2) Number of pyramid levels and their scales. (3) Nonlinearity applied to filter outputs (squaring to get energy). (4) Histogram angular bin quantization. (5) Histogram smoothing. (6) Histogram 'peak silence' threshold(s) γ_c. (7) Contrast compensation: choice of neighborhood size, choice of function.

Let $\vec{\theta}_{h_I}$ be the vector of orientations found to be dominant by a human for a given pattern I and let $\vec{\theta}_{c_I}$ be the vector for the computer algorithm. There are many possible criteria to consider when comparing the results. One goal is that the humans and computer algorithm should agree on the number of dominant orientations, i.e. equal dimensions of the vectors:

$$\min_{\mathscr{C},\mathscr{H}} \sum_I (\dim \vec{\theta}_{h_I} - \dim \vec{\theta}_{c_I})^2. \tag{7}$$

The sum is over all image patterns in the test, and the minimization is with respect to all the variables in the two lists above. This criterion may be sufficient in quick search applications, perhaps just to determine if further comparison is worthwhile, or perhaps to also determine which particular model, say structured or statistical, should be fitted for further comparison.

A more exact criterion takes into consideration the strengths perceived by the human and the positions found by both human and computer. There are many ways to formulate this goal. If $\vec{\theta}_{h_I}$ and $\vec{\theta}_{c_I}$ already have equal dimensions then the following objective form provides a tighter match:

$$\min_{\mathscr{C},\mathscr{H}} \sum_I (\vec{\theta}_{h_I} - \vec{\theta}_{c_I})^T \mathbf{W} (\vec{\theta}_{h_I} - \vec{\theta}_{c_I}), \tag{8}$$

where $\mathbf{W}$ is a diagonal matrix of weights, with diagonal entries w_{ii} equal to the strengths assigned by the human to each of the orientations in $\vec{\theta}_{h_I}$. This formulation is minimized when the strongest perceived orientations are located at the same position by both the computer and human. One can continue setting up similar objective functions depending on the stated goals.

Because of the large number of variables, their infinite set of combined possible values, and the complex nonlinear interactions among them, the optimizations posed here need additional constraints to make their solutions tractable. Many such constraints have already been imposed by the choices described above in the computer algorithm (such as using 256 × 256 images, four pyramid levels, etc.). The analysis below is based on constraining most of the variables to a set of fixed values, and varying the choices of 'peak salience' thresholds for both the computer and humans. The results achieved for these choices are discussed in the next section. However, research is continuing in this general framework, with the aim of finding a small set of 'universal' values for which the algorithm most closely approximates the human.

5.2. *Analysis of human visual data*

The main objective of the visual experiment was to get human data $\vec{h}_I$ with which to compare the results of the computer algorithm $\vec{c}_I$. Disagreements between results for the humans and the algorithm can then be analyzed to learn how to improve the algorithm, and to learn what any limitations of this basic approach are.

As mentioned earlier, the human data consists of one orientation histogram for each subject for each pattern. The computer data consists of one orientation histogram for each of four levels of scale for each pattern. Before comparing the human and computer results, it is necessary to transform the two sources of data into a more comparable form.

For the initial comparison of the data, the orientations chosen by subjects for each test image will be quantized to 10 deg bins. Considering that on test images such as Fig. 5(c) the human's responses spread about 6 deg, the 10 deg quantization should not be a significant loss. The strengths associated with orientations falling in a particular bin are summed for that bin.

A total of three variables are computed from the human response to each test image I. The elements of $\vec{\theta}_{h_I}$ are quantized orientations as described above. Each element of $\vec{\gamma}_{h_I}$ is normalized by the maximum strength that could be given any element of $\vec{\theta}_{h_I}$, i.e. 10 (maximum strength) × the number of subjects responding to that test image. 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.

Table 1.
Notation for recording and comparing human responses and computer outputs.

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

Human-picked orientations: Matched to $\vec{\theta}_{c_I}$	Human-picked orientations: Rejected	Computer-picked orientations: Matched to $\vec{\theta}_{h_I}$	Computer-picked orientations: Rejected
$\vec{\theta}^{M}_{h_I}$	$\vec{\theta}^{R}_{h_I}$	$(\vec{\theta}^{M}_{c_I})$	$(\vec{\theta}^{R}_{c_I})$
$(\vec{\gamma}^{M}_{h_I})$	$\vec{\gamma}^{R}_{h_I}$	$\vec{\gamma}^{M}_{c_I}$	$\vec{\gamma}^{R}_{c_I}$

5.3. *Analysis of computer orientation histograms*

The computer algorithm produces four histograms, their peaks, and peak salience values. This section describes a procedure for reducing this information to one vector, $\vec{\theta}_{c_I}$, which can be compared to the orientations $\vec{\theta}_{h_I}$ for humans. Since the histogram values over all levels of the steerable pyramid are normalized, they can be compared to each other. The salience measures can then be ranked from lowest to highest magnitudes across the levels of the pyramid. The first M highest salience measures and their associated orientations can then be compared with the human visual data.

The average dimension of vector $\vec{\theta}_{h_I}$ is five over all test images. The average number of orientations picked by subjects over all the images was 1.3. The dimension of $\vec{\theta}_{h_I}$ is greater than this average since two humans may have each picked one orientation, but if they picked them slightly apart, then they will be recorded as two different orientations. Hence, choosing $M = 5$ as the initial number of dominant orientations to collect by computer is reasonable. M is a variable which can be increased in the future if the orientations found by the filters are not sufficient to match the human data. The maximum number of orientations picked by any subject on any single image was seven. The maximum number of orientations found by computer analysis for any single image was eight.

Orientations existing over more than one pyramid level are only picked once. It was found by looking at images where orientations existed in more than one level of the pyramid that an allowance of a 10 deg spread insures that orientations are not picked more than once in the ranking. A 10-deg spread means that, in the ranking, an orientation is not included if it is closer than 10 deg to the orientations picked in the other steerable pyramid levels.

Two variables $\vec{\theta}_{c_I}$ and $\vec{\gamma}_{c_I}$ are computed from the computer algorithm's response to each test image I; these are summarized in Table 1 with the human variables. The elements of $\vec{\theta}_{c_I}$ are the orientations whose salience measures are included in the top M measures over the four levels of the pyramid. The elements of $\vec{\gamma}_{c_I}$ are the salience measures for these orientations.

5.4. *Choosing variables for evaluation*

The focus now is on comparing the data from the humans 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 image I.

During the comparison, there are four cases that can occur; these are summarized in Table 1. The new vector $\vec{\theta}^{M}_{h_I}$ consists of the elements of $\vec{\theta}_{h_I}$ for which matches were found in $\vec{\theta}_{c_I}$. For the present algorithm, this is all orientations within 10 deg of those found by the filters. (This considers only position at this point, not saliency.) 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 deemed important by the humans. The values in $\vec{\gamma}^{R}_{c_I}$, 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 $\vec{\gamma}^{R}_{c_I}$ over all the images is used for the initial salience threshold, $\gamma_c = 0.085$ to decide if a peak is a dominant orientation. A few cases where $\vec{\gamma}^{R}_{h_I}$ is important (the humans found orientations that the computer could not match) are discussed in the 'difficult cases' in Section 5.7. The values of Table 1 shown in parentheses are not necessary in the results reported here.

To determine the threshold γ_c consider:

$$r_c = \text{number of elements of } \vec{\gamma}^{M}_{c_I} < \gamma_c, \tag{9}$$

a measure of the 'wrong' rejections caused by this threshold. In other words, orientations found by both the computer and human should have corresponding strengths greater than γ_c, i.e. it is desirable that $r_c = 0$. A similar rejection measure can be formed for the human data:

$$r_h = \text{number of elements of } \vec{\gamma}^{R}_{h_I} < \gamma_h, \tag{10}$$

where γ_h is a (typically very low) value used to ignore some stray low-strength orientations from the human data. The threshold γ_h is set to 0.15 in this initial evaluation. Choosing this value for the threshold means that the strength of the values $\vec{\gamma}_{h_I}$ should be at least 15% of the maximum strength that they can be assigned. A couple of the significant cases are discussed in the sections below.

5.4.1. *Case 1: Effect of γ_c on choosing peaks at different levels of pyramid.*

The computer made orientation decisions for the 111 Brodatz images using the method above by applying one threshold γ_c to the salience measures at all levels of the pyramid. The orientations (number and position) found by computer were compared with the same values found by humans. The results of this are shown in the first line of Table 2. A total of 59/111 of the images achieve 100% agreement between the humans and the computer.

Table 2.
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 data from the human study and computer algorithm.

Dominant orientations computer found	0	1	2	3	4
With fixed threshold, γ_c	23/32 = 72%	16/43 = 37%	17/28 = 61%	1/4 = 25%	2/3 = 67%
With different γ_{c0} – γ_{c3}	32/41 = 78%	18/44 = 41%	15/19 = 79%	1/4 = 25%	2/2 = 100%
Gain of As (agreement) (+)	+9	+2	0	0	0
Loss of As (−)	−1	−1	−2	0	0

The study revealed that images found by the computer to have no dominant orientations showed the best agreement. Images for which the computer found peaks showed more disagreement. A careful comparison of the strengths assigned by humans and the salience measures found by computer revealed that raising the salience threshold on the highest pyramid level was necessary to get better agreement with the human data. It was also found that better results could be achieved at levels 0–2 by lowering the threshold. In other words, at coarser scales the orientation needs to be stronger for it to match the humans' perception as being dominant.

To pick the new level-dependent values of γ_c: γ_{c0} – γ_{c3}, an iterative procedure was followed with the goal of providing the maximum agreement between humans and computer on the 111 images. The resulting values of the pyramid-adaptive thresholds, γ_{c0} – θ_{c3}, are shown in Table 3. Several details on this comparison can be found in Gorkani (1993).

5.5. *Results with different thresholds for different pyramid levels*

Table 2 summarizes the agreements between human and computer when a fixed threshold is used for all levels of the pyramid, and when a different threshold is used at each level. Agreement occurs when $r_c = 0$ and $r_h = 0$, with $\gamma_h = 0.15$ and

Table 3.
Threshold values chosen for different pyramid levels 0–3. Notice they are greatest at the coarsest (top) level.

Threshold	Values
γ_{c3}	0.418
γ_{c2}	0.075
γ_{c1}	0.034
γ_{c0}	0.0098

with $\gamma_{c0} - \gamma_{c3}$ taking the values in Table 3. Each column of the third row shows 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 plus and minus signs start to balance, one expects the thresholds are nearer the critical points.

It can be seen from Table 2 that using multiple thresholds enables better agreement with the results for humans. These results seem to confirm that coarser scale orientations must be correspondingly stronger than fine ones to be perceived as dominant. Detection of textures with no dominant orientations and with from two to four dominant orientations is significantly improved by this method. However, the algorithm still has problems classifying textures to have one dominant orientation. (The poor results for the three-dominant-orientation case are not conclusive since the sample size in that case is minuscule.)

Many observations have been made during analysis of the disagreements in this form of the algorithm, details of which can be found in Gorkani (1993). In the rest of this paper we will focus on the analysis of the best performing version of the algorithm, the above version augmented with contrast compensation.

5.6. *Results after incorporating contrast normalization*

Figures 6–12 show all the Brodatz test images grouped by their number of dominant orientations as found by the computer, using the different thresholds at each level and the nonlinear contrast normalization. An 'A' underneath an image denotes *agreement* between the human and computer on choosing the number of orientations[4].

After the nonlinear contrast normalization described in Section 3.3 was applied to all 111 textures, the following new images agreed with the humans: Tests18, 25, 50, 52, 53, and 96. Thus the contrast normalization helped in six cases, while introducing no new disagreements.

Table 4 summarizes the improved results obtained using the multiple thresholds and contrast compensation. Other issues that might be considered are as follows: (1) If there is only one low-contrast orientation in the texture then it may not be found using this current method. (2) Only the relative values of the contrast normalized histograms have been considered so far; more of the peak shape information could be compared. (3) The contrast normalized histogram can also be used at higher pyramid levels.

Table 4.
Table summarizing current results. The maximum for each entry is 111.

	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

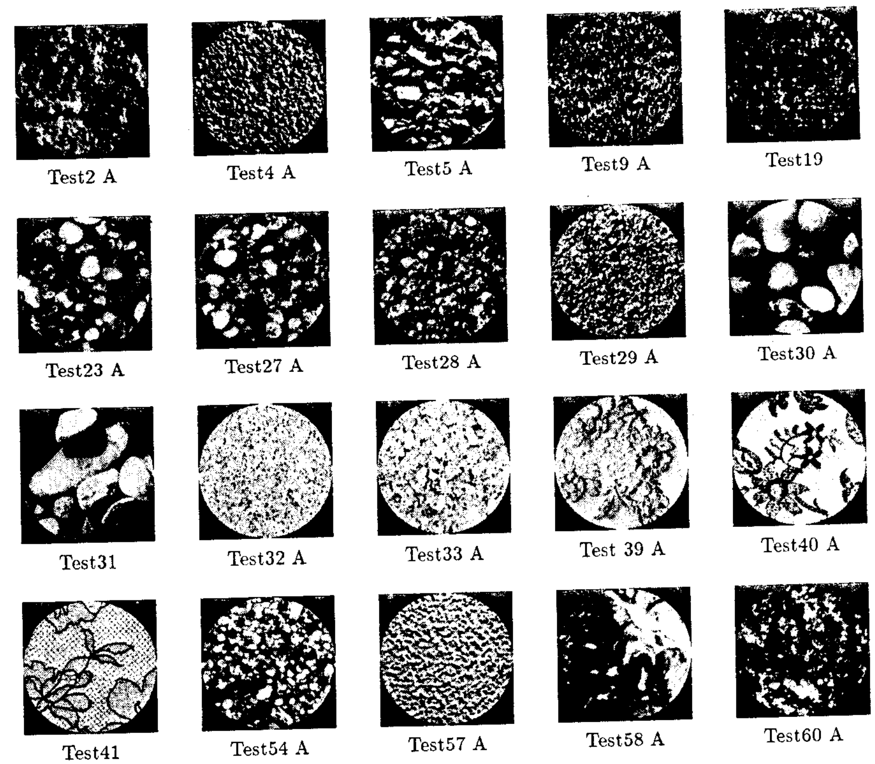


Figure 6. No dominant orientations were found by the computer in these. An 'A' under an image indicates agreement with the human study data.

5.7. *Difficult cases and future directions for comparison*

There are a number of difficult cases which remain. This section analyzes the algorithm's weaknesses, to help researchers better predict where it should fail or succeed.

Two images where the algorithm agreed somewhat with the humans, but not completely, are the lizard-skin patterns in Tests22 and 35 (see Fig. 8). For both images, the humans picked three orientations with relative strengths much bigger than γ_h. However, the corresponding peaks found by the filters were too broad and not sufficiently prominent. One of the reasons that peaks are not prominent (and perhaps the key reason in these cases) is that the textures are non-homogeneous over the region being analyzed. Repeated analysis over smaller subregions may produce better peaks. (The human eye may also wander over these subregions.) Finding the ideal-sized region over which to analyze orientation is a difficult problem; it may be best studied jointly with attention.

Orientation information is known to be useful in helping detect symmetry (Bigün and du Buf, 1992). There are two cases in this study where failure of the computer algorithm to agree with the human subjects appears to be caused by the subjects confusing symmetry with dominant orientation. The first was Test6 (Fig.

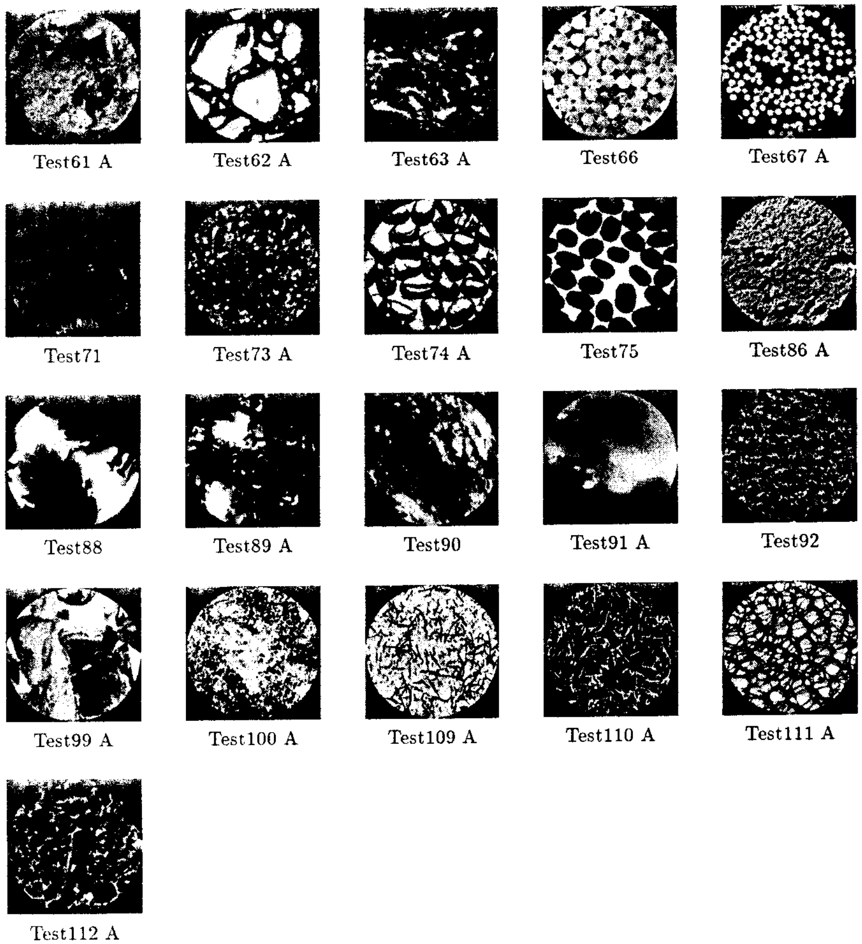


Figure 7. Figure 6 continued.

10) where using the computer compensates for contrast and picks vertical and horizontal, but the subjects picked vertical, horizontal and the diagonals. The second was Test14 (Fig. 10) where the computer decides no contrast compensation is needed and picks vertical and horizontal, but the subjects picked vertical, horizontal and the diagonals. In both cases, very few subjects picked the diagonals, but those who did gave them high strength.

In both these cases, it is possible that the diagonal orientations picked by the humans could be found by a cascade of the steerable filters, i.e. apply once, then apply some nonlinearity to their outputs, then apply steerable filters to this subsequent output. Consider also the hypothetical case of a texture composed of zero mean white noise modulated with a sinusoid. For a broad range of the modulation frequency, directionality is easily perceived by the human, but cannot

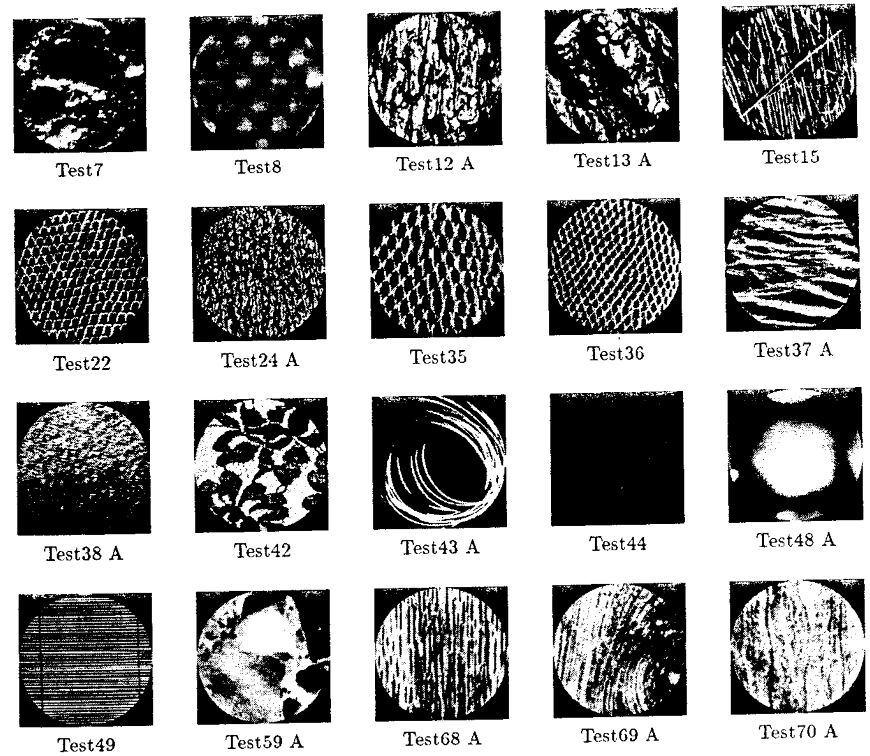


Figure 8. One dominant orientation was found by the computer in these. An 'A' under an image indicates agreement with the human study data.

be detected by the algorithm we have presented, even though the algorithm low-pass filters and combines information over multiple scales. However, if the stimulus contrast is first rectified, then the orientation is easily detected by the algorithm here. How best to combine orientation detection with nonlinearities is an important area open for research.

There were 111 − 95 = 16 textures that did not agree in any dominant orientations. Of these, one disagreed (Test44) because a strongly oriented part of the pattern is hidden by the circular disk, and hence visible to the filters (run on the square image) and not to the humans (who saw only the round images). This was the only pattern which seemed to be dramatically affected by this choice of implementation.

The 16 'difficult' images are shown in Figs 13 and 14. Beside each image in the figures the following information is given:

- How many orientations were chosen by human?
- How many orientations were chosen by computer?
- What were the orientations chosen by humans (the relative strength indicated in parentheses)?

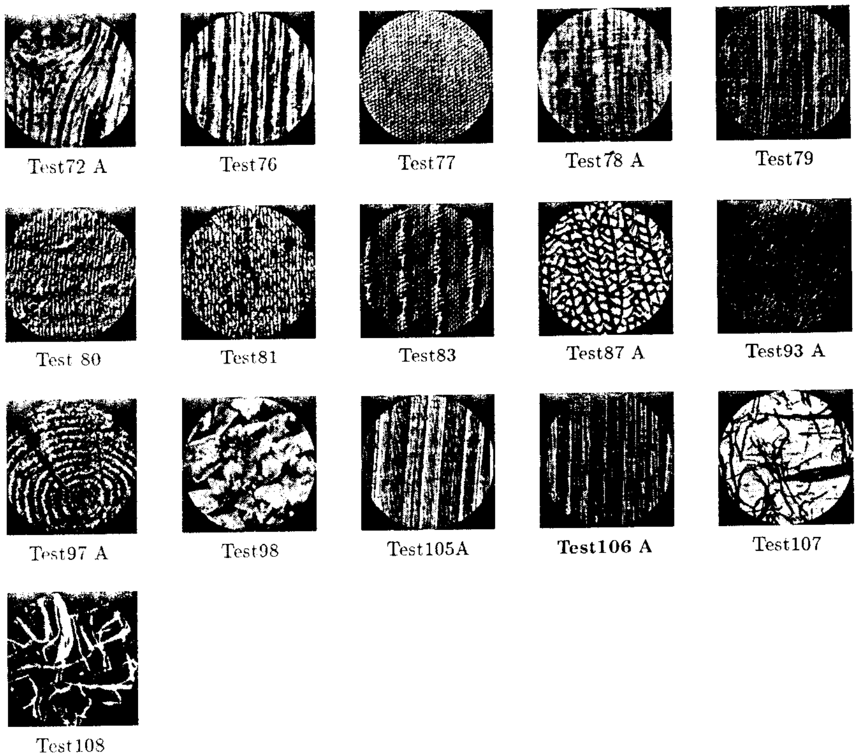


Figure 9. Figure 8 continued.

- 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 parentheses.
- Speculation about why there were no matches.

A set of size 16 is small considering the large variety of data in the original set of size 111. Also the results here are only from a first attempt at optimization. Nonetheless, there are significant problems raised in this set of images. One of these problems is the filter shape and size. In images such as Test77 the orientation information is fine and should be detected at the bottom level of the pyramid. However, the filters do not detect a salient 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 narrower, perhaps achievable by using higher Gaussian derivatives. A similar problem appears to occur in Test90, but at the other extreme, where the filters are not coarse enough.

Several of these patterns contain evidence for processes that may be higher-level than simple orientation detection. In particular there is evidence for grouping of objects (Tests66, 75), of coarse-level segmentation of objects (Tests41, 88, 90), and perhaps also of object identification with shadow removal

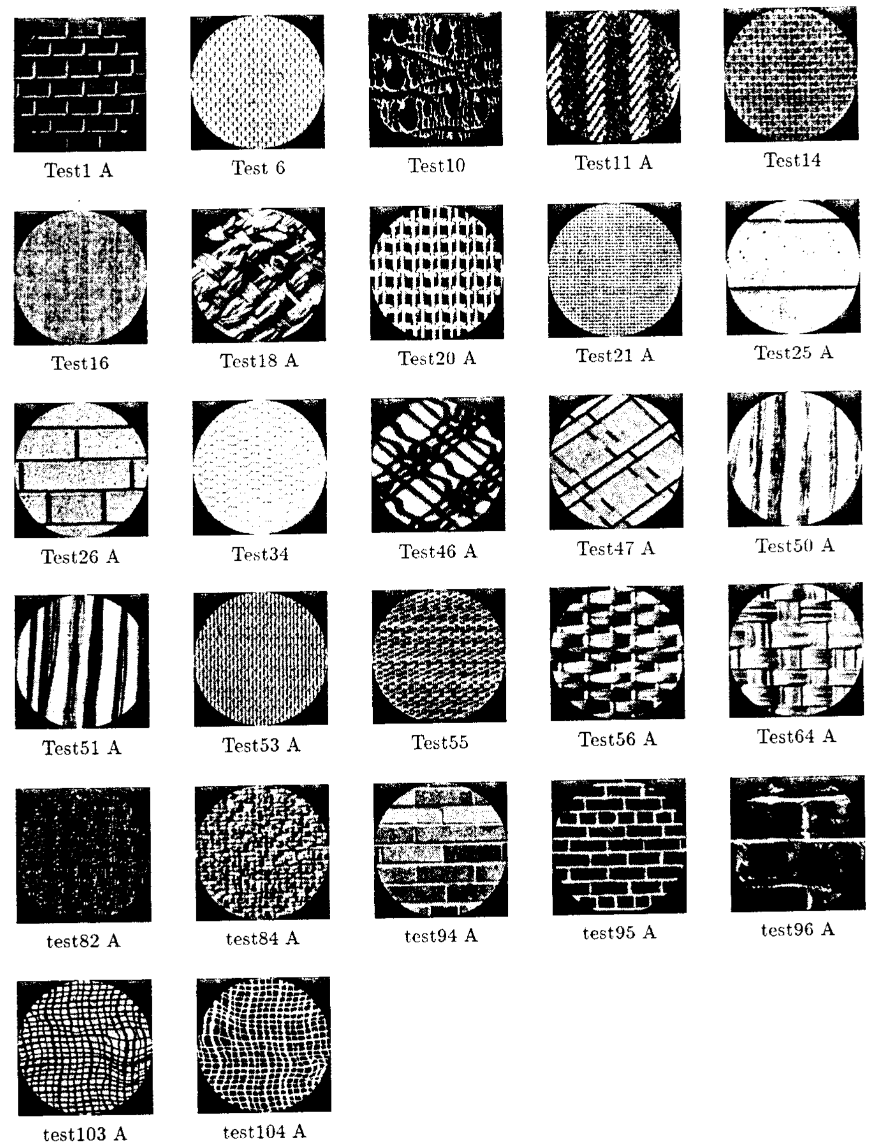
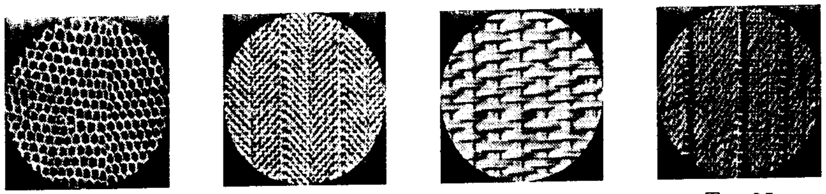
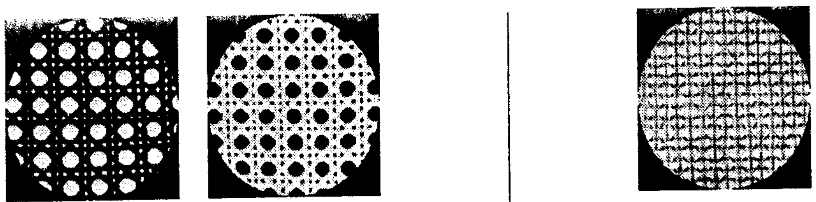


Figure 10. Two dominant orientations were found by the computer in these. An 'A' under an image indicates agreement with the human study data.



Test3 A Test17 Test65 Test85

Figure 11. Three dominant orientations were found by the computer in these. An 'A' under an image indicates agreement with the human study data.



Test101 A Test102 A Test52

Figure 12. Four dominant orientations were found by the computer in the two images on the left, and five in the rightmost image. An 'A' under an image indicates agreement with the human study data.

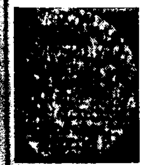
(Test31). The latter two phenomena may be strongly influenced by semantic interpretation, although this is a difficult cause to verify. There may also be some enhancement of the human-perceived orientation in Test88 due to the high contrast along the diagonal where the image meets the black background.

Many of the other difficulties present in 'the difficult 16' look like they may be fixed by improvements in the contrast handling of the algorithm, especially by including contrast handling at higher pyramid levels, or by adjusting the processing more locally. Local processing should help especially for images like Test71 where the directional pattern is inhomogeneous.

6. SUMMARY

This study has compared the 'dominant orientations' perceived by forty human subjects with the dominant orientations found by a multi-scale orientation detection algorithm using contrast normalization. It was found that using different thresholds with four levels of a steerable pyramid is sufficient to detect all the perceptually dominant orientations of 68 out of 111 textured images without even using contrast normalization. Using contrast normalization brings this number up to 74. If the algorithm is asked to detect at least one perceptually dominant orientation chosen by the human subjects then the success rate rises to 95 out of 111.

These numbers are significant, considering that: (1) The comparisons were made on a large and diverse set of data—111 different images as opposed to the typical small set of under twenty images. (2) The textures were all 'natural'; hence many were inhomogeneous and contained complex borderline 'non-textural'



Test19

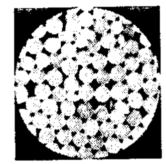
Humans: 2
Computer: 0
Hθ = $-5°$ (0.19), 85° (0.52)
Cθ: level0 (0.0051), level0 (0.0003)
Speculation: contrast

Test31

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

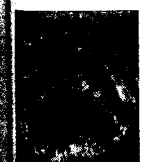
Test41

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



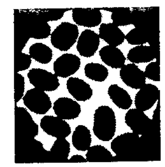
Test66

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



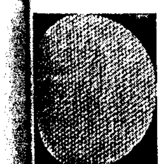
Test71

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



Test75

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



Test77

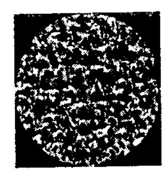
Humans: 2
Computer: 1
Hθ = $-65°$ (0.51), $-5°$ (0.445)
Cθ: level2 (0.0001), level3 (0.0452)
Speculation: filters

Test88

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

Test90

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



Test92

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

Figure 13. '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
$C\theta = 77°$, level=3 (1.5094)
$H\theta$: (0.10)
Speculation: contrast

Test42

Humans: 0
Computer: 1
$C\theta = 83°$, level=0 (0.0108)
$H\theta$: no match
Speculation: contrast

Test44

Humans: 0
Computer: 1
$C\theta = -45°$, level=0 (9.6)
$H\theta$: no match
Missing information in corner

Test98

Humans: 0
Computer: 1
$C\theta = -53°$, level=0 (0.0107)
$H\theta$: (0.02)
Speculation: contrast

Test107

Humans: 0
Computer: 1
$C\theta = -83°$, level=2 (0.2003)
$H\theta$: (0.09)
Speculation: contrast

Test108

Humans: 0
Computer: 1
$C\theta = 11°$, level=0 (0.0156)
$H\theta$: (0.09)
Speculation: contrast

Figure 14. 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.

patterns. These are much harder than synthetic textures which tend to be homogeneous. (3) The semantic meaning of the textures was not removed (for example by filtering), and yet it did not seem to cause much interference.

Rigorous conclusions are hard to make in an empirical study like this. Nonetheless, there are three conclusions that this study suggests: (1) Since the proposed computational algorithm performed closer to the humans when different thresholds were used at each scale, this suggests that maybe the human eye requires more 'saliency' from orientations at coarser levels before they are perceived to be dominant. (2) The semantic meaning of the textures in the Brodatz album does not interfere significantly with the 'low-level' processing in the proposed orientation detection algorithm. (3) Four levels of steerable pyramid with the proposed histogram analysis, and nonlinear contrast compensation and decision making, provide a reasonable first approximation to the human perception of detecting dominant orientations in natural textures.

The first conclusion could be tested by a follow-up study where the humans identify orientations at the various pyramid scales. Regarding the second conclusion, it is widely assumed that higher-level cognitive models influence low-level vision, but the size and nature of this influence is unknown. It is not far-fetched to consider 'dominant orientation' as a type of high-level description since non-vision researchers understand it with minimal explanation. Subsequently, the success of the reported low-level algorithm in matching the human's high-level orientation perception indicates that the semantic information recognizable in the patterns (e.g. 'this is lizard skin') did not significantly influence low-level orientation analysis. Only in a few cases (Section 5.7) did these influences play a suspect role in the failure of the steerable pyramid algorithm.

There is still room for improvement. Several areas for continuing research have been discussed, all in the context of a large optimization problem. An open question is 'how universal are the current thresholds—will they work on a still larger and more diverse set of real data'? There are fundamental constants in fields like physics; are there fundamental ones at work in visual perception? And finally, when a general orientation detector is found, one that works in a similar way to human orientation perception, how much of machine perception can be built upon it?

Acknowledgements

We would like to express our gratitude to O. Yip for his help developing and running the human experiment, to B. Horowitz for his X-window expertise, to E. H. Adelson and A. Bobick for helpful discussions during the design of the human test, to W. T. Freeman for the steerable filters code and many helpful insights, and to the reviewers for their helpful suggestions on this manuscript. This work was supported in part by BT (formerly British Telecom), UK.

NOTES

1. D45 was missing from the tape of digitized images we received.
2. For all the histograms shown in this paper, the angles along the horizontal axis are the angles at which the filters detect orientation changes. Thus the angles are all 90 deg from the actual structure seen at the orientation. For example, a vertical line would produce a strong peak in H at 0 deg.
3. The analysis of the synthetic teaser images will appear in a future report.
4. Since the reproduced images in this document are smaller and of poorer quality than those displayed to the human subjects during the test, not all the orientations clearly seen on the monitor can be clearly seen in the figures. The original images as well as data collected during this study are available for those interested in verifying our results.

REFERENCES

Aloimonos J. and Shulman, D. (1989). *Integration of Visual Modules: An Extension of the Marr Paradigm.* Academic Press, New York.

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

Bajcsy, R. (1973). Computer description of textured surfaces. *International Joint Conference on Artificial Intelligence*, pp. 572–578.

Bergen, J. R. and Adelson, E. H. (1988). Early vision and texture perception. *Nature* **333**, 363–364.

Bergen, J. R. and Landy, L. S. (1991). Computational modeling of visual texture segregation. In *Computational Models of Visual Processing.* M. S. Landy and J. A. Movshon (Eds). MIT Press, Cambridge, M.A. pp. 253–271.

Bigün, J. and du Buf, H. J. M. (1992). Geometric image primitives by complex moments in Gabor space and application to texture segmentation. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*

Bigün, J. and Granlund, G. H. (1987). Optimal orientation detection of linear symmetry. In *Proceedings of the 1st International Conference on Computer Vision,* London, England, pp. 433–438.

Brodatz, P. (1966). *Textures: A Photographic Album for Artists and Designers.* Dover, New York.

Chaudhuri, S., Nguyen, H., Rangayyan, R. M., Walsh, S. and Frank, C. B. (1987). A Fourier domain directional filtering method for analysis of collagen alignment in ligaments. *IEEE Trans. Biomed. Eng.* **34,** 509–517.

Choe, Y. and Kashyap, R. L. (1991). 3-D shape from a shaded and textural surface image. *IEEE Trans. PAMI* **13**, 907–919.

Cohen, H. A. and You, J. (1992). A multi-scale texture classifier based on multi-resolution 'tuned' mask. *Pattern Recognition Lett.* **13,** 599–604.
Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Trans. PAMI* **13,** 891–906.
Gorkani, M. (1993). Designing an orientation finding algorithm based on human visual data. Master's thesis, M.I.T. Perceptual Computing TR **#222,** MIT Media Laboratory.
Graham, N., Beck, J. and Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Res.* **32,** 719–743.
Graham, N., Sutter, A. and Venkatesan, C. (1993). Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation. *Vision Res.* **33,** 1893–1911.
Hamey, L. G. C. (1992). On human perception of regular repetitive textures. Technical Report **92-0094**C, Macquarie University, NSW, Australia.
Haralick, R. (1979). Statistical and structural approaches to texture. *Proc. IEEE* **67,** 786–804.
Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In: *Computational Models of Visual Processing.* M. S. Landy and J. A. Movshon (Eds), MIT Press, Cambridge, MA. pp. 119–133.
Hubel, H. D. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195,** 215–243.
Jain, A. K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition J.* **24,** 1167–1186.
Julesz, B. (1991). Early vision and focal attention. *Rev. Mod. Physics* **63,** 735–772.
Kass, M. and Witkin, A. (1987). Analyzing oriented patterns. In: *Readings in Computer Vision.* M. A. Fischler and O. Firschein (Eds), Morgan Kaufmann, Los Altos, CA, pp. 268–276.
Malik, J. and Perona, P. (1990). Preattentive texture discrimination with early visual mechanisms. *J. Opt. Soc. Am.* **7,** 923–932.
Perona, P. (1991). Deformal kernels for early vision. Technical Report MIT-LIDS-P-2039, MIT Laboratory for Information and Decision Systems, MIT, Cambridge, MA.
Phillips, G. C. and Wilson, H. R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *J. Opt. Soc. Am. A* **1,** 226–232.
Picard, R. W. and Kabir, T. (1993). Finding similar patterns in large image databases. In: *IEEE Proc. ICASSP,* Minneapolis, MN. **5,** 161–164.
Rao, A. R. and Lohse, J. (1992). Identifying high level features of texture perception. Computer Science RC17629 #77673, IBM.
Rao, R. and Schunck, B. G. (1991). Computing oriented texture fields. *CVGIP Graphical Models Image Processing* **53,** 157–185.
Rosenfeld, A. and Kak, A. C. (1982). *Digital Picture Processing,* Vol. 2. Academic Press, Orlando, FL.
Shapley, R. (1990). Retinal regulations of visual contrast. *Optics Photonics* **2,** 17–23.
Shepard, R. N. and Cooper, L. A. (1982). *Mental Images and their Transformations.* MIT Press, Cambridge, MA.
Simoncelli, E. P., Freeman, W. T., Adelson, E. H. and Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Trans. Information Theory* **38,** 587–607.
Tamura, H., Mori, S. and Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Trans. SMC* **8,** 460–473.
Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychol.* **12,** 97–136.
Webster, M. A. and De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Am. A* **2,** 1124–1132.
Wolfe, J. M., Cave, K. R. and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol.: Human Percept. Performance* **15,** 419–433.
Wright, G. A. and Jernigan, M. E. (1986). Texture discriminants from spatial frequency channels. In *IEEE International Conference SMC,* **1,** pp. 519–524.
Young, R. A. (1986). The Gaussian derivative model for machine vision: Visual cortex simulation. Technical Report GMR-5323, General Motors Research Laboratories.
Ziemer, R. E. and Tranter, W. H. (1990). *Principles of Communications: Systems, Modulation and Noise.* Houghton Mifflin, Boston, MA.
Zucker, S. W. (1985). Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision Graphics Image Processing* **32,** 74–103.

APPENDIX

USING THE STEERABLE PYRAMID TO ESTIMATE LOCAL ORIENTATION

To find the main orientation in a local neighborhood, an oriented energy $E(\theta)$ is calculated for image I at each level of the pyramid where $E(\theta)$ is defined as:

$$E(\theta) = (f^\theta * I)^2 + (h^\theta * I)^2, \qquad \text{(A1)}$$

where f^θ is the directional filter for the steerable pyramid and h^θ is its approximate Hilbert transform. Since f^θ and h^θ can be expressed as a linear combination of their basis functions, $E(\theta)$ can be expressed in the following way:

$$\begin{aligned} E(\theta) &= \left(\left(\sum_{l=0}^{3} k_{fl}(\theta) f_l\right) * I\right)^2 + \left(\left(\sum_{n=0}^{4} k_{hn}(\theta) h_n\right) * I\right)^2 \\ &= \left(\sum_{l=0}^{3} k_{fl}(\theta) f_l * I\right)^2 + \left(\sum_{n=0}^{4} k_{hn}(\theta) h_n * I\right)^2, \end{aligned} \qquad \text{(A2)}$$

where k_{fl}, $0 \le l \le 3$ are the interpolation functions to steer f_l, the basis filters for f, and k_{hl}, $0 \le n \le 4$ are the interpolation functions to steer h_n, the basis filters for h. As can be seen in Eqn (A2) to calculate $E(\theta)$, image I at each level of the pyramid just has to be convolved with the basis filters for f and h and the interpolation functions for a particular θ can be used to give the oriented energy.

The dominant orientation θ_0 can be found by maximizing $E(\theta)$. The solution for θ_0 was found by Freeman and Adelson (1991) to be:

$$\theta_0 = \frac{\arg(C_3, C_2)}{2}, \qquad \text{(A3)}$$

where C_2 and C_3 are combinations of the basis filter outputs for f and h. The strength of the orientation estimation S is defined as:

$$S = \sqrt{C_2^2 + C_3^2}. \qquad \text{(A4)}$$

The approximation stated in Eqn (A3) 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 this approximation is not correct and other methods have to be used to find the orientations. These issues are reviewed in Gorkani (1993). Throughout this paper the above method is used.