

Energy processing and coding factors in texture discrimination and image processing

TERRY CAELLI

University of Alberta, Edmonton, Alberta, Canada

In this paper two specific questions are considered: (1) When image energy is defined in terms of orientation and frequency components of the image spectrum, how much energy is detected as a function of contrast? (2) Does the visual system have a limited resolution for image orientation and frequency components which define the psychophysical upper limits for two-dimensional image coding units? To examine these two questions, experiments were conducted with both black/white and gray-scaled images. The results indicate a monotonic relationship between image contrast and energy processed. Finally, results demonstrate that the upper limits (or resolution) for discriminable spectral orientation and spatial frequency are approximately 5 deg and 1/8 octave, respectively. Image domain demonstrations confirm these results.

This paper is concerned with the question of what information is detected in a briefly exposed image. Clearly, the question can be quantitatively answered only when specific image dimensions are considered—dimensions which have known selectivities within the visual system both at neurophysiological and psychophysical levels. The image dimensions—spatial frequency and orientation—were chosen for this study for two main reasons: (1) It is now clear from a number of studies that one can *define* the response characteristics of individual cortical units in terms of two-dimensional band-limited filters (Andrews & Pollen, 1979; Daugman, 1980; DeValois & DeValois, 1980; Rose, 1979). (2) A large number of psychophysical studies with individual grating and band-limited signals suggest that there exist channels within the visual system that are selectively sensitive to both spatial frequency and orientation components of an image (Julesz & Schumer, 1981). However, to this stage, the precise number, bandwidths, and energy processing characteristics of such coding units have not been fully investigated.

Such representations for receptive field properties and psychophysical channels do not imply that the visual system actually performs a Fourier analysis of the image. Rather, such spectral definitions of image coding units provides a way of investigating how such receptive and “perceptive” fields may operate, as a linear system, in processing general image properties. Here the term “perceptive” field is used to represent (after Jung & Spillman, 1970) the image domain profiles of two-dimensional (band-limited)

channels determined by psychophysical methods (Mostafavi & Sakrison, 1976) to function independently.

One important consequence of viewing image coding in these terms is that we can investigate the amounts of orientation and spatial frequency information¹ detected in an image as a function of different display conditions. Of specific interest in our first experiment is the issue of orientation and frequency processing as a function of image contrast, since it is clear from single one-dimensional grating studies (e.g., Georgeson & Sullivan, 1975) that (absolute) threshold frequency sensitivity measures do not necessarily reflect suprathreshold responses. It is not known to what extent we can predict the sensitivity to two-dimensional image frequency components in terms of the threshold-derived contrast sensitivity function, or in terms of what energies are present in an image of each frequency band. Equally, it is not known whether or not our sensitivity to image-orientation-specific information is equal to that of spatial (radial) frequency. Again, the benefit of observing information processing with respect to spectral representations of the image is that we can compare directly both image dimensions with respect to energy detection characteristics.

Experiment 1 is concerned with these latter issues. Here we have used an adaptation of the original Julesz (1962) texture-discrimination paradigm in which an original image and an energy-reduced (filtered) version are briefly exposed simultaneously and we determine the contrast level *below* which the observer *cannot* discriminate between them. In this way, we can determine the relationship(s) between image energy detection and contrast as a function of different regions of the image-frequency spectrum with respect to selective orientation and radial frequency (polar coordinates of the spectrum) dimensions.

This research was funded by a grant from the Australian Research Grants Committee. The author acknowledges the help of Kim Harrison and Paul Bevan. The author's mailing address is: Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.

EXPERIMENT 1 Orientation and Frequency-Specific Energy Processing

Method

Stimuli and Apparatus. Due to the large number of digital images required, image resolution was restricted to 128×128 pixels, with each pixel (image element) having 4-bit intensity resolution (16 intensity levels). Each image was displayed in a circular 2-deg visual-angle aperture, although all filtering was enacted on the original 128×128 square matrix. Although more than 16 gray levels would be ideal (see Caelli & Hübner, Note 1), our discrimination paradigm at least guaranteed equal restrictions on intensity quantization for both images.

Figure 1 shows the two image types employed: a scene and random texture. The textures (of which sections are shown in Figure 2) were generated as random contrasts having different element sizes ($T_2=2 \times 2$, $T_4=4 \times 4$, $T_8=8 \times 8$ pixels), which determined "granularity" differences. Fourier log power [$\log(\text{Amp}^2)$] spectra plots were determined for both radial frequency and orientation distributions (Figure 1). Here the radial frequency values were averaged over orientations and the orientation values were averaged over radial frequencies.

For the purpose of probing different radial frequency ranges, three disjoint regions corresponding to low ($f \leq 2$ cycles/deg: cpd—L), middle ($2 < f \leq 8$ cpd, center at 4 cpd—M), and high ($8 < f \leq 32$ cpd, center at 16 cpd—H) were considered. With the orientation analyses, each annulus (Figure 2) was divided into 18 (for low), 36 (for middle), and 72 (for high) orientation-specific regions of 20 deg (± 10 deg about each equally spaced center), 10 deg (± 5 deg), and 5 deg (± 2.5 deg) widths, respectively. This increase in the number of orientation regions, from the low- to high-frequency bands, was used to keep the equivalent two-dimensional (orientation \times radial frequency) areas comparable. However, as will be seen from the results of this experiment, it is the percentage energy reduction ratio, rather than the specific type of reduction, which is critical for discrimination.

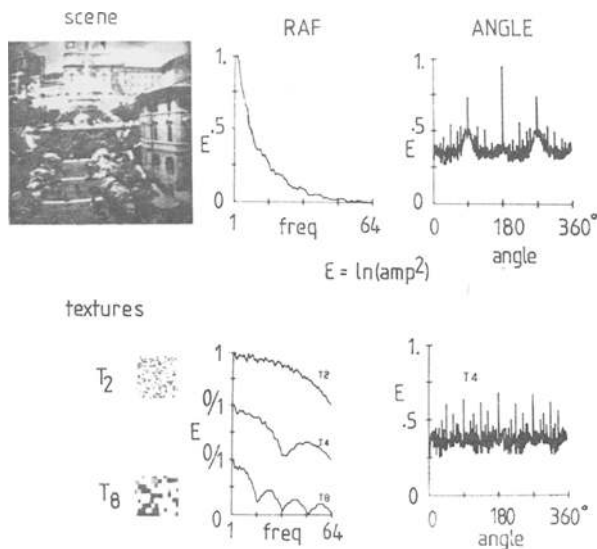


Figure 1. Scene and textured images used in Experiment 1. Relative log (energy) (E : 0-1) distributions are shown as a function of radial frequency (RAF) in picture cycles and orientation (angle: deg). T_i refers to a texture whose granularity is determined by a basic element of size $i \times i$ (see text).

The image spectra were then reduced in energy (bandpass filtered) along the (radial) frequency dimension in the following way. For each of the three regions defined above (L, M, H), the bandwidth (annulus size) was reduced (with respect to a logarithmic frequency measure: octave units) about the center frequency by 10%, 20%, 40%, 60%, 80%, and 90% of the original energy within that region of the spectrum (Figure 2a). It should be noted that each of the three regions contained different amounts of energy, as illustrated in Figure 1. Identical percentage energy reductions were made in each region (L, M, H) along the orientation dimension by simultaneously reducing the widths of the orientation regions (determined by the number of segments) about each unit's center (Figure 2b). Each region was filtered (or reduced in energy) while the other two bandwidths were left alone, such that the original and filtered image pairs differed only in a selective loss of information within the target region. Two examples are shown in Figure 3. Such orientation or (radial) frequency-specific filtering operations correspond to (rectangular) ideal bandpass filters, and so do not strictly correspond to physiologically determined profiles as discussed by Daugman (1980). They were employed, however, to simply investigate energy processing as a function of two-dimensional frequency bandwidths without employing specific tuning profiles. Such bandpass filtering is known to produce "ringing" (side-bands) in the image, which, in these studies, is presumed not to be a significant factor in determining discrimination thresholds.

These operations produced differing amounts of total energy reductions over the various images and filter types. However, an extra combined (L+M+H) case guaranteed total energy reductions of the above specified percentages (Figure 2). The images (3 textures + 1 scene) filtered by each of the above energy-reducing filters [6 amounts \times (3 frequency bands + 1 total) \times (radial frequency, orientation specific)] were inverse Fourier transformed to result in $48 \times 4 = 192$ filtered images. Each image was set at the same space average luminance (12 cd/m^2) as the original images.

Procedure. An experimental trial consisted of the presentation of an original image (scene or texture: center image in the plates shown in Figure 3) and a corresponding energy-reduced version. They were presented adjacent to each other (separated by .25 deg) and exposed for 120 msec at a given contrast level. In this case, contrast was defined as $[(I_{\max} - I_{\min}) / (I_{\max} + I_{\min})] \times 100$, and so determined by the extreme values of the 4-bit pixels and an analogue contrast modulator coupled to the display device (an Electrohome 17-in. TV monitor). The effect of contrast modulation on the 16 intensity levels is illustrated in Figure 4, and, again, it should be emphasized that both images were treated identically in this respect.

The subject's task was to indicate whether the two images were identical (with respect to "quality, clarity, and fidelity") or not. Contrast was increased and decreased according to a random staircase technique, and a contrast threshold for discrimination was determined when the response ("same"/"different") alternated three times in 1% steps about a limiting contrast value. This procedure was repeated three times for each of the total 192 image pairs. Original and filtered images were allocated randomly to the left and right display positions. The subject was instructed to fixate between the two images before each trial and to not attend to any one area of the image.

Three subjects, who had emmetropic vision with no known history of visual defects, were paid as observers. Experimental periods were restricted to 1 h in duration, the total experiment being conducted over a 1 month period.

Results and Discussion

In order to convert the "contrast for discrimination" into a direct discriminability index, the measure DS (discrimination sensitivity) was used to index responses, where $DS = 1/\text{contrast for discrimination}$.

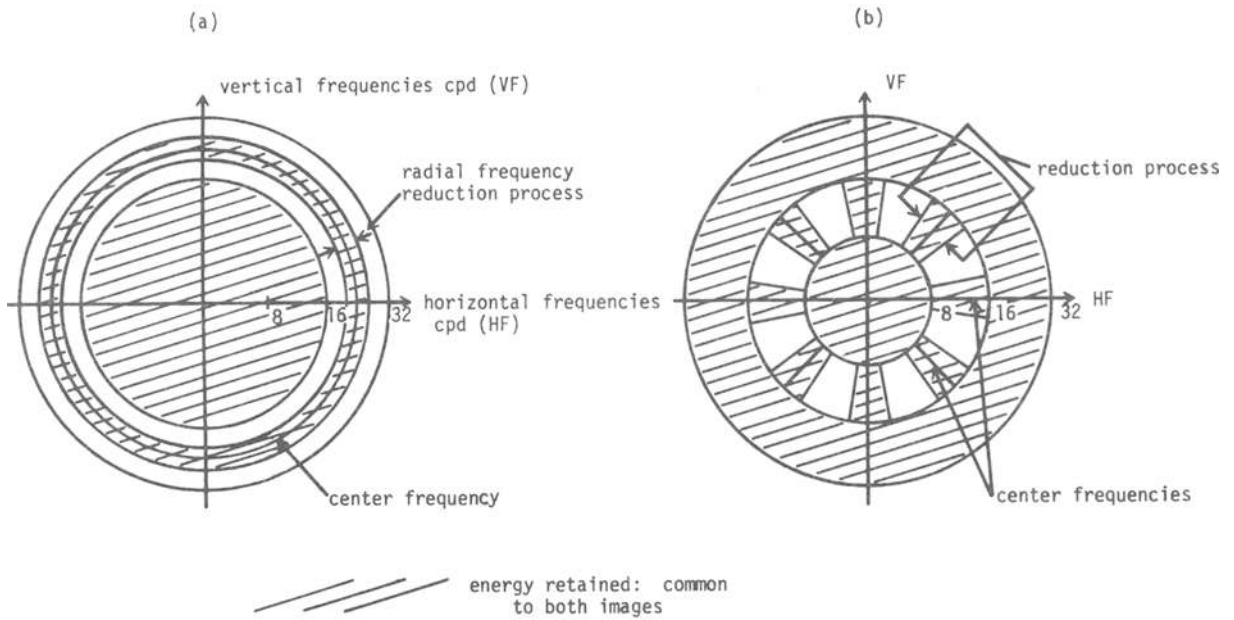


Figure 2. Method used to selectively delete frequency components in the images shown in Figure 2: (a) the reduction process along the radial frequency dimension; (b) the reductions along the orientation code.

Figure 5 shows the relationship between DS (averaged over trials and observers) and the energy reduction values (ER in percentage) with the scene and textured images. It is important to note that, although fixed percentages of energy reductions were used *within* each bandwidth (10% to 90%), the total reduction over the full spectrum (as with the images) varied with the proportionate amounts of energy contained within each bandwidth. That is, total energy reductions of precisely 10% to 90% occurred only in the combined (L + M + H) case, as illustrated in Figure 5. For this reason, a three-way repeated-measures analysis of variance was computed [4(image type) × 2(type of reduction: radial or orientation) × 24(energy reduction values) × subjects (3)]. Re-

sults yielded a significant effect only on the energy-reduction factor [$F(23,46) = 4.71, p < .01$]. It should be noted that observers could not discriminate the low energy-reduction values within the low-frequency band, and these three points are not represented in Figure 5, although, for purposes of the above analysis, they were set to a DS value of 0.01.

The main result of this experiment is that sensitivity to image-energy reduction appears to be independent of frequency bandwidth, the type of reduction strategy (orientation or radial frequency), and pattern type. In summary, Figure 5e shows the relationship between the overall image contrast value and the amount of energy that may be reduced without the observer's detecting it.

Although this experiment does not point to a specific class of coding mechanisms for two-dimensional image processing, it does demonstrate that in image perception we do not process all the frequency components that exist in the stimulus. We have proved this indirectly by determining the amount (and type) of spatial frequency information that can be deleted before discrimination occurs. This result is consistent with the suggestion that visual perceptive fields are delimited (in their linear response characteristics) by two-dimensional band-limited energy-processing units whose bandwidths are determined by their retinotopic properties (Kulikowski, Marcelja, & Bishop, 1982; Daugman, Note 2). This would explain why reducing energy in a *distributed* fashion over the spectrum would produce images nondiscriminable from the original scene, since all appropriate detector

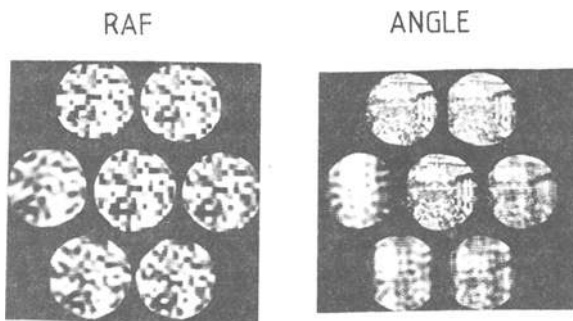


Figure 3. Original center texture and scene images and (clockwise from top left) energy reductions of 10%, 20%, 40%, 60%, 80%, and 90% over all three (low, middle, high) spectral regions for (a) radial frequency (RAF) and (b) orientation (Angle) dimensions.

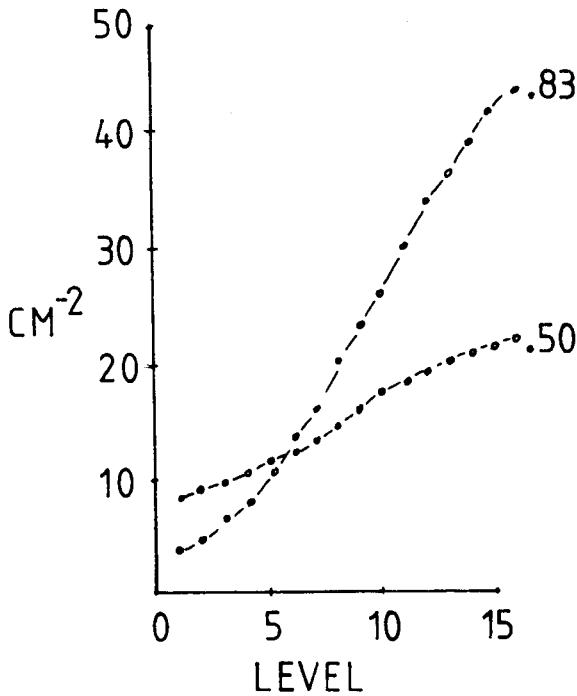


Figure 4. The relationships between 16 gray levels (4-bit pixel size) and luminances (cm^{-2}) at two contrast levels ($l_{\max} - l_{\min}$) / ($l_{\max} + l_{\min}$): .83, .50.

units could still be functional. Such units, by definition, would not be sensitive to individual frequency components of the image but rather to a range of frequencies, all of which could adequately stimulate a given detector.

Consider the images in Figure 6. The top row shows two original 256×256 4-bit pixel images, with the texture-amplitude spectrum shown on the upper right. The frequency components going down the rows have been transformed in the following way. Each row defines a uniform set of two-dimensional spectral regions varying from $1/4$ octave \times 10 deg wide to 2 octaves and 10-deg-wide units. Within each unit, the average amplitude of all the composite frequencies has been assigned to each member. Upon calculating the inverse Fourier transforms, it is obvious that no image degrading occurs if we restrict such units to the $1/4$ octave and 10° wide, while broader values induce perceivable distortions at this high contrast range.

Such low-frequency resolution limits are also consistent with the results of a recent increment threshold experiment conducted by us (Caelli, Brettel, Rentschler, & Hilz, 1983). In this experiment, the ability of observers to discriminate gratings in terms of both their spatial frequency and their orientations was determined. Taken together, these data suggest that the type of two-dimensional spatial frequency information which is representative of what can be

perceived in an image is certainly not defined by individual frequency components. Rather, the upper limits for energy coding seem to be limited to the sets of narrow-band two-dimensional spectral regions defined above. It is these two-dimensional areas which similarly delimit energy processing in "perceptive field" profiles, although their precise spectral tuning characteristics are not demonstrated here (see Daugman, 1980).

The results of Experiment 1 indicate that such energy-coding units would have threshold response characteristics determined at least partially by the contrast of the full image. This seems reasonable, since an increase in overall contrast increases the total energy of the stimulus. What we do not know is whether these limited energy-coding characteristics occur under "maximum" contrast conditions. In the following experiment, this question was investigated

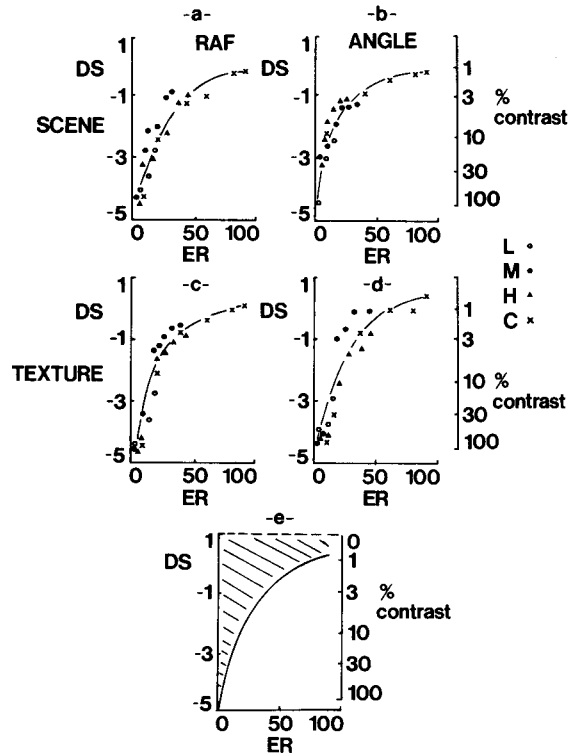


Figure 5. Discrimination sensitivity (DS: left ordinate in log units) and contrast for discrimination (right ordinate) for scene and texture images as a function of total image energy reductions (ER: percentage). H, M, H, C represent the spectrum region of energy reduction (low, middle, high, combined). Left-side figures a and c refer to radial frequency (RAF) and right-side figures b and d refer to orientation (angle) specific energy reductions. The lines in a-d define the consistent trend. In e are shown averages over all conditions and the relationships between contrast and energy reductions not detected. All points represent averages over observers and observations. Lined area corresponds to regions where energy reductions can be made without detection below the specified contrast level.

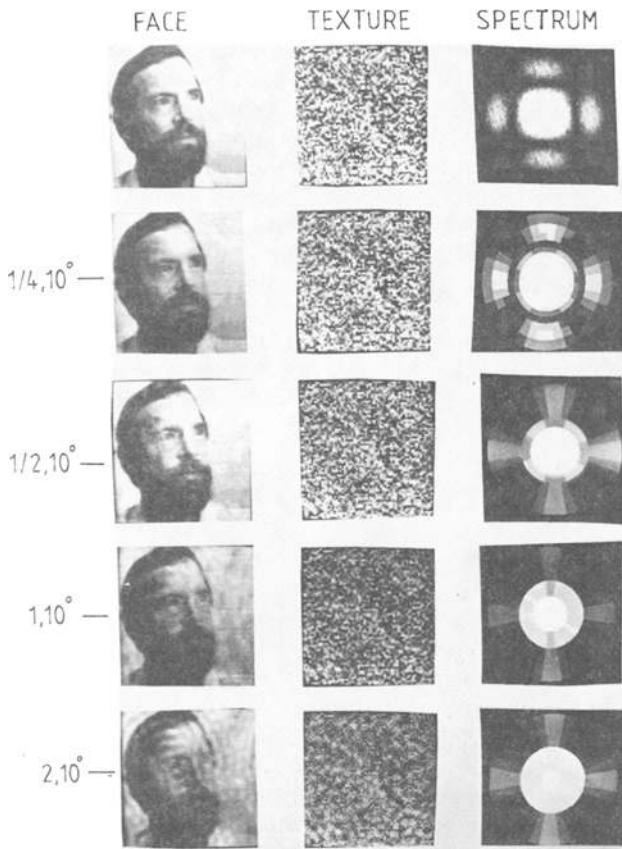


Figure 6. Top row shows two original images digitized as 256×256 , 4-bit pixels and the amplitude spectrum for the texture. Rows 2 through 5 represent the inverse Fourier transforms of the amplitude-averaging process illustrated for the texture case in the right-hand column. The unit radial frequency widths (octaves: RF) and orientation tuning (deg:°) are shown on the left (RF, deg).

with respect to the limits for orientation processing—using black/white textures.

EXPERIMENT 2 Orientation Coding in Black/White Textures

The number of permissible orientations about a point varies as a function of length on a discrete grid. Specifically, for lines of n units radial length embedded on a $(2n + 1) \times (2n + 1)$ submatrix, there are $8n$ orientations. In the following experiment the aim was to explore whether the perceptual orientation resolution limit was correlated with this physical limitation or whether it was somewhat below, consistent with the observations with contrast-modulated images.

Method

Stimuli and Apparatus. Texture pairs, as illustrated in Figure 7, were generated for three different line lengths (subtending .1, .2, and .3 deg of visual angle). For each length, three different rectangular distributions of orientations were employed: 180, 90,

and 40 deg. As mentioned above, the number of possible orientation states on an $n \times n$ grid is a function of the line length. So, although the theoretical orientation distributions for the 45 lines was 4, 2, and 0.89 deg, respectively, this did not occur with all line lengths. Specifically, the minimum orientation resolutions for the three line lengths were (.1:9 deg, .2:4.5 deg, .3:2.25 deg). Consequently, the rectangular distributions of orientations over 180, 90, and 40 deg were subject to these constraints. For example, with .1-deg length lines, the orientation resolution is restricted to 9 deg, etc. (Figure 7). For each orientation and length condition, the orientation resolution of one texture was decreased by the minimum resolvable step size defined above (as varying with the line length). For example, with the .3-deg length, the "original" resolution was 2.25 deg. Hence, all changes in resolution were restricted to multiples of this value. This limitation of orientation states implies that our estimates of orientation resolution with the small-length element (.1 deg) are, by definition, coarser than those of the longer lines.

Each texture pair was displayed on a HP 1310A CRT under control of a PDP-11/40 computer interfaced with a HP 1350A graphics translator with a $1,000 \times 1,000$ point grid structure. The two textures subtended a visual angle of 4 deg from a viewing distance of 1 m, being composed of two 5×9 micropatterns (Figure 7).

Procedure. Three naive observers, without any known visual defects, were used. An experimental trial consisted of the presentation of a texture pair for 120 msec, after which the observers were to indicate whether they had perceived one or two textures—whether the right half appeared different from the left (Figure 7a). Original textures and "orientation-resolved" versions were allocated randomly to the left and right positions over trials. A random staircase procedure was used to ascertain the orientation resolution threshold (discrimination resolution, DR) for each condition—orientation distribution (180, 90, and 40 deg) and line length. Such threshold estimates were repeated three times per subject per condition.

Results

Figure 8 shows mean resolution for discrimination as a function of the orientation ranges and lengths used.

A two-way analysis of variance (angle \times length \times subjects) yielded significant lower orientation resolutions as a function of both angle (the distributions of 180, 90, and 40 deg) and length parameters [angle, $F(2,16) = 21.19$, $p < .05$; length, $F(2,16) = 20.52$, $p < .05$] and a significant interaction effect [$F(4,32) = 9.15$, $p < .05$].

Although these results are interpretable across orientation ranges (40, 90, and 180 deg), they are not

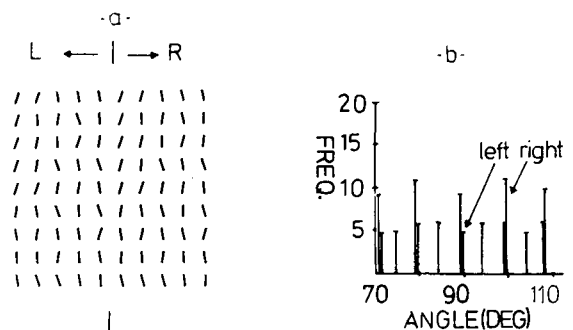


Figure 7. (a) Two textures used in Experiment 2 and (b) their associated orientation distributions.

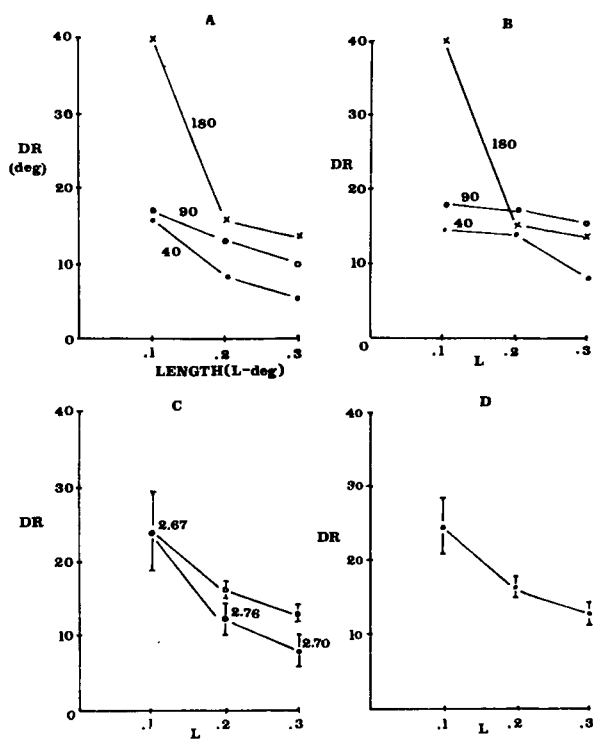


Figure 8. (a) Orientation resolution for discrimination (DR) as a function of initial orientation distribution width and line length. (b) Comparison case when initial resolution for all sizes is set at 10 deg. (c) Orientation resolution thresholds averaged over distributions [●, average of (a); ○, average of (b)]. (d) Average over all conditions. Note the constant (approx. $e = 2.72$) ratio of threshold to maximum resolution permitted.

directly comparable across line lengths due to the basic orientation resolutions of the original textures, as discussed above. Consequently, a further replication of the experiment was performed with a constant 10-deg resolution of the original textures—independent of length. The orientation resolution thresholds were estimated by the identical procedure as before. Results are shown in Figure 8b, and averages over angles are shown in Figure 8c. In this case, only angle and length factors were found to be significant [angle, $F(2,16) = 5.07$, $p < .05$; length, $F(2,16) = 10.60$, $p < .05$].

What is interesting here is that the ratio of orientation resolution (for discrimination) to the limiting physical orientation resolution (determined by the lengths) turns out to be a constant and approximately $e(2.72)$, suggesting that Weber's law may well apply. This was not so for the fixed 10-deg resolution condition, since, by definition, the orientation resolution of the original textures was fixed above the limiting physical values.

Taking these factors into account, the results demonstrate that: (1) the perceived orientation resolution (or perceived number of orientation "states") is less

than what is physically present in the textures and possibly defined by a constant multiple of the physical resolution (e); (2) the best resolution for orientation was no better than ± 5 deg (occurring with the largest line); and (3) as line length increases, or grid resolution increases, orientation resolution increases only to the constant factor defined in (1).

In terms of the previous discussion (see p. 000), the most parsimonious orientation coding properties with these results would entail a set of detectors whose tuning widths are functionally limited to ± 5 deg. This limit, ascertained under asymptotic contrast conditions, with black/white textures (contrast of 100%) is comparable with those recently documented in Caelli and Hübner (1983) with gray-scale textures and scenes.

GENERAL DISCUSSION

The experiments reported here were aimed at answering two questions which arise naturally from recent findings about coding mechanisms in the visual system. Is there a systematic relationship between image contrast and energy processed? Are there a finite number of energy coding states detected and discriminated in briefly displayed spatially distributed information? Although, in one sense, these questions have been indirectly asked and investigated with one-dimensional grating stimuli and single oriented line segments, there has not been, to this date, a direct study with genuinely two-dimensional images. In this way, the results of these experiments may only be regarded as confirmatory of already existing conjectures.

However, the perspective of this project is to develop and demonstrate the finite energy processing characteristics of human observers which are amenable to modeling in terms of linear systems via two-dimensional filtering procedures.

These experiments have been concerned with the analysis of radial frequency and orientation processing characteristics of digital images. Experiment 1 indicates the relationship between energy detected and contrast—irrespective of whether the energy is orientation or radial frequency specific. These data equally suggest that this paradigm is appropriate for studying energy processing but not the fine issues concerned with two-dimensional coding strategies.

However, combining the results of these investigations, we conclude that orientation and radial frequency coding are limited to ± 5 deg orientation and $\pm 1/8$ octave widths, respectively, with distributed spatial information (as in natural scenes). The results of Experiment 1 indicate that as image contrast is lowered, the response of such units also decreases to result in a loss in the detection of specific orientation and radial frequency energies.

A number of recent studies of the spatial frequency response range of individual cells in the cat and monkey (area 17) indicate quite wide tuning of above 1 octave range (DeValois & DeValois, 1980, p. 322). Equally, half-amplitude response ranges of such units for orientation are often wider than ± 5 deg (Daugman, 1980). These data are not to be confused with the coding investigation of interest here. Whether it be due to the cross-correlation of many widely tuned units, or to the existence of specifically different narrow-band units in the visual system, we know from these psychophysical studies that size and orientation sensitivity is considerably finer than previously described broadly tuned channels.

REFERENCE NOTES

1. Caelli, T. M., & Hübner, M. *On the number of intensity levels detected in textures*. Manuscript submitted for publication, 1983.
2. Daugman, J. *Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters*. Unpublished manuscript, 1981.

REFERENCES

- ANDREWS, B., & POLLEN, D. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology* (London), 1979, **284**, 163-179.
- CAELLI, T. M. *Visual perception: Theory and practice*. Oxford: Pergamon Press, 1981.
- CAELLI, T. M., BRETTEL, H., RENTSCHLER, I., & HILZ, R. Discrimination thresholds in the two-dimensional spatial frequency domain. *Vision Research*, 1983, **23**, 129-133.
- CAELLI, T. M., & HÜBNER, M. On the efficient two-dimensional energy coding characteristics of spatial vision. *Vision Research*, 1983, in press.
- DAUGMAN, J. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 1980, **20**, 847-856.
- DEVALOIS, R., & DEVALOIS, K. Spatial vision. *Annual Reviews of Psychology*, 1980, **31**, 309-341.
- GEORGEON, M., & SULLIVAN, G. Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology* (London), 1975, **252**, 627-656.
- JULESZ, B. Visual pattern discrimination. *IRE Transaction on Information Theory IT-8*, 1962, 84-92.
- JULESZ, B., & SCHUMER, S. Early visual perception. *Annual Review of Psychology*, 1981, **32**, 575-627.
- JUNG, R., & SPILLMAN, L. Receptive-field estimation and perceptual integration in human vision. In F. A. Young & D. B. Lindsley (Eds.), *Early experience and visual information processing in perceptual and reading disorders*. Washington, D.C.: National Academy of Sciences, 1970.
- KULIKOWSKI, J. J., MARCELJA, S., & BISHOP, P. O. Theory of spatial position and spatial frequency resolutions in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics*, 1982, **43**, 187-198.
- MOSTAFAVI, H., & SAKRISON, D. Structure and properties of a single channel in the human visual system. *Vision Research*, 1976, **16**, 957-968.
- ROSE, D. Mechanisms underlying the receptive field properties of neurons in cat visual cortex. *Vision Research*, 1979, **19**, 533-544.

NOTE

1. Here the two-dimensional spatial frequency domain is represented in polar coordinate form. That is, vertical and horizontal frequencies are replaced by orientation and radial frequency determinants. The benefit of specifying image orientation information in the frequency domain is that it is translation invariant (Caelli, 1981).

(Manuscript received April 12, 1982;
revision accepted for publication May 20, 1983.)