Investigative Ophthalmology & Visual Science, Vol. 32, No. 8, July 1991

Image Enhancement for the Visually Impaired

Simulations and Experimental Results

Eli Peli,*†‡ Robert B. Goldstein,* George M. Young,* Clement L. Trempe,*†§ and Sheldon M. Buzney*†§

Digital image enhancement has been proposed as an aid for the visually impaired. The capability of two enhancement techniques to improve recognition of images by patients with central scotoma or cataracts was evaluated using image-processing simulations and direct patient testing. Enhancements and simulations were based on measurements of contrast sensitivity loss for patients with macular disease. Contrast sensitivity loss was measured using Gabor-type localized stimuli and paradigms that are appropriate for analyzing form perception. The simulations using the contrast sensitivity data suggested that patients with moderate visual loss (20/70–20/200) may have difficulty recognizing faces and may benefit from enhancement by both of the techniques used. Ability to recognize celebrities from enhanced images improved for 39 of the 46 patients tested. The improvement was significant ($P < 0.05$) for 16 of the 38 patients with central visual loss and for 3 of 8 patients with anterior segment media opacities tested. The simulations suggest that the benefits of image enhancement may be similar or even greater for recognition of other types of images. Invest Ophthalmol Vis Sci 32:2337–2350, 1991

Low-vision patients have difficulties not only reading small print but also deciphering continuous-tone images of natural scenes or photographs. High-contrast photographs have been used to facilitate access of patients to such images.[1,2] Digital image processing was proposed as an aid for the visually impaired.[3] Image enhancement may be used to improve visibility of printed pictures and video images for these patients. For instance, television programs can be enhanced either at a central broadcasting location or at the patient's receiver. The same technology could be used to enhance the images presented on the patient's closed-circuit television magnifying system. It may be possible to develop a portable system with a head-mounted, closed-circuit television system to aid mobility.[4]

A conceptual preemphasis model of image enhancement for the visually impaired was proposed by Peli et al.[5] The model suggests that, if the transfer characteristics of the visual system of a low-vision patient can be measured, images then can be preprocessed before presentation to the patient to compensate for the degradation caused by the patient's visual impairment.

The potential value of image enhancement was evaluated previously using simulations of effects of enhancement with normal observers.[5–7] Photographs of enhanced images taken using optically simulated cataracts[5] appear to provide detail needed for recognition. In testing normal observers, images that were enhanced using local histogram equalization reduced the contrast required for face discrimination by a full octave.[6] Others[7] showed that increasing the contrast of low-pass-filtered text increased the reading rate for normal observers.

Enhancement of text using filters based on each patient's contrast sensitivity function (CSF) modestly reduced magnification demands for reading for three patients with central scotoma[8] and increased the reading rate substantially for the same three patients.[9]

Although these results show considerable promise, the value of image enhancement in improving recognition of gray-scale images by visually impaired patients has not been demonstrated. We wanted to determine if enhancing images of human faces improved recognition of facial features. In using faces, we restricted the infinitely diverse range of possibilities to a class of images for which a large body of knowledge is available.[10] Moreover, difficulty with face recognition is a frequent, early complaint of many patients with macular disease,[11] and older normal observers reportedly have had difficulties recognizing faces under low-contrast conditions.[12]

Faces may be recognized even when blurred consid-

From the *Eye Research Institute, †Department of Ophthalmology of Harvard Medical School, ‡Department of Ophthalmology of Tufts University School of Medicine, and §Retina Associates, Boston, Massachusetts.

Supported in part by grant R01-EY05957 (EP) from the National Institutes of Health, Bethesda, Maryland, and a grant from the Teubert Charitable Trust Foundation, Huntington, West Virginia.

Submitted for publication: June 25, 1990; accepted February 20, 1991.

Reprint requests: Eli Peli, Eye Research Institute, 20 Staniford Street, Boston, MA 02114.

erably.[13] A good level of discrimination of face images from a restricted sample of test faces is possible even when the images are low-pass filtered to such an extent that only head shape, neck-and-shoulder geometry, and gross hairline are all that remain.[14] This, however, does not represent the level of performance in face recognition tasks required for everyday social conditions. Ginsburg[15] claimed that low spatial-frequency information is sufficient for face identification. Others[16] found that high-frequency information is not redundant but is sufficient by itself to enable face discrimination. Although low frequencies convey most of the relevant information for processing images of faces, high-frequency information is not redundant.[17] High frequencies seem to benefit performance in tasks that require accessing the identity of a face compared with discriminating among a small sample of face images. What constitutes high or low frequencies in terms of face recognition? For faces spanning 3–4°, information above 15[16] or 16 cycles/face[18] was undetectable by normal observers and, thus, did not aid in face recognition. However, spatial information above 5[16] or 4 cycles/face[18] considerably improved recognition. The critical band width for recognizing faces with varying facial expressions was reported to be 6 cycles/face even for a small test group of ten faces.[19] Thus, spatial information above 4 cycles/face and up to 15 cycles/face appears to be important for recognition.

Originally, we proposed a linear preemphasis model,[4,5] adapted by Lawton[9] for the enhancement of text. In this model, the reduction in contrast sensitivity at all frequencies is used to determine the level of the required enhancement. Using all frequencies, however, is unnecessary and also may be disadvantageous for patients with central scotoma. Suprathreshold features at lower spatial frequencies in the image do not need to be enhanced because they are perceived at their correct contrast due to contrast constancy.[20] Such constancy was shown to hold for parafoveal and peripheral retinal eccentricities.[21] Enhancing such features will, therefore, result in a distorted appearance and, at the same time, reduce the display range available for enhancement of higher frequency subthreshold features. Measurements of patients' CSF are, nevertheless, required to determine the range of frequencies that need to be enhanced. These are frequencies that can be seen by the patient but only at contrasts higher than are available in the face image.

We evaluated digital image enhancement for face recognition in two steps. First, the effect of enhancement was assessed using linear and nonlinear simulations of the appearance of the enhanced and unenhanced images. Linear processing was used to simulate images presented to patients with cataracts, and nonlinear processing was required to simulate the appearance of the same images presented to patients with central scotoma. These nonlinear simulations were required to predetermine whether the enhancement parameters estimated from patients' CSF measurements would improve the appearance of face images for the observers sufficiently. Second, we determined the degree of improvement in recognizing enhanced face images by visually impaired patients. Patients with central scotoma due to macular disease and another group of low-vision patients with various anterior segment disorders were tested. To evaluate actual face recognition rather than the ability to discriminate among a limited selection of test faces (a matching task), we tested the patient's ability to recognize celebrities. Recognition of familiar faces is a task that is more robust[13] than the recognition of recently memorized faces. Familiarity is an important factor in the observer's ability to make many of the natural face recognition transfers encountered every day, such as a change in head gear.[13] Two sources discuss the differences between recognition of familiar and unfamiliar faces.[22,23] Readers who are more interested in our results of enhancement testing with patients than in our methods used to tune the enhancement for this population can proceed to the section "Patients' Recognition of Enhanced Images of Faces."

Materials and Methods

Measurement of Patients' Visual Loss

Most previous measurements of spatial CSFs have been based on stimuli that give maximal contrast sensitivity.[24] However, as discussed subsequently, the use of such CSFs in simulating vision in normal observers[15] and low-vision patients[25,26] or in designing image-enhancement techniques as visual aids[5,9] may be inappropriate. We measured CSFs using stimuli that are better suited as bases for simulating and analyzing perception of images.[24]

A few features of the grating stimuli need to be adjusted for simulation purposes. Repeated, continuous cycles at any frequency are rarely seen in natural images. Only when periodic patterns (such as a picket fence) appear in the image do we find more than two consecutive, high-contrast cycles in such images. Thus, the sensitivity added by spatial summation is unlikely to represent visual perception of images other than gratings. The use of localized stimuli believed to be matched to an impulse response of simple cortical cells is a natural way to obtain CSFs that minimize spatial summation.

The temporal presentation of the traditional gratings stimuli is usually a gradual fade in and fade out rather than flash on and off. The flash presentation technique results in higher sensitivity measured at low spatial frequencies,[27] which is considered artifactual. In free viewing, however, the saccadic scan path pattern of eye movements results in abrupt changes of eye position. Thus, the local presentation of stimuli to receptive fields usually does not follow the gradual fade-in and fade-out presentation common to most grating detection tests. Therefore, thresholds measured with abrupt presentation of grating patches better represent the perception of localized contrast in complex images. We showed[26] that the difference in sensitivity measured with the two types of temporal presentation is reduced when the number of cycles in the stimuli is increased.

Detection thresholds commonly are measured using paradigms in which only the presence of the target is recorded without the need for the subject to perceive any form, such as orientation, profile, or contrast. Such detection thresholds are sensitive but do not necessarily represent the sensitivity required for useful image interpretation or recognition. Thresholds used for pattern perception analysis should include form discrimination.[28]

In view of these arguments, we designed our CSF measurements to be used for simulation of low vision and image enhancement based on visual loss in the following way. Stimuli presented were patches of sinusoidal gratings in a Gaussian envelope (Gabor functions) of 1-octave band width. Stimuli were flashed on for 1 sec, either at horizontal or vertical orientation, and removed abruptly. The subject's task was to determine the orientation of the gratings.

Subjects: Contrast thresholds were measured for 14 normal observers and 10 visually impaired patients. The normal observers were paid volunteers, who were undergraduate and graduate students with normal or corrected-to-normal visual acuity. All patients had central visual loss due to the following: age-related maculopathy, diabetic retinopathy, central retinal vein occlusion, or retinal detachment. Central visual field loss was documented with central fields in five patients. The five other patients were determined to have central visual field loss by clinical examination. Visual acuity of the tested eye ranged from 20/60–20/136. Informed consent was obtained from all subjects before testing.

Equipment: Stimuli were generated with an Adage 3000 image processor using ten-bit digital-to-analog converters on a Tektronix 690SR, 60-Hz noninterlaced monitor. A calibrated table was used to correct for the nonlinear response of the display.

Stimuli patterns: The Gabor function stimuli were composed of a patch of either horizontal or vertical sinusoidal gratings in cosine phase, multiplied by a two-dimensional Gaussian envelope. The 1-octave stimuli contained $12/\pi \cong 4$ cycles, but only about 2 cycles were visible due to the rapid decline of the envelope. The surrounding screen luminance and the luminance before and after the patch presentation were matched with the mean luminance of the test patch (46 cd/m^2). The normal observers were tested also with a more common stimulus using a fixed-size screen ($4° \times 4°$). In this case, the number of cycles of grating presented varied with the spatial frequency. The background surrounding the grating matched the grating's mean luminance. The temporal presentation and testing paradigm were unchanged.

Procedure: Stimuli were presented in a two-alternative forced-choice paradigm. The subject pressed a button to present a grating patch and they then indicated whether the grating patch was horizontal or vertical and received auditory feedback. The psychophysical procedure was a hybrid method consisting of three successive steps. The sequence started with a staircase procedure that ran to the second reversal of direction. The procedure then was changed to a modified parameter estimation by sequential testing (PEST) method.[29] In the second phase, the stimulus contrast was controlled by the staircase, but data were collected and analyzed by the PEST algorithm. The third step started when an initial threshold estimate was determined by the PEST routine (within a confidence interval >40%), and then the stimulus contrast control was switched also to PEST. This modification prevented long, random walks that occur occasionally at the beginning of a PEST routine.[30] With this method, 50–80 presentations were required to reach the termination criterion. Seven spatial frequencies (0.5–10 cycles/degree) were interleaved randomly. The contrast was changed in 0.1-log unit steps. After termination, a psychometric function was fitted to the data to obtain threshold estimation and standard deviation (on the order of 0.1 log units).

Calculation of contrast spectra: To assess the effect of patients' visual loss on their perception of images and to determine the required enhancement, the measured CSFs were compared with contrast spectra of face images. Radially averaged power spectra were used to evaluate the spectral content of natural scene images[31] and of face images.[16] In both cases, power (amplitude squared) levels were calculated and normalized. Normalized power spectra provide information only regarding the relative power at different frequencies and cannot be used for direct comparison with the absolute values of contrast sensitivity thresholds. We used a measure of the amplitude rather than the power to evaluate the contrast spectral content.

The magnitude of the Fourier transform of five face images was averaged. The averaged Fourier magnitude then was averaged radially across all orientations. The resultant one-dimensional function of spatial frequency represents the averaged magnitude at each spatial frequency. Dividing the magnitude by the mean gray level yields an estimate of contrast (amplitude divided by mean) at each frequency. Because these spectra were averaged, the contrast available in the image locally was underestimated. Note that face spectra were calculated and discussed in terms of object frequencies in cycles/face, and therefore, a viewing distance or the span of the face in degrees was set before the relationship to retinal frequencies expressed in cycles/degree can be established.

Results: The calculated face spectrum was compared with the observers' contrast threshold curves to determine the range of spatial frequencies in the face image seen by each observer. Figure 1 illustrates the contrast thresholds (expressed as mean ± standard error of the mean) measured in 14 normal observers using the different stimuli. The threshold recorded with our localized 1-octave stimuli was 0.5–1.0 log unit higher than the threshold obtained with the fixed-size stimuli. The shape of the CSF curve was changed from the typical band pass obtained, with the fixed-size stimuli, to low-pass characteristics with our patch stimuli.

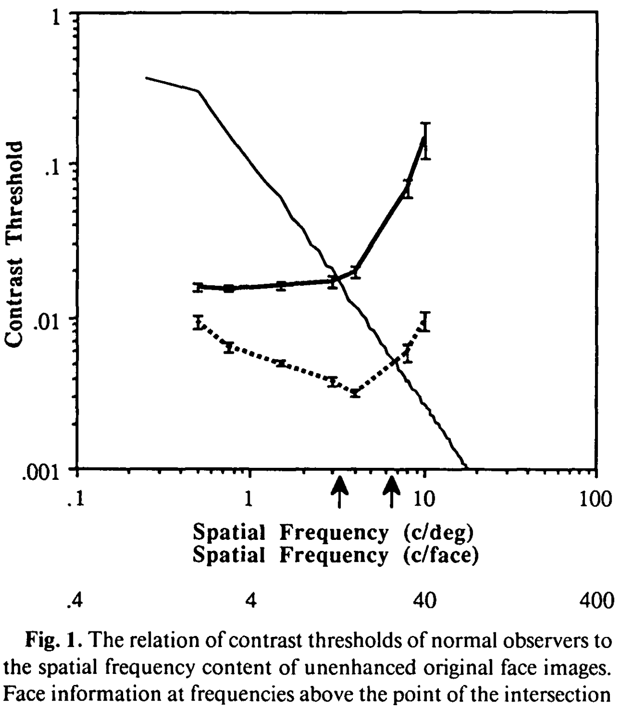


Fig. 1. The relation of contrast thresholds of normal observers to the spatial frequency content of unenhanced original face images. Face information at frequencies above the point of the intersection of the threshold curve and image spectrum (arrows) will not be visible to the corresponding observer. The mean radially averaged amplitude spectra of five faces are indicated by a thin line. The mean contrast threshold of eyes of 14 normal observers measured with localized Gabor patch stimuli (thick line) suggests that details above 13 cycles/face will not be visible. The mean contrast threshold of the same 14 normal observers measured with a fixed size screen is presented for comparison (dashed, thick line). Error bars represent standard errors of the mean.

The contrast threshold measurements were compared with the radially averaged amplitude ("contrast") spectra from five face images (Fig. 1). Each single face spectrum varied only slightly from the presented average. The face images were assumed to span 4° of visual angle (from chin to hairline). An average face spans 4° at a distance of about 2.5 m. Patients frequently complain of difficulties in recognizing faces on the street and at the office from this distance.

Spatial frequency content at a frequency higher than the point of intersection between the contrast threshold curves and the averaged spectrum curve for the face image was not detectable by the corresponding observer. Thus, the threshold measured using our localized stimuli suggested that normal observers would be unable to see face information above 13 cycles/face. This finding agrees with the limit of 15[16] and 16 cycles/face[18] found by others. The contrast sensitivity measured with the commonly used fixed-size stimuli (Fig. 1), however, suggested that higher spatial frequencies should be visible up to 25 cycles/face. This difference supports our contention that 1-octave patch stimuli result in a CSF that is more representative of pattern perception than the CSF obtained with the fixed-size, multiple-cycle grating stimuli.

Because averaged radial spectra underestimate the local contrast in the image, both types of CSF would predict even higher cutoff frequencies. We demonstrated[32] that the method used by others[16] somewhat underestimated the high-frequency contrast in the face image. Thus, frequencies above 15 cycles/face could be detected in a face image. Filtration of all high-spatial information above 32 cycles/face resulted in barely detectable changes in a 4° face image. Furthermore, when a face image spanned only 2°, many fine details were blurred. This effect too was predicted from simulations using the CSF obtained with 1-octave patch stimuli but not with the extended grating fixed size stimuli.[26] These arguments and examples supported the use of the 1-octave patch CSF as a guide for the enhancement design.

Of the ten patients, eight with central retinal visual loss provided measurable CSFs. Their contrast thresholds are compared in Figure 2 with the averaged radial spectra of the original face images and enhanced images. Three of these patients also participated in the face recognition experiment described later. Contrast threshold data from one of the eight patients were selected arbitrarily for use in the various simulations described here. This patient was a 60-year-old man

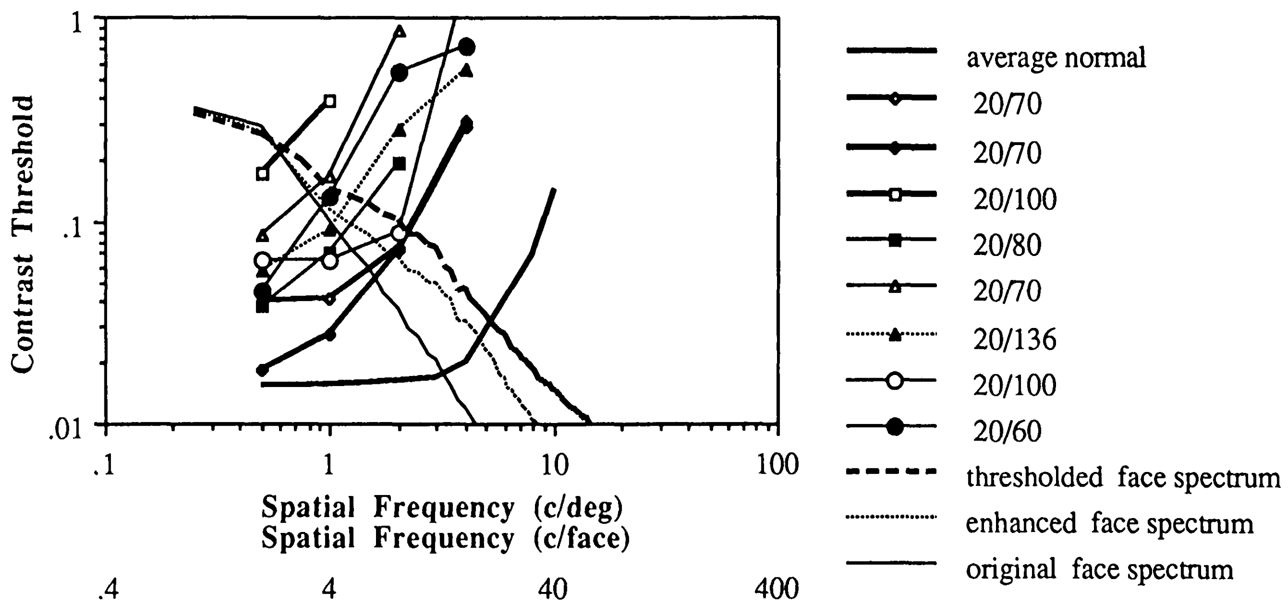


Fig. 2. The relation of patients' contrast threshold to the spatial frequency contents of enhanced and unenhanced face images. Thresholds for eight patients (curves with symbols) compared with the mean thresholds for 14 normal observers (*reproduced from Fig. 1*) show reduced sensitivity at all frequencies. The relation of patients' data to the face spectra suggests that many of these patients could benefit from both types of enhancements used at the critical range of 4–8 cycles/face. Three of these patients (bold lines) participated in the face recognition experiment described below. One of the patient's curve (triangles on dotted line) was arbitrarily selected for use in the simulations. Patients' visual acuity is noted in the legend.

with a chorioretinal scar at the fovea secondary to age-related maculopathy (visual acuity, 20/136).

The patients' sensitivity functions, compared with spectra of the original, unenhanced faces, implied that, under the same conditions, these patients would be unable to detect face information above 4–6 cycles/face. Thus, these patients were expected to have difficulty in recognizing faces at normal street-encounter distance when the face spans 4° of visual angle or less. We tuned our image-enhancement procedures to increase the contrast for frequencies above 4 cycles/face. The averaged radial spectra for the enhanced images obtained from the two different techniques described later (Fig. 2) suggested that the visibility of critical facial details at 4–8 cycles/face would be improved for many of the patients. However, because of the averaged, global nature of these amplitude spectra, this analysis should be viewed cautiously. Nonlinear simulations that consider both the local and threshold nature of contrast perception can provide a better assessment of the effect of enhancement on the appearance of images to patients with central scotoma.

Simulation of the Appearance of Enhanced Images

The Fourier analysis of images in the context of image perception frequently suggests that the CSF can be implemented as a modulation transfer function of the system in the frequency domain for filtration of images.[15,25] When applied to the simulation of appearances of images with normal[15] or abnormal[24] vision, this linear process multiplies the Fourier transform of the image by the value of the CSF at the corresponding frequency. This approach ignores the well-known, highly nonlinear characteristics of the visual system. Despite large differences in contrast thresholds for different frequencies and at different eccentricities, the appearance of suprathreshold gratings is constant or almost constant.[21,33] Therefore, the CSF should not be used to determine the apparent contrast of suprathreshold features in the image. This property of visual perception was maintained in the nonlinear simulations described here. Regardless of the nonlinear nature of the visual system, it is possible to use linear processing for simulating the appearance of images through cataract. It is appropriate in this case because the cataract can be represented as a linear optical filter. The ratio of a patient's CSF to a normal observer's CSF is assumed to measure the optical transfer function of the cataractous lens. This ratio is used as the linear filter applied to the image. (The nonlinearity of the patient's visual system further affects his or her perception. However, this effect is much smaller than that of the linear filtering applied in this case.) The more familiar linear simulations are provided mainly for comparison with our nonlinear simulations of the appearance of images to patients with central scotoma.

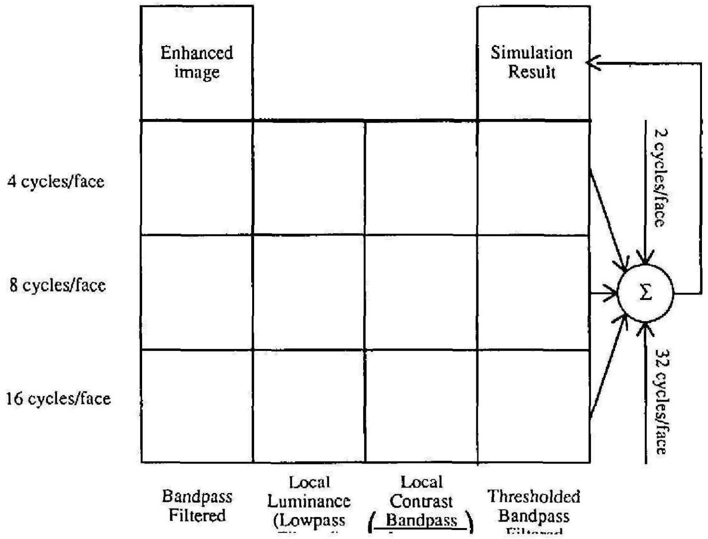
In simulating the vision of patients with central scotoma, a black area usually was used to cover parts of the image. This approach assumed that patients willingly or reflexively place the scotoma surrounding the location of their fovea over the target of interest. This assumption was disputed by some authors,[34] using the scanning laser ophthalmoscope, who showed that patients point a retinal area adjacent to the scotoma toward the target of interest; thus, keeping the scotoma from obscuring their vision. We saw similar behavior when patients looked at face images presented by the scanning laser ophthalmoscope. Therefore, our simulations assume that patients use their residual retinal function outside the scotomatous area to inspect both the CSF test stimuli and any target of interest presented to them. Thus the perception of patients was affected by visual capacities of the peripheral retina, not by the presence of a scotoma.

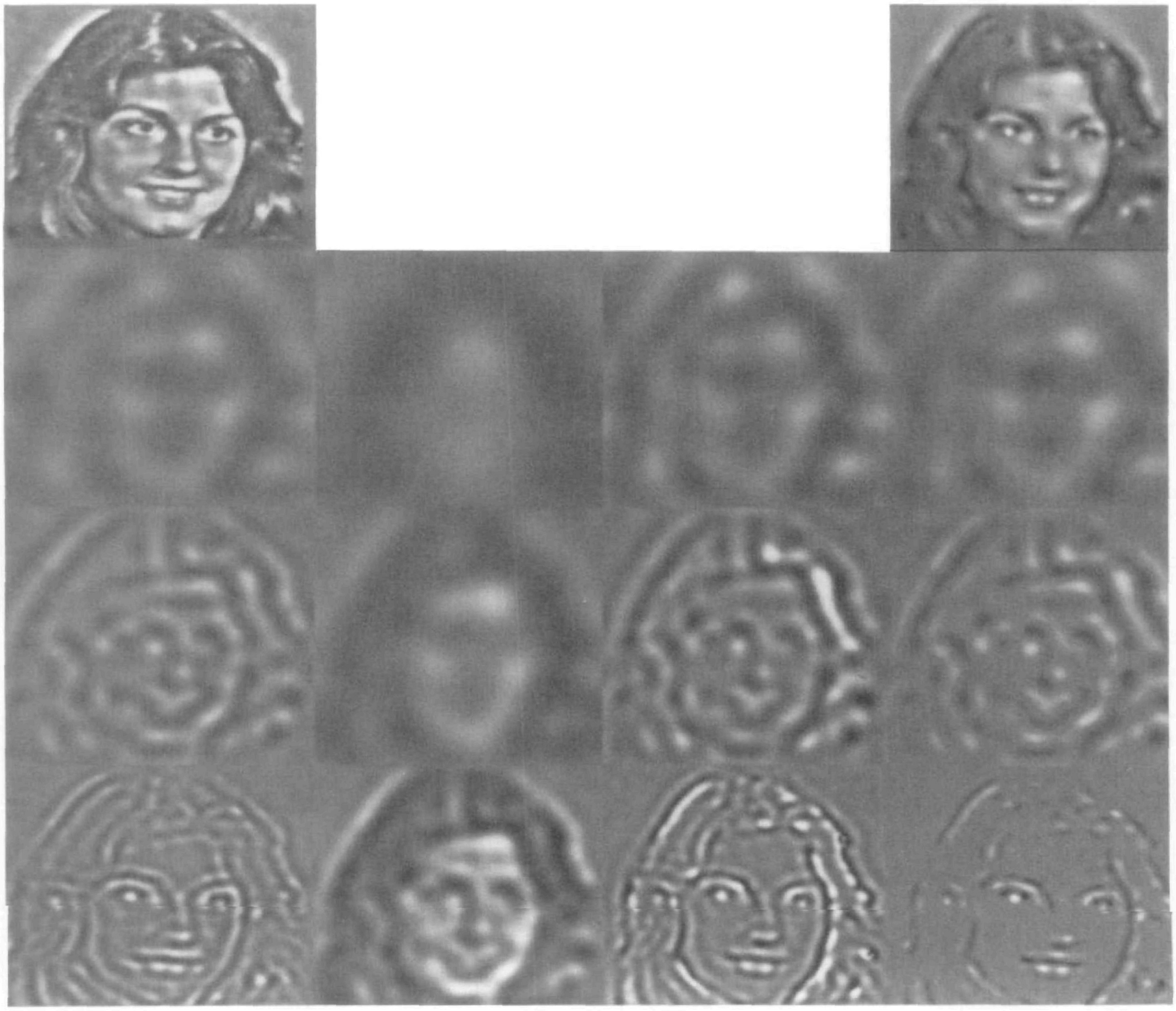
The nonlinear characteristics of the visual system are such that different thresholds need to be applied at different spatial frequencies, or scales. Therefore, our simulations were based on the pyramidal structure of local, band-pass-filtered contrast.[32] This pyramidal image-contrast structure enabled us to use threshold processing to simulate the appearance of images point-by-point and for every spatial frequency in the image. The images were sectioned in the frequency domain into 1-octave band width sections. The contrast at each spatial position was calculated by dividing the band-pass-filtered value by the low-pass-filtered value at the same point. At each level of the pyramid, every point was compared with the appropriate contrast threshold. If the contrast at that point was higher than the threshold, the amplitude of this point was not affected. If the contrast at this point was below the threshold, the amplitude was set at zero. This process is illustrated in Figure 3, using the CSF of a patient with a central scotoma due to macular disease, measured as described earlier (the patient's CSF is illustrated in Fig. 2). Thus, the final image in Figure 3B (upper right) simulates how the original image appeared to this patient. The image was processed with the stipulation that the face height would span 4° of visual angle. On this scale, this patient's visual loss had little effect on information at 4 cycles/picture (Fig. 3B, second row from top). A small effect was noted at 8 cycles/picture (Fig. 3B, third row from top), and a substantial effect on the information content was seen at 16 cycles/picture (Fig. 3B, bottom row). The complete processing also included the band of 2 cycles/picture and the band of 32 cycles/picture (data not shown). The simulated image obtained in this manner maintains the full contrast reported by patients with central visual loss and clear media and is not faded or washed out, unlike the cloudy appearance of images seen through cataracts.

Enhancement methods: Two different image-enhancement techniques, previously found to be useful with optical simulation of vision through cataracts, were used: adaptive filtering[5] and adaptive thresholding.[35] These techniques do not lend themselves to direct application of the linear-enhancement process based on each patient's CSF. Nevertheless, the enhancement can be tuned to cover the spatial frequency range above 4 cycles/face that our previous analysis suggested was important for improved face recognition for our patient population. In addition, these techniques, if found useful, may be used with a video frequency to process television images at the rate at which they arrive.

Adaptive filtering: The adaptive filtering technique was developed originally for images degraded by cloud cover or shadows.[36] The technique implements a high-pass filter that may be modified locally based on local image brightness. The adaptability of the technique was not used in our study; instead, we used this technique as a fast-space domain-based algorithm for high-pass filtering. Because the adaptive-enhancement technique has been elaborated elsewhere,[3,36] it will be described briefly here. First, the image is separated into spatial low- and high-frequency components. The low-frequency component is obtained by calculating (for each pixel) the average brightness level in a small window around the pixel. The high-

Fig. 3. Illustration of the method used for simulating the appearance of an enhanced face image (spanning 4° of visual angle) to a low-vision patient whose contrast sensitivity function (CSF) is illustrated in Fig. 1. Use the line drawing (*top*) as a map of the composite image (*bottom*). *Upper left,* the enhanced image; *upper right,* the final simulated appearance of the same image to the patient. The three rows represent processing at different spatial frequencies on the pyramid. Left-most column in each row represents the band-pass-filtered image obtained from the original image. The second column in each row represents the corresponding low-pass-filtered version for the same scale, ie, all the energy below the band represented in the first column. Third column represents the contrast images. Contrast arrays are bipolar and DC level of 128 has been added arbitrarily to present those arrays as images. Images in the fourth, or right-most, column represent the thresholded band-pass-filtered images. For each image in the third column, each point was tested against the threshold value illustrated in Fig. 1 for the corresponding spatial frequency. If the contrast of the image at that point is above threshold, the corresponding point from the left-most image is maintained and reproduced on the right-most column. If the contrast at a point is below threshold, the corresponding point is set to zero (gray) at the right-most image. Simulated image (*upper right*) is generated by summing all the images in the right-most column. Actual processing included two more rows, one at 2 cycles/picture and another at 32 cycles/picture (not shown).


Enhanced image
Simulation Result
4 cycles/face
8 cycles/face
16 cycles/face
2 cycles/face
32 cycles/face
Σ
Bandpass Filtered
Local Luminance (Lowpass
Local Contrast (Bandpass
Thresholded Bandpass



frequency component is obtained by subtracting the low-frequency component from the original image and is then amplified. We used an amplification factor of 5 for all images used in this study. If the original, unenhanced image occupies most of the display's dynamic range, the amplified high-frequency component will exceed the available range. To provide the dynamic range required for this amplification, the local luminance level or the low-frequency content also was modified. The alternating current (AC) portion of the low-frequency content was multiplied by a factor of 0.9, permitting an additional range for the amplified high-frequency component. These two modified components then were added to produce the final image. The size of the small window (21 × 21 pixels) was tuned to amplify components above 4 cycles/face (Fig. 2). Note, however, that due to the averaged nature of the spectra, at no frequency was the change as large as the applied amplification factor of 5.

Adaptive thresholding: The adaptive-thresholding technique[35] usually is not considered an enhancement technique but may serve as such, especially for the visually impaired. Thresholding is a method of transforming a gray-tone image into a binary one (ie, an image with only two levels, black and white). The binary image has inherent high contrast, and if it maintains the original image's information in a satisfactory way, it may be useful as an enhancement tool. This technique was found to be effective with optically simulated cataracts.[4] In addition, binary display devices with high brightness and contrast may be less expensive and, thus, provide the medium for such an enhancement system.[37] Simple thresholding sets all points in the image above a threshold value to white and all darker points to black. However, in many cases, no threshold value will give a binary image of sufficient detail or clarity. Therefore, techniques using variable and adaptive thresholding that change across the image as a function of local image properties are needed. Our technique[35] was a modification of one such adaptive-thresholding technique.[38] The image was divided into small, nonoverlapping windows. The gray-level histogram of all pixels in the window was tested for bimodality, ie, two distinct lobes or peaks. If such lobes were found, the value between them was set as the proper threshold for this local area. When all bimodal windows were assigned a threshold value, the thresholds over the entire image were interpolated for all windows (including those without bimodal histogram) using a two-dimensional linear interpolation. Based on those interpolated, smoothed thresholds, the threshold value was calculated for each point, and subsequently it was designated as white or black. The two parameters used in our modified version, the window size (8 × 8 pixels) and the bimodality threshold value (3), were set empirically to obtain detailed thresholded features. On examination of sample spectra, they were found to result in enhancement of spatial frequencies above 4 cycles/face (Fig. 2).

All images processed with both techniques were enhanced using the same parameters for the following simulations and for patient testing procedures.

Results: Figure 4 simulates appearances of original and enhanced face images for patients with macular disease and cataract. The CSF used in both cases was the same one used here and illustrated in Figure 2. One case simulated vision with central scotoma, a nonlinear simulation. Another case simulated an image seen through a cataract, a linear optical filter. The simulation suggested that face recognition may be improved substantially for both categories of patients and for both types of enhancement. The results of enhancement were even more favorable for an image of a street scene containing more relevant high-frequency information than a face image (Fig. 5).

Because the simulation removed high frequencies in the same range enhanced by both techniques, it may seem trivial that the appearance of the enhanced images was improved. However, the local nature of image contrast and threshold nonlinearity of the simulated appearance may result in enhanced contrast not being manifested in the simulation if it remains subthreshold. Thus, our simulations suggested that if the measured contrast threshold represents patient perception, the applied enhancement should improve image perception as simulated. This verification was important because the ratio of the patients' CSF to normal observers' CSF was much higher than the enhancement applied, and higher levels of enhancement were not practical due to the limited dynamic range of the display. These simulations were used as pilot experiments to test the effects of enhancement with various parameters before actual testing of patients began.

Patients' Recognition of Enhanced Images of Faces

With these encouraging simulation results, we tested the direct effect of enhancement on face recognition of low-vision patients.

Methods: Thirty-eight patients (age range, 27–91 yr; mean, 64) with central visual loss due to macular disease in one eye and a visual acuity of <20/70 were selected. Preference was given to patients with good visual acuity in one eye of >20/40 so that familiarity with celebrities could be verified. Their diagnoses were heterogeneous, as is typical of a low-vision population, and included various types of age-related maculopathy (13), diabetic retinopathy with macular le-

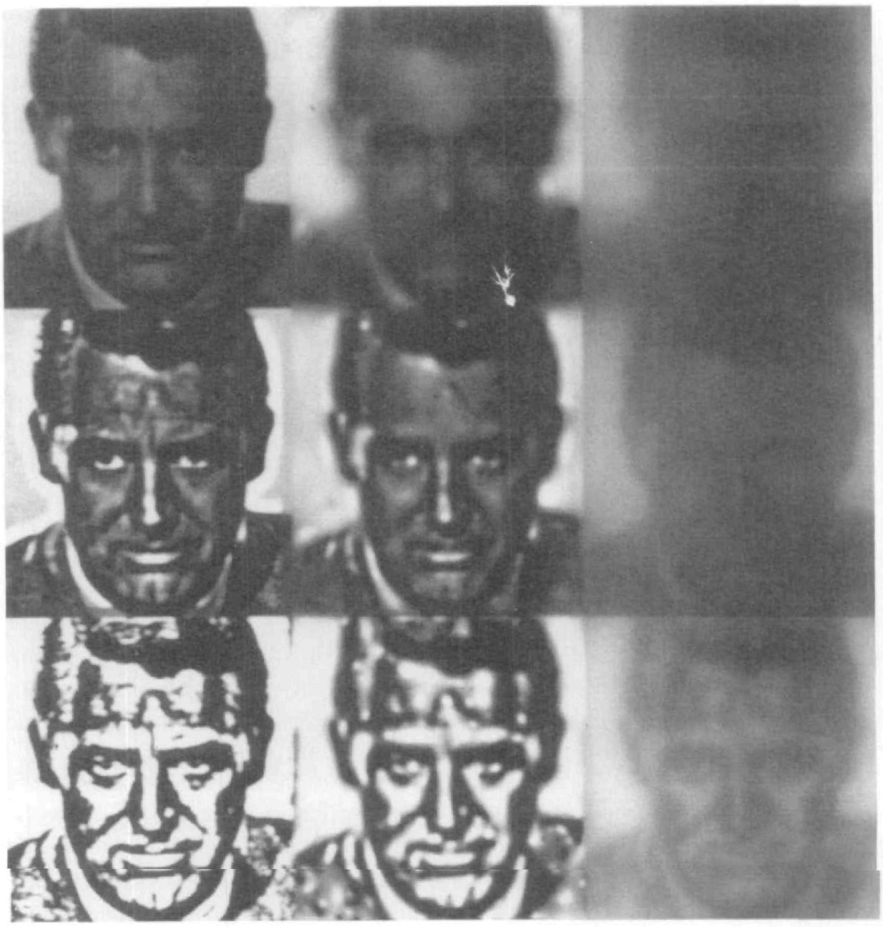


Fig. 4. Simulations of appearances of original images and enhanced face images (spanning 4° of visual angle) for patients with central scotoma and with cataract. Both sets of patients were assumed to have the same contrast sensitivity function. The left column represents what persons with normal vision see. The middle column simulates the appearance to a patient with central scotoma (nonlinear processing). The right column simulates the appearance of the same images to a patient with cataract (linear filtering). The top row is the original, unenhanced image; the middle row, the adaptively enhanced image; and the bottom row, the adaptively thresholded image. Note the improvement in visibility of detail for both sets of patients and both types of enhancement.

sions (8), retinal vein occlusion (8), detached retina including the fovea (5), and various other foveal lesions. The visual acuities in the affected eyes ranged from 20/70–20/200 (mean, 20/130). Central visual fields were recorded frequently before referral; such fields, indicating central scotoma to 6/1000 target on the Autoplot (Bausch and Lomb, Rochester, NY) were available for 28 of the 38 patients. The other ten

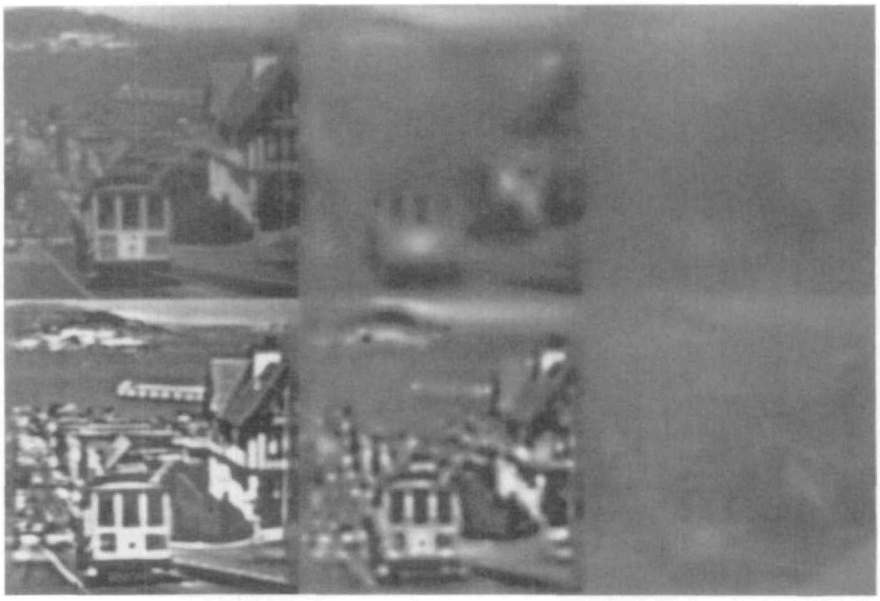


Fig. 5. Comparison of the simulated appearances of an image with high spatial frequency content with the appearance of enhanced image for patients with central scotoma and with cataracts. The columns are arranged as in Fig. 3. The top row is the original image, which is degraded greatly. The bottom row is the adaptively enhanced image, providing more details than the original image for the visually impaired. Here too the images are assumed to span 4° of visual angle, and the contrast sensitivity function is the same one applied in Figs. 2 and 3.

patients were determined to have central field loss based on their clinical examination including fundus appearance, measured acuity, and clarity of the optical media. Many of the patients had more than one ocular disorder and were classified according to the condition determined clinically to account for most of the demonstrated visual loss. We tested 21 of these patients, selected randomly, with the adaptive-enhancement[36] technique and 17 with the adaptive-thresholding technique.[35] A smaller group of eight patients with various anterior segment opacities including cataracts and corneal dystrophies (visual acuity, 20/60–20/200) were tested also: four with adaptive thresholding and the rest with adaptive enhancement. Informed consent for participation in the study was obtained from each patient before testing.

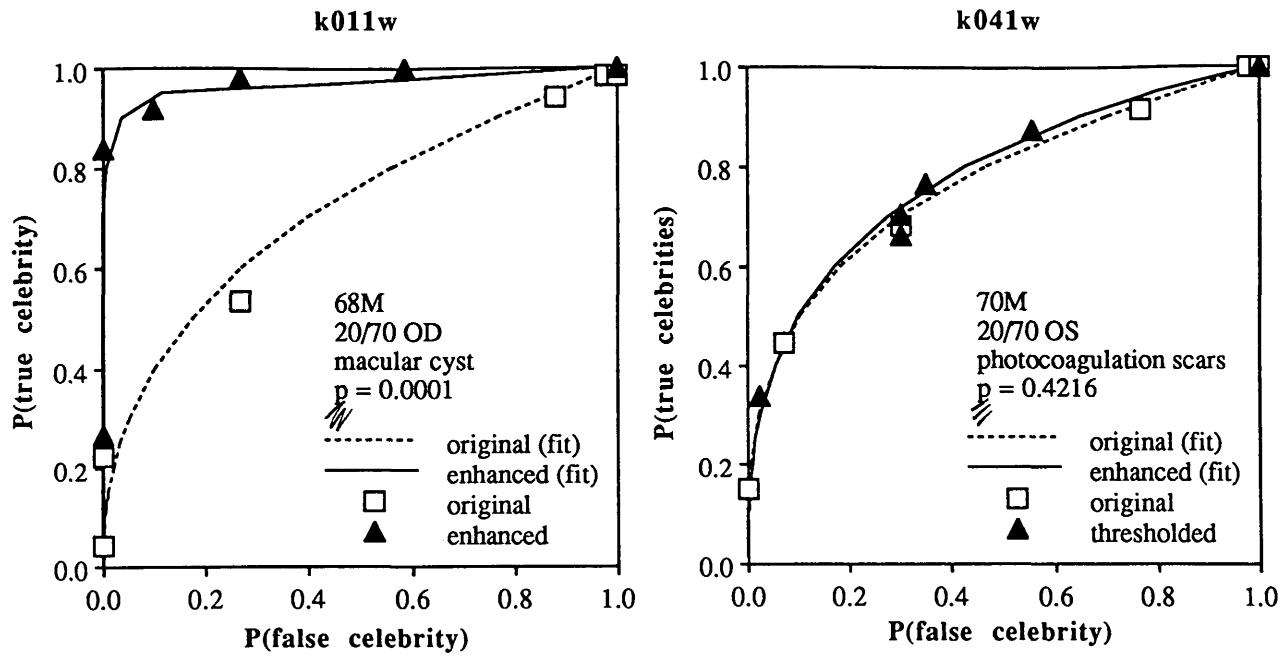
Images: Photographs of 50 celebrities and 40 unfamiliar people were used. The celebrity photographs were expected to be familiar to most patients in our study who were older than 60 yr of age. All photographs presented a face in full frontal position. Celebrities included many television and screen actors and actresses. Most were young although a few older actors and television personalities were included. Other older famous celebrities included politicians, mostly presidents or presidential candidates. The unfamiliar people also were from two age groups: young models and residents of a retirement home. All were photographed by the same photographer for a study on facial changes with aging. Most faces were photographed on a bland, bright background, but a few photographs had some detail in the background. Transparencies of celebrity and unfamiliar faces were digitized at a resolution of 256 × 256 pixels and at 256 gray levels. Illumination was adjusted to obtain good dynamic range and clear visibility of all images. All images were digitized under the same magnification and illumination conditions. All faces were cropped to a square area from the chin to the hairline, and the external upper outline usually was removed. Cycles per face height were used rather than cycles per face width because in our format all faces had the same height, but the width varied across the set. The face images were enhanced with both the adaptive-enhancement algorithm[5] and the adaptive-thresholding technique.[35] The same parameters used for the simulations were used for processing all images.

Procedure: The images were presented to the subject sitting in a lighted room on a 60-Hz noninterlaced video monitor with their display size adjusted to 4° × 4°. Original (unenhanced) and enhanced images were intermixed and presented in random order by the computer. Enhanced images presented to each patient were of one type only (adaptive enhancement or adaptive thresholding). Thus, a total of 180 images were presented to each subject. Subjects indicated their level of confidence in recognizing a face as a celebrity on a scale of 1–6. A rating of 1 meant that the subject was positive that the face belonged to a celebrity, whereas 6 meant that the face was clearly visible but not recognizable as a celebrity. A rating of 2 indicated that the subject was fairly sure but not positive that the face was a celebrity; 5 signified that the subject was fairly sure that the face was not a celebrity. The ratings 3 and 4 were used when features were difficult to discern. A score of 3 meant that the subject had an inkling that the image was a celebrity; 4 signified that the image was not clear but judged not to be a celebrity. An even number of rating levels was used to reduce the tendency of patients to select the middle level and to force a choice in each case. After testing the computer presented to the patient's better eye the celebrity images that were not recognized in both enhanced and unenhanced modes. If the patient could not recognize a particular celebrity with the better eye, that celebrity was redefined as an unfamiliar person in analyzing that patient's responses.

Patient responses were used to calculate receiver operating curves[39] (ROCs), plotting the probabilities of true celebrity versus false celebrity. The probability of true celebrity was the probability that celebrities were identified correctly as such by the patients. The probability of false celebrity was the fraction of noncelebrities identified as celebrities by the patients. For the first data point on the curve, the fraction of celebrities receiving the rating of 1 was plotted against the fraction of noncelebrities receiving the same rating. For other points on the curve, the cumulative fractions are calculated. Separate curves were calculated for original and enhanced images.

Because the same faces were presented in both forms, the responses for each face were assumed to be correlated, requiring a correlated ROC analysis.[40] The area under the ROC (A_z) was taken as a measure of recognition.[39] If the enhancement improved face recognition, the area under the ROC for the enhanced image should be greater than that for the original image. The level of correlation was used in determining the significance of the difference between the two areas.

Results: Most of the patients (39 of 46) had improved face recognition with the enhanced images compared with the original, unenhanced images (Fig. 6). Analysis of the difference between the two areas under the ROC curves indicated a statistically significant increase in recognition for 9 of the 21 patients with macular disease tested with the adaptive-enhancement technique ($P < 0.05$). For the two patients in this group of 21 whose recognition decreased with enhancement, the difference was not significant. Of



a b

Fig. 6. Receiver operating curves (ROCs) for two patients comparing their recognition of adaptively enhanced face images vs unenhanced, original images. A 68-year-old man with foveal cyst OD (visual acuity 20/70). The area difference is significant ($P = 0.0001$) (**A**). A 70-year-old man with age-related maculopathy OS (visual acuity 20/70) (**B**). The difference between the areas under the curves is not significant ($P = 0.421$). Note the difference in performance between the two patients despite their equal visual acuity.

the 17 patients with macular disease tested with the adaptive-thresholding technique, 6 had a significant increase in recognizing the enhanced images, and 1 had a significant decrease.

Three of the eight patients with anterior segment disorders also had a significant increase in recognizing the enhanced images (Fig. 7). In response to questions and often spontaneously, patients reported that the enhanced images were significantly clearer, sharper, and easier to see. Although many patients noted that the binary thresholded images appeared distorted and cartoon-like, they were still clearer and easier to recognize. Frequently celebrities who were not recognized without enhancement were recognized easily when the image was enhanced.

Presentation of the ROC curves was usually the final stage in the analysis. This was sufficient when only a few ROC curves were presented. Because we reviewed and presented the results from more than 40 observers, we calculated a measure of the improvement (or decrement) in performance attained with the enhanced images. The improvement was measured as the ratio of areas under the curves ($A_z[enhanced]/A_z[original]$). The results for all patients are given in Figure 7. The level of improvement varied from one patient to the next. This improvement was correlated negatively with the area under the curve for the original image ($n = 46$, $r = -0.39$, $P = 0.018$). The most improvement with enhancement can be realized only by patients whose original performance was degraded considerably by his or her visual impairment. The improvement, however, was not correlated with visual acuity ($r = 0.0093$, $P = 0.16$).

Because the potential improvement in face recognition for each patient depended on their performance without the enhancement, a comparison across patients required normalization. We normalized the measurement of improvement by calculating *gain* as the ratio of improvement and maximal possible improvement:

$$\text{gain} = \frac{A_z(\text{enhanced}) - A_z(\text{original})}{1 - A_z(\text{original})} \qquad (1)$$

The mean *gain* was 65% of the maximal possible improvement for the nine patients with macular disease who had significant improvement with the adaptive-enhanced images and 41% for the seven patients who showed significant improvement with adaptive thresholding. The mean gain for the three patients with anterior segment opacities who had significant improvement was 74% of the maximal possible im-

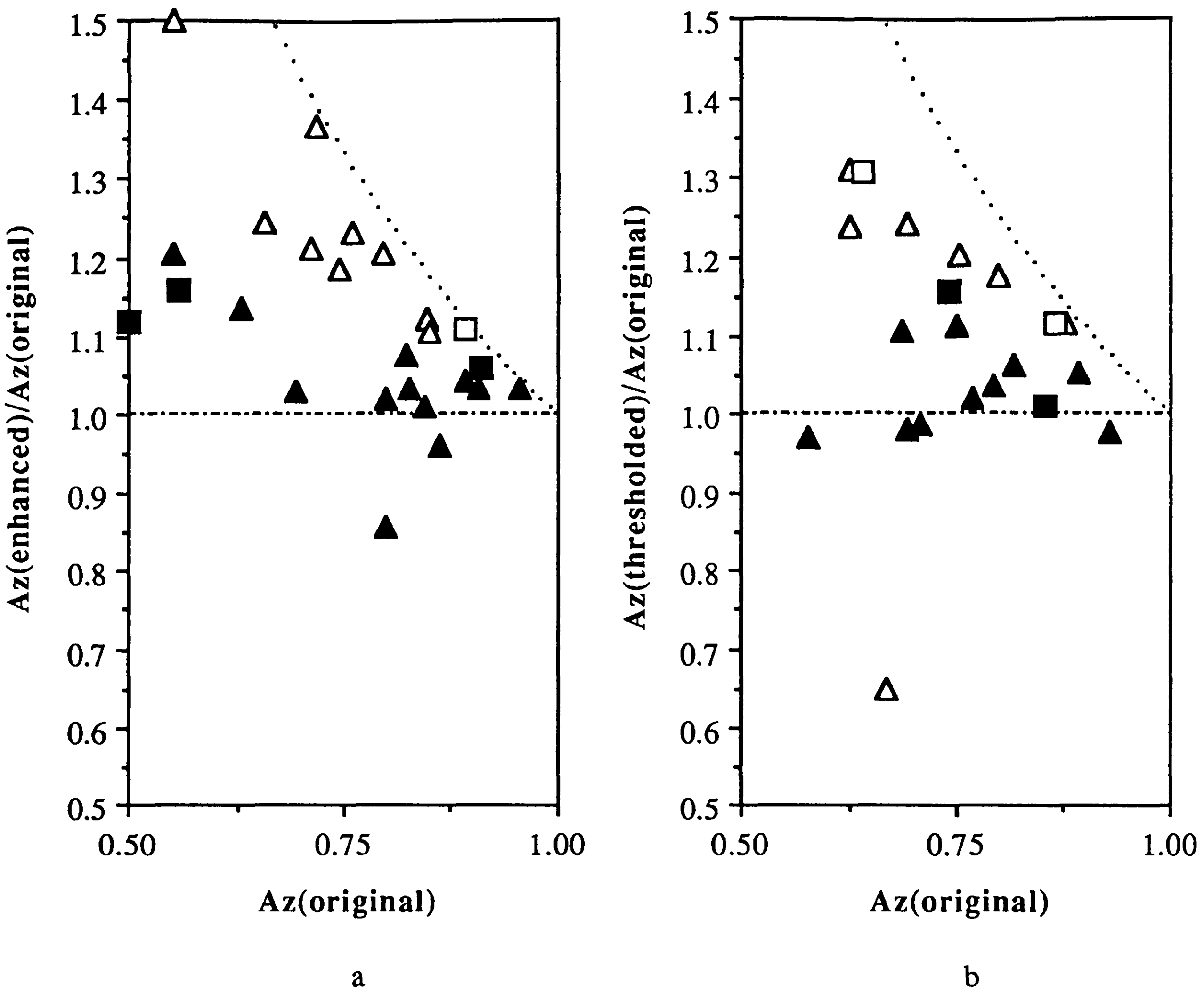


Fig. 7. Change in recognition for all patients as a function of recognition without enhancement. Data points above dashed horizontal line represent improvement, but those below represent a decrement in recognition with the enhanced images. Results of patients tested with adaptive enhancement (A), results of patients tested with adaptive thresholding (B). Patients with central scotoma secondary to macular disease (triangles), patients with anterior segment opacities (squares). Open symbols represent statistically significant change ($P < 0.05$). Filled symbols represent nonsignificant changes. The dotted curve delineates the maximum improvement possible for each level of performance without enhancement.

provement. For many of the patients the improvement was close to the possible maximum (Fig. 7); however, the *gain* was not correlated significantly with their initial performance (r = 0.461, $P = 0.307$).

Discussion

Image enhancement can provide a measurable, statistically significant improvement in face recognition for the visually impaired. For many, improvement was considerable and approached the maximum. These encouraging results suggested that the technology of image processing may be useful for the visually impaired. Although the effect was significant only for about one half of the patients, we believe that this is sufficient to justify further investigation. Before the technique is applied to broadcast television or other live video applications, we will first investigate whether better results can be obtained using other enhancement techniques. Specifically, the value of enhancement based on the individual patient's visual loss should be evaluated. The benefit from enhancement tailored to the individual patient should be sizable to justify the complexity and expense of this approach.

Enhancement tailored to the individual patient does not refer to the linear preemphasis model in which the CSF was used to determine enhancement throughout the spatial spectrum. As discussed earlier, such enhancement is unnecessary and inefficient. Rather, a narrow band of frequencies in the image just below the patient detection threshold should be enhanced. Lower frequencies, suprathreshold features, do not need to be enhanced but may need to be attenuated to enable enhancement of the higher frequency range. The specific range of frequencies to be enhanced depends on both the patient's CSF and the image contrast spectrum. Thus, the optimal enhancement may need to be varied with changes in the image content or observation distance. A patient may need to retune his or her system occasionally when the image spectra change drastically. In addition to determining the optimal enhancement technique, the effects of image motion should be evaluated, and an application to color images should be developed.

Using band-pass-filtered face images, others[41] compared face recognition performance of such images with center frequencies at 4.0, 11.3, and 32.0 cycles/face width. They found that the performance of normal observers did not depend on the center frequency of the band-pass-filtered image and concluded that no critical spatial frequency could be enhanced to improve face identification for low-vision observers. Our results contradict this conclusion.

Despite these discrepancies, both sets of conclusions may be reconcilable. In the other study, subjects' responses were evaluated in a forced-choice paradigm, and the authors considered results that were significantly above chance as the ability to recognize faces.[41] However, subjects doing a threshold task are usually unsure and displeased with their performance during the test. It is possible that, for the purpose of social interaction, patients require a higher level of recognition than is achieved at threshold. Our paradigm using the rating system directly assessed the level of confidence that each subject had in recognizing faces. This difference in method may explain why the conclusions in our study differed from those of the previous one.[41]

Although spatial frequencies of 4–6 cycles/face may provide sufficient information for face recognition in matching tasks, performance of this task is virtually impossible with filtration below this level. We showed that many patients with central scotoma lose contrast sensitivity to the point that filtration of face images (in a normal street-encountered distance) will be below 4 cycles/face. These patients, therefore, could benefit from enhancement of these critical frequencies.

Some of the quantitative differences may be due to differences in the measuring units of cycles/face. The previous study[41] used an ear-to-ear measurement as the face width. We used face height, ie, from the chin to the beginning of the hairline, as our unit. Furthermore, these authors used images that were elongated vertically, creating a height-to-width ratio of about 2:1. Thus, their 4 cycles/face width images actually contained 8 cycles/face height in our units.

One limitation of our testing paradigm is that it can be applied to each patient only once before all the faces become familiar. Thus, although results with the two enhancement techniques (Fig. 7) appear similar, they were not compared for effectiveness in the same patient. In any case, they were not considered as competing methods of enhancement for the same display device but rather as two separate alternatives for two different types of devices, and both were found useful to a similar extent.

Because many of our patients were tested with their worse eye, their performance represents an unadapted, untrained response. Such responses may be expected from recently disabled, visually impaired persons and may not be the same for a patient with long-standing binocular impairment. We chose these patients to verify their familiarity with the celebrities. We compared their performance under the two conditions of enhanced and unenhanced images, and we believe that a similar level of improvement will occur for the adapted patient if the deficit is sensory and the adaptation is mainly cognitive in nature. The performance of the eight patients with binocular visual loss we tested was similar to the performance of the 38 patients with monocular visual loss.

We tested patients for the effects of enhancement only on face recognition, but our simulations suggest that the improved visibility afforded by enhancement also may be valuable for viewing various scenes (Fig. 4). The improvement in recognition may be greater for images in which the high spatial frequencies contain more important and critical information than in face images. In many situations, spatial information in much higher frequencies (1-2 octaves higher than in the case of faces) may be required to improve target discrimination.[42] In these situations, high spatial frequencies were more important for identifying complex objects than the low frequencies.

Evaluating the performance of low-vision patients with visual aids frequently is difficult as a result of the large variability in the initial patient performance without an aid or with an alternative aid. Normalizing improvement by the maximal possible performance (*gain*) appears to resolve this difficulty and to enable comparisons across patients and across patient categories. This technique may provide a useful tool for comparing different enhancement algorithms or imaging and diagnostic or treatment modalities.

Key words: contrast sensitivity, face recognition, low vision, macular disease, visual aids

Acknowledgments

The authors thank S. Lubars for valuable technical help, Dr. D. Getty for excellent advice regarding data analysis, Dr. C. Metz for generous permission to use his software, and Dr. C. Belcher for referring the patients with anterior segment opacities.

References

1. Covington G: Let Your Camera Do the Seeing: The World's First Photography Manual for the Legally Blind. Washington, DC, National Access Center, 1981.
2. Kenney AP: A range of vision: Museum accommodation for visually impaired people. J Vis Impair Blind 77:325, 1983.
3. Peli E and Peli T: Compensatory image enhancement for low-vision publication. Optom Vis Sci 59:41P, 1982.
4. Peli E, Arend LE Jr, and Timberlake GT: Computerized image enhancement for visually impaired persons: New technology, new possibilities. J Vis Impair Blind 80:849, 1986.
5. Peli E and Peli T: Image enhancement for the visually impaired. Optical Engineering 23:47, 1984.
6. Isenberg L, Luebker A, and Legge GE: Image enhancement for normal and low vision. ARVO Abstract. Invest Ophthalmol Vis Sci 30(Suppl):396, 1989.
7. Rubin GS, Turano K, and Schuchard RA: Effect of contrast stretching in spatial frequency range on reading grades: OSA Annual Meeting 1988. *In* Technical Digest Series, Vol 11. Washington, DC, Optical Society of America, 1988, pp. 97.
8. Lawton TB: Improved word recognition for observers with age-related maculopathies using compensation filters. Clin Vis Sci 3:125, 1988.
9. Lawton TB: Improved reading performance using individualized compensation filters for observers with losses in central vision. Ophthalmology 96:115, 1989.
10. Ellis HD, Jeeves MA, Newcombe F, and Young A, editors: Aspects of Face Processing. Boston, Martinus Nijhoff, 1986.
11. Rubin GS and Schuchard RA: Does contrast sensitivity predict face recognition performance in low vision observers? *In* Technical Digest of the Topical Meeting on Noninvasive Assessment of the Visual System, Vol 3. Washington, DC, Optical Society of America, 1990, pp. 130–133.
12. Sekuler R, Owsley C, and Hutman L: Assessing spatial vision of older people. Optom Vis Sci 59:961, 1982.
13. O'Toole AJ, Millward RB, and Anderson JA: A physical system approach to recognition memory for spatially transformed faces. Neural Networks 1:179, 1988.
14. Harmon LD: The recognition of faces. Sci Am 229:71, 1973.
15. Ginsburg AP: Visual information processing based on spatial filters constrained by biological data. Doctoral dissertation, Cambridge University, Cambridge, England, 1978.
16. Fiorentini A, Maffei L, and Sandini G: The role of high spatial frequencies in face perception. Perception 12:195, 1983.
17. Sergent J: Microgenesis of face perception. *In* Aspects of Face Processing, Ellis HD, Jeeves MA, Newcombe F, and Young A, editors. Boston, Martinus Nijhoff, 1986, pp. 17–33.
18. Woodhouse JM: The application of contrast sensitivity to studies on image detection and recognition. PhD thesis, University of Cambridge, 1975.
19. Rubin GS and Seigel K: Recognition of low-pass filtered faces and letters. ARVO Abstract. Invest Ophthalmol Vis Sci 25(Suppl):71, 1984.
20. Georgeson MA and Sullivan GO: Contrast constancy: Deblurring in human vision by spatial frequency channels. J Physiol (Lond) 252:627, 1975.
21. Cannon MW Jr: Perceived contrast in the fovea and periphery. J Opt Soc Am [A] A2:1760, 1985.
22. Bruce V: Recognizing familiar faces. *In* Aspects of Face Processing, Ellis HD, Jeeves MA, Newcombe F, and Young A, editors. Boston, Martinus Nijhoff, 1986, pp. 107–117.
23. Hay DC, Young AW, and Ellis AW: What happens when a face rings a bell? The automatic processing of famous faces. *In* Aspects of Face Processing, Ellis HD, Jeeves MA, Newcombe F, and Young A, editors. Boston, Martinus Nijhoff, 1986, pp. 136–144.
24. Watson AB, Barlow HB, and Robson JG: What does the eye see best? Nature 302:419, 1983.
25. Lundh BL, Derefeldt G, Nyberg S, and Lennerstrand G: Picture simulation of contrast sensitivity in organic and functional amblyopia. Acta Ophthalmol (Copenh) 59:774, 1981.
26. Peli E, Goldstein R, Young G, and Arend L: Contrast sensitivity functions for analysis and simulation of visual perception. *In* Technical Digest of the Topical Meeting on Noninvasive Assessment of the Visual System, Vol 3. Washington, DC, Optical Society of America, 1990, pp. 126–129.
27. Breitmeyer B and Julesz B: The role of on and off transients in determining the psychophysical spatial frequency response. Vision Res 15:411, 1975.
28. Fleck H-J: Measurement and modeling of peripheral detection and discrimination thresholds. Biol Cybern 61:437, 1989.
29. Lieberman HR and Pentland AP: Microcomputer-based estimation of psychophysical thresholds: The best PEST. Behavioral Research Methods of Instrumentation 14:21, 1982.
30. Klein SA and Manny RE: Efficient estimation of thresholds with a small number of trials. *In* Technical Digest of the Topical Meeting on Noninvasive Assessment of the Visual System, Vol 7. Washington, DC, Optical Society of America, 1989, pp. 80–83.
31. Field DJ: Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am [A] 4:2379, 1987.
32. Peli E: Contrast in complex images. J Opt Soc Am [A] 7:2030, 1988.
33. Cannon MW Jr and Fullenkamp SC: Perceived contrast and stimulus size: Experimental simulation. Vision Res 28:695, 1988.
34. Timberlake GT, Peli E, Essock EA, and Augliere RA: Reading with macular scotoma: II. Retinal locus for scanning text. Invest Ophthalmol Vis Sci 28:1268, 1987.
35. Peli E and Lahav M: Drusen measurements from fundus photographs using computerized image analysis. Ophthalmology 93:1575, 1986.
36. Peli T and Lim JS: Adaptive filtering for image enhancement. Optical Engineering 21:108, 1982.
37. Peli E: Visual issues in the use of head-mounted monocular display. Optical Engineering 29:883, 1990.
38. Chow CK and Kaneko T: Automatic boundary detection of the left ventricle from cineangiograms. Comput Biomed Res 5:388, 1972.
39. Swets JA and Pickett RM: Fundamentals of accuracy analysis. *In* Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York, Academic Press, 1982, pp. 15–45.
40. Metz CE, Wang P, and Kronman HB: A new approach for testing the significance of differences between ROC curves measured from correlated data. *In* Proceedings of the VIII Conference on Information Processing in Medical Imaging, Deconinck F, editor. The Hague: Martinus Nijhoff, 1983, pp. 431–445.
41. Schuchard RA and Rubin GS: Face identification of bandpass filtered faces by low vision observers. ARVO Abstract. Invest Ophthalmol Vis Sci 30(Suppl):396, 1989.
42. Norman J and Ehrlich S: Spatial frequency filtering and target identification. Vision Res 27:87, 1987.