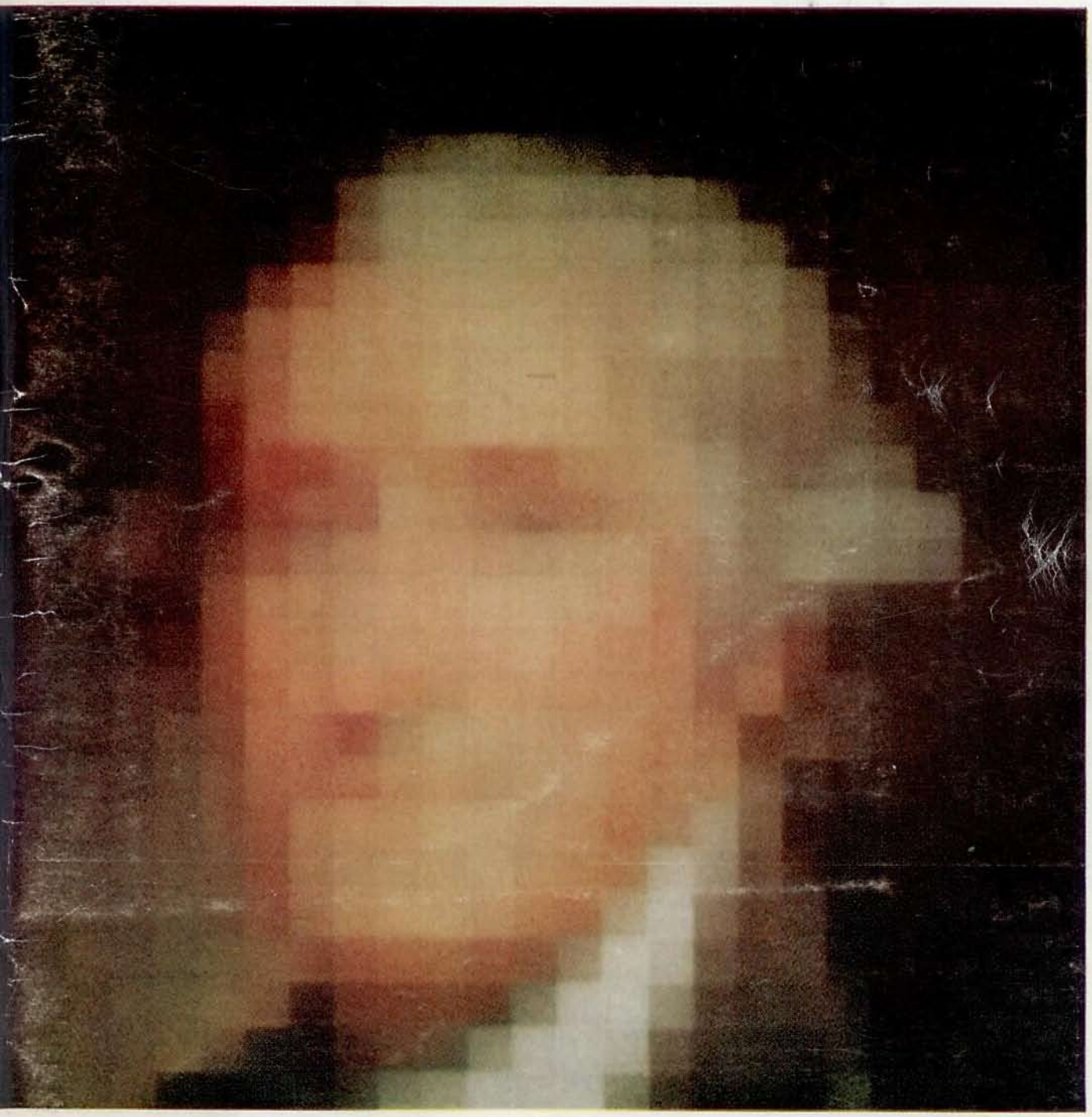
SCIENTIFIC AMERICAN



THE RECOGNITION OF FACES

ONE DOLLAR

November 1973

THE RECOGNITION OF FACES

One of the subtler tasks of perception can be investigated experimentally by asking how much information is required for recognition and what information is the most important

by Leon D. Harmon

Faces, like fingerprints and snowflakes, come in virtually infinite variety. There is little chance of encountering two so similar they cannot be distinguished, even on casual inspection. Unlike fingerprints and snowflakes, however, faces can be recognized as well as discriminated. It is possible not only to tell one from another but also to pick one from a large population and absolutely identify it, to perceive it as something previously known, just as in reading one not only can tell that an *A* is different from a *B* but also can identify and name each letter.

Why are faces so readily recognized? In seeking the answer to this question my colleagues and I posed several related but more modest questions that we believed would be more amenable to experimental investigation: How can a face be formally described? Given a verbal description, how well can a particular face be identified? To what extent is recognition impaired when the image of a face is blurred or otherwise degraded? What kinds of image degradation most seriously affect recognition? Can faces be classified and sorted as numerical data?

This inquiry was inspired by yet another question: How can a computer be made to recognize a human face? This question remains unanswered, because pattern recognition by computer is still too crude to achieve automatic identification of objects as complex as faces. Machines can recognize print and script, craters and clouds, fingerprints and pieces of jigsaw puzzles; the recognition of human faces, however, is a much subtler task.

Even though machine recognition of faces has not been attained, the investigation of how it might be done has led to a number of related issues that in themselves are worthwhile (and tractable) areas of research. Several new approaches to problems in the manipulation of visual data have emerged. I shall recount here four series of experiments that were directed to an understanding of recognition. The first is concerned with how artists reconstruct faces from descriptions and how closely the resulting portraits resemble the person described. Next I shall comment on a set of experiments in which faces were identified from pictures that had limited information content. The third approach examines the recognition of faces from formal numerical descriptions. Finally, I shall describe a system in which man and computer interact to identify faces more efficiently than either could alone.

If one could devise an objective formulation of the criteria used by an artist in drawing a portrait, a set of properties useful for automatic recognition might emerge. One kind of art that we thought might provide useful information is the sketches drawn by police artists (called face-reconstruction artists) from descriptions provided by witnesses. (Another promising possibility is the caricature, but we have not yet studied it.)

Verbal descriptions are rarely used in the drawing of police sketches. Few observers, unless they are specially trained, can give satisfactory clues to appearance in words. Most can point to features similar to those they remember, however, and that is how the reconstruction artist usually begins. Our initial experiments were intended to test the effectiveness of this procedure and to gain some preliminary notions of what features are considered important in describing or recognizing a face.

Frontal-view photographs were shown to an experienced artist, who compiled a written description of each face; the description included references to facial features in a catalogue of faces made up of photographs of various head shapes, eye spacings, lip thicknesses and so on, organized by feature type. Thus a large part of the description consisted of "pointing to" similar features on other portraits. The completed description was given to another artist, whose task was to reconstruct the face from the written description [*see illustration on next page*].

The first attempt, although obviously resembling the original photograph, differed from it in the depiction of important features and proportions. When limited feedback was allowed, however, there was rapid improvement. The describing artist, with the initial sketch in hand, provided simple verbal corrections, such as "The hair should be bushier at the temples"; with this information the reconstructing artist was able to draw a much more accurate likeness. Finally, to find the limit of improvement, that is, to discover just how faithful a portrait could be drawn, the reconstructing artist was given the photograph to work from. Under those conditions he was able to produce a strikingly realistic representation. Some sketches, in fact, were judged to look more like the per-

LEONARDO'S "MONA LISA," rendered as a "block portrait," consists of 560 squares, each of which is uniform in color and brightness. The transformation of the familiar painting was accomplished in the same way as that of the portrait of George Washington on the cover of this issue of SCIENTIFIC AMERICAN. Recognition can be enhanced by rapidly moving the page, by squinting at the image or by viewing it from a distance of 10 feet or more.

son than the photograph did. Presumably the artist enhanced recognition by in some way emphasizing significant detail.

All the sketches were shown to test subjects who, as fellow employees, had seen the "suspect" often. Almost half of the sketches drawn from descriptions were correctly identified and about 93 percent of the drawings made directly from photographs were recognized.

Our work with face-reconstruction artists was a pilot experiment we hoped would lead, through informal observation, to a better understanding of the problems confronted in the recognition of faces and to the formulation of further experiments. Some of the incidental information derived from the study was indeed interesting. For example, we found that several of the faces were outstandingly easy to recognize in the sketches. Presumably those subjects were more easily described than the others, or perhaps they possessed certain features that are conspicuous or rare. Several subjects remarked that the nose and eyes in one sketch were important to identification, yet for the same face other subjects observed that although the nose, mouth and hair were well drawn, the eyes were not and did not aid recognition.

Another way to study recognition is to ask how little information, in the informal sense of "bits," or binary digits, is required to pictorially represent a face so that it can be recognized out of a finite ensemble of faces. We explored this "threshold" of recognition with portraits that had been precisely blurred.

The type of blurring commonly encountered in photographs is caused by an improperly focused optical system; it reduces the information content of the picture, but it proved unsuitable as a technique in our investigations because the degree of blurring cannot be precisely specified or controlled. A more measurable method degrades the image in quantifiable steps through a relatively simple computer process.

In our experiments a 35-millimeter transparency of a conventional portrait photograph is scanned by a beam of light moving in a raster pattern of 1,024 lines. The variations in the intensity of the beam caused by the varying transparency of the film are detected by a photomultiplier tube. The analogue signals produced by the photomultiplier are converted into digital form by sampling each line in the raster at 1,024 points and assigning a brightness value to each point, so that the completed image consists of $1{,}024^2$ (or 2^{20}) discrete points, about four times the resolution of the commercial television image. Each of the points may have 1,024 brightness values, or tones of gray. The dissected image is stored in the magnetic-tape memory of a digital computer.

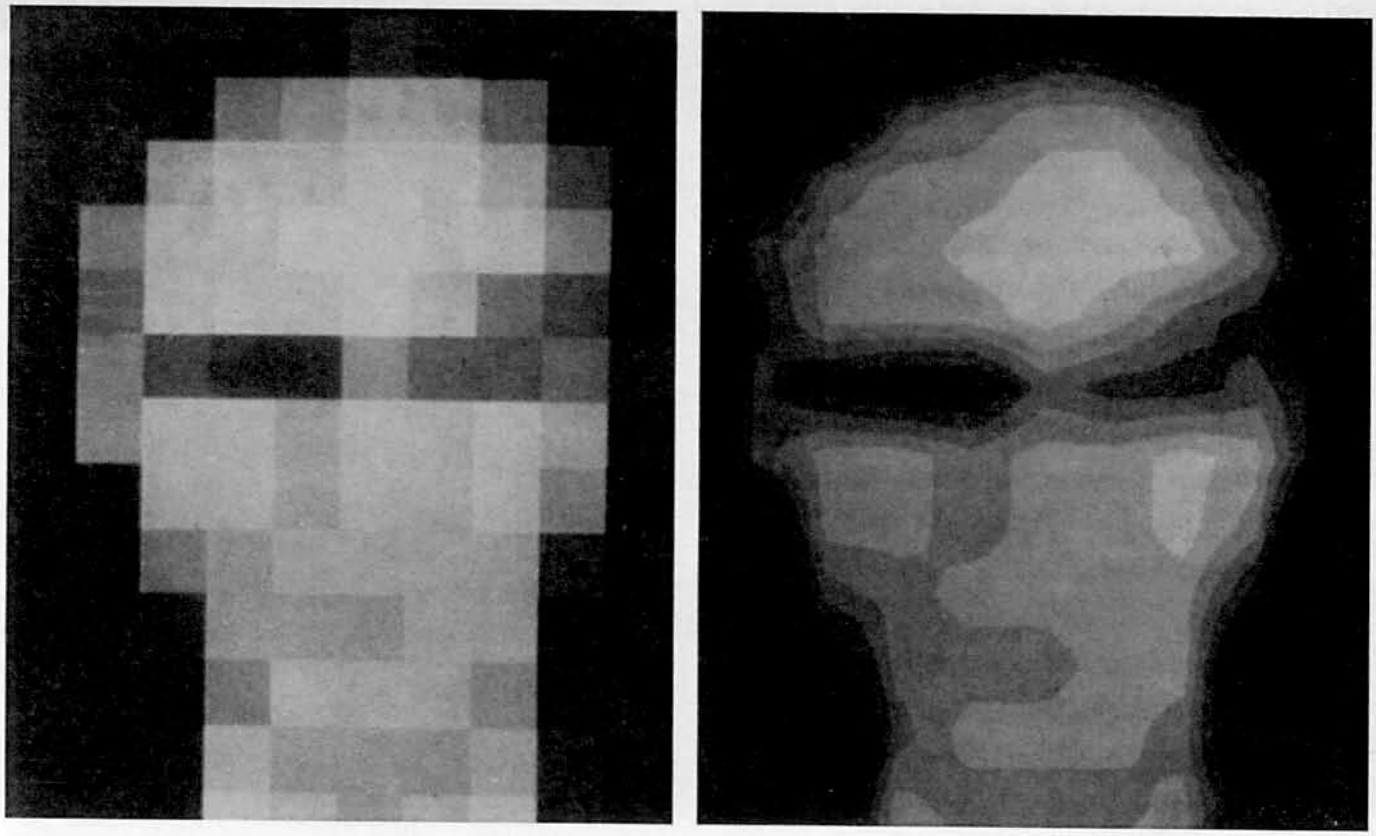
To create the degraded image the computer divides the picture into $n \times n$ squares of uniform size and averages the brightness values of all the points within each square. For example, if a photograph is to be made into an array of 16×16 squares, each square will contain 64×64, or 4,096, points; the brightness to be assigned to the entire square will be found by averaging the values of these points. In a final step the number of brightness values is reduced to eight or 16 by assigning to each square the gray tone closest to its original averaged value.

The computer stores the digital information comprising the picture on magnetic tape and the tape controls a cathode-ray-tube monitor, which then displays the completed portrait. A photograph of this display constitutes the finished product. Alternatively, the magnetic tape can be used to control a facsimile printer that produces a print of the processed image without the intermediary cathode ray tube [*see bottom illustration on opposite page*].

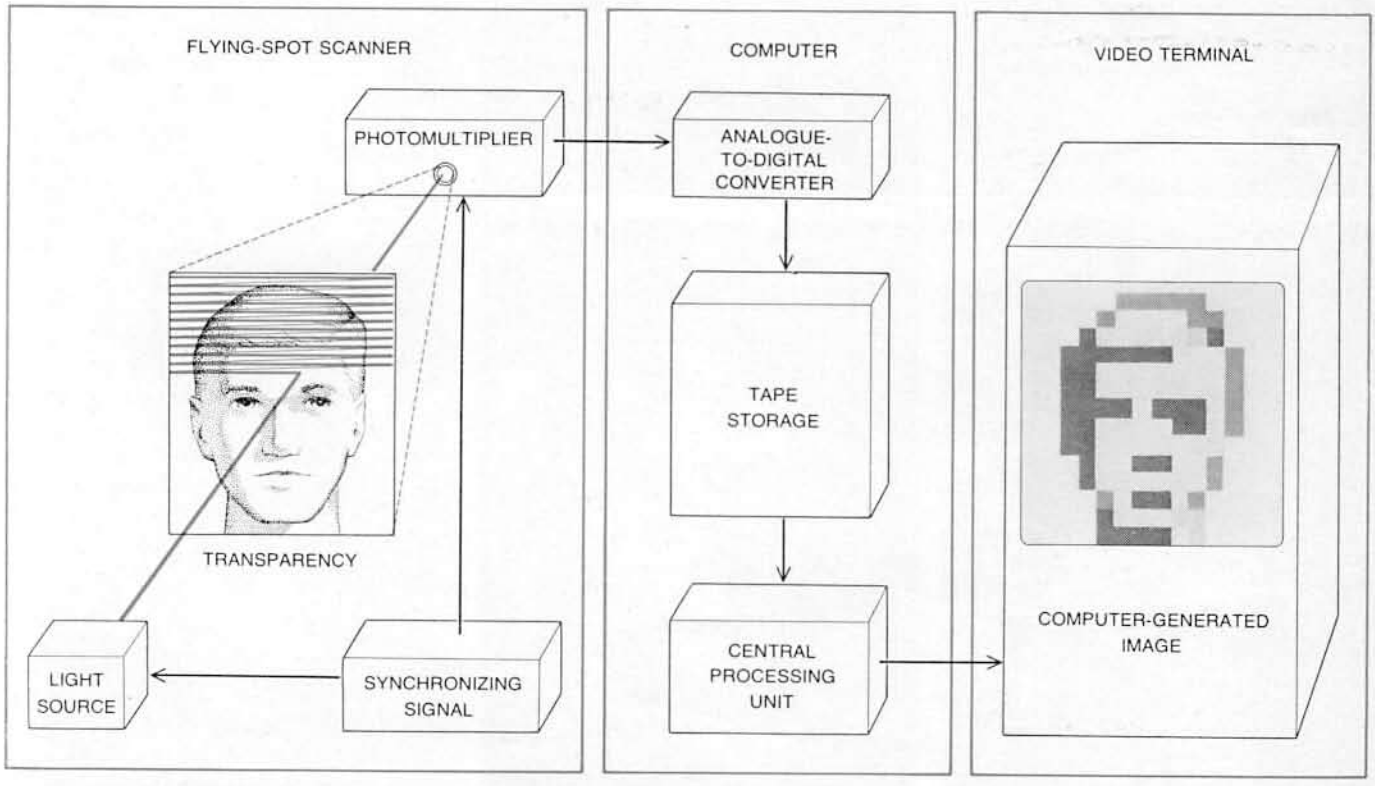
Viewed from close up, these "block portraits" appear to be merely an assemblage of squares. Viewed remotely, from a distance of 30 to 40 picture diameters, faces are perceived and recognized.

Preliminary experiments were made to select the coarsest image that might be expected to yield about 50 percent accuracy of recognition. For some kinds of picture, resolution of only a few thou-

SKETCHES FROM DESCRIPTIONS were made by a "face-reconstruction artist" skilled in drawing portraits from information provided by witnesses. At top left is the photograph from which the three sketches are derived. For the first drawing (*top right*) a written description of the face, including references to illustrations in a catalogue of facial features, was presented to the artist. A better likeness was produced (*bottom left*) when simple verbal corrections were provided. For the final version (*bottom right*) the artist was given the photograph; the resulting portrait represents the limit of accuracy of the process.



REDUCED-INFORMATION-CONTENT PORTRAITS were generated by a computer. The picture at left is a block portrait; it is an array of 16 × 16 squares, each one of which can assume any one of 16 levels of gray. Not all the 256 squares are required to represent the face. The contoured representation at right was produced by filtering the block portrait to remove high frequencies.



SYSTEM FOR MAKING BLOCK PORTRAITS uses a flying-spot scanner, a device similar to a television camera. The image, usually in the form of a 35-millimeter photographic transparency, is scanned in a raster pattern of 1,024 lines. In the analogue-to-digital converter each line is sampled at 1,024 points and the brightness of each point is assigned one of 1,024 values. Using this information stored on magnetic tape, the central processing unit divides the image into $n \times n$ squares and averages the brightness values of all the points within each square. The number of permissible brightness values is then reduced to eight or 16. The resulting image is displayed on a video terminal (a television screen) and photographed. The computer can also be made to operate a facsimile printer, which produces a finished picture directly. Most of the portraits used in these experiments were made by the latter process.

sand elements provides acceptable quality; the limits of recognition for photographs of faces, however, have not been reported. Our informal investigation revealed that a spatial resolution of 16×16 squares was very close to the minimum resolution that allows identification.

Tests were also made to determine the useful limits of gray-scale representation. The relation between gray-scale and spatial resolution is an interesting one: either factor can serve as a limit to recognition. It was not the object of our experiments to document this relation, however, and so only a few gray-scale tests were made once the 16×16 spatial pattern was decided on. For 16×16 portraits gray scales of either eight or 16 levels yielded eminently recognizable portraits; consequently our experiments used those levels exclusively. (The allowed gray levels can be expressed in terms of bits. A gray scale of eight levels requires three bits of information; a scale of 16 levels calls for four bits.)

Fourteen of the block portraits were shown to 28 subjects. Each subject was given a list of 28 names, including the names of the 14 persons depicted. The experiment was intended to investigate the effects of changing the gray scale from a three-bit to a four-bit one, as well as to test identification performance.

Overall recognition accuracy was found to be 48 percent. (Random guessing would produce such a result only four times in a million trials.) The result was essentially indifferent to the resolution of the gray scale. Thus the number of bits required for approximately 50 percent accuracy of recognition was no more than 16×16 squares times three bits, or 768 bits. None of the portraits, however, filled all the squares in the 16×16 grid; therefore fewer than 256 squares made up each face. An average of 108 squares was needed.

Recognition of particular faces ranged from 10 percent to 96 percent. In these experiments too some faces were always easy to identify, although, as will be seen, the reasons are peculiar to the conditions of the experiment. Two portraits received outstanding recognition and four were rarely identified correctly.

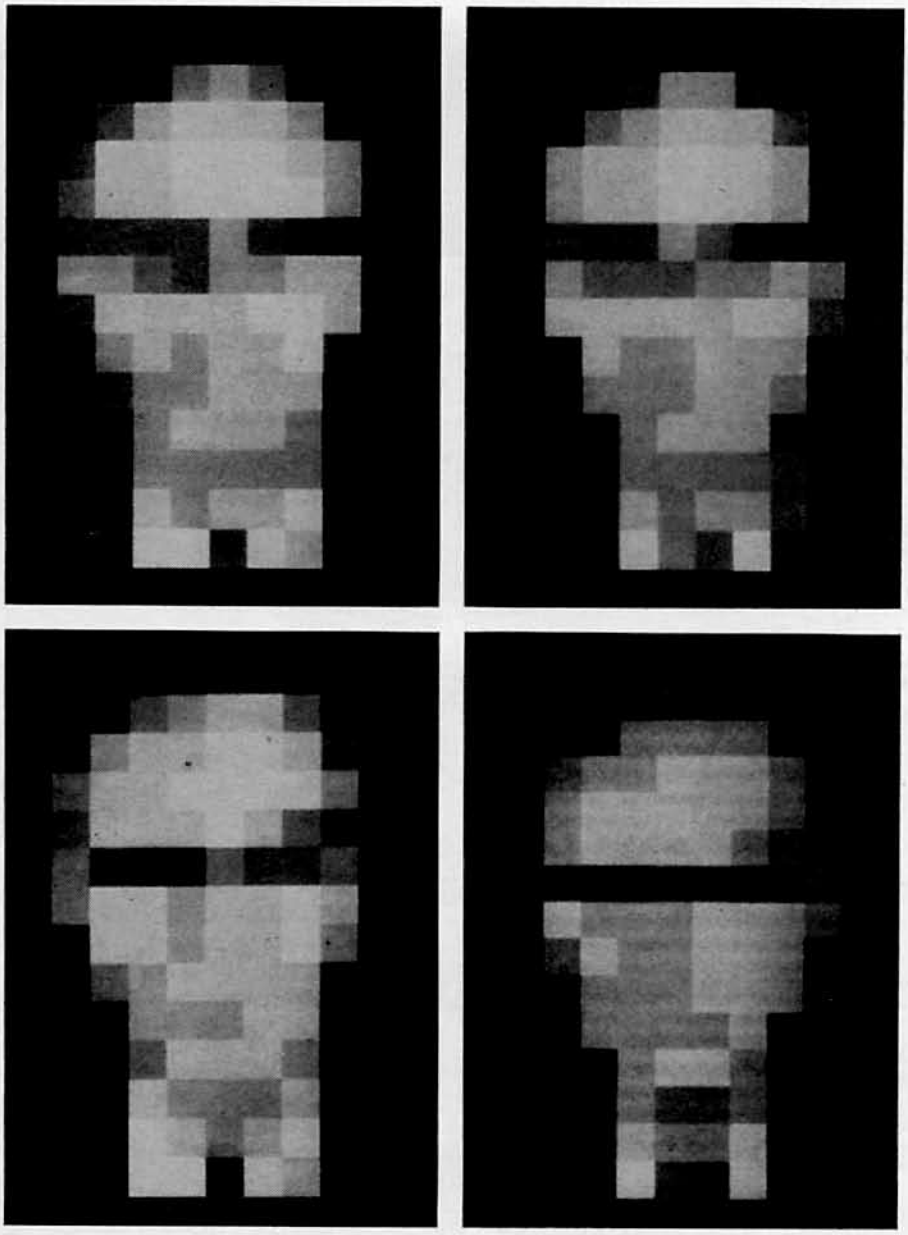
Two possible explanations of these disparities were suggested. First, some faces, because of the peculiar arrangement of their features, respond notably well or particularly poorly to coarse spatial presentation. Second, the grid, arbitrarily positioned over a given face by the scanning process, may land luckily or unluckily for adequate representation. For example, a square might just bracket an eye, or it might land half on and half off. The latter possibility was judged to be the more likely. I hypothesized that those pictures that were recognized well probably had a fortuitously placed grid.

To test the hypothesis each portrait was reprocessed by shifting the 16×16 matrix with respect to the original block portrait. Three new pictures were made: one shifted a half-square to the right, one a half-square down and a third a half-square to the right and down [*see illustration at left*].

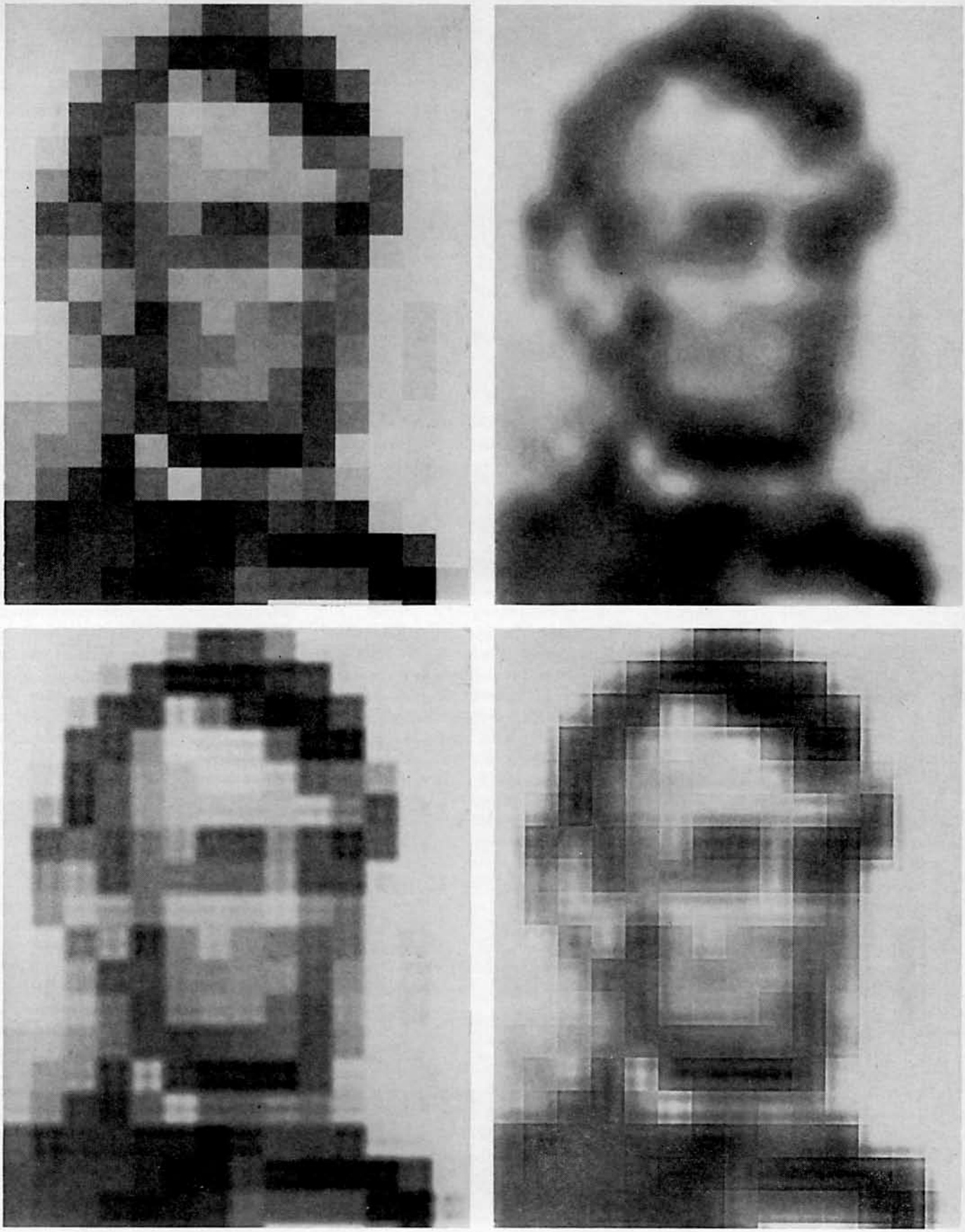
Recognition of the sets of four shifted pictures was tested. The subjects were given the identity of each photograph; their task was to rank the four portraits in each set in order of pictorial accuracy. My hypothesis predicted that in these tests those pictures that were readily identified in the earlier experiment would be ranked first in their set and that those scoring worst initially would be ranked near the bottom. So it turned out; both correlations were confirmed.

This result led us to believe that if the best grid positions had been found and used in the earlier experiments, the average accuracy of recognition might have been closer to 100 percent than to 50 percent. A new experiment confirmed this: performance rose to 95 percent.

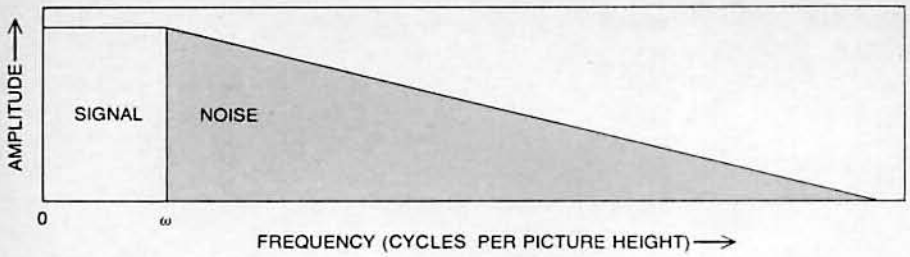
An interesting and provocative characteristic of block portraits is that once recognition is achieved more apparent



EFFECTS OF GRID PLACEMENT on recognition are illustrated by four block portraits of the same face. The original is at top left. Alternative versions were made by shifting the grid placement one half-block to the right (*top right*), one half-block down (*bottom left*) and one half-block right and down (*bottom right*). When portraits made with optimum placement replaced those made with random placement, recognition accuracy doubled.



SELECTIVE FREQUENCY FILTERING influences the ease with which block portraits are recognized. The original block portrait of Abraham Lincoln is at top left. It consists of the photographic "signal," whose highest spatial frequency is 10 cycles per picture height, and noise frequencies extending above 10 cycles. As was anticipated, filtering out all spatial frequencies above 10 cycles (*top right*) greatly enhances recognition. Selective removal of only part of the noise spectrum, however, reveals which frequencies most effectively mask the image. At bottom left all frequencies above 40 cycles have been removed; even though the sharp edges of the squares are eliminated, perception is improved only slightly. When the two-octave band from 10 to 40 cycles is removed (*bottom right*), the face is more readily recognized. The phenomenon apparently responsible for this effect is critical-band masking.



BLOCK-PORTRAIT SPECTRUM consists of a signal that extends to some finite spatial frequency ω, corresponding to the block-sampling frequency and noise of frequencies above ω. The amplitude of the noise typically decreases with increasing spatial frequency.

detail is noticed. It is as though the mind's eye superposes additional detail on the coarse optical image. Moreover, once a face is perceived it becomes difficult not to see it, as if some kind of perceptual hysteresis prevented the image from once again dissolving into an abstract pattern of squares. The observation that is most intriguing, however, is that recognition can be enhanced by viewing the picture from a distance, by squinting at it, by jiggling it or by moving the head while looking at it. The effect of all these actions is to blur the already degraded image.

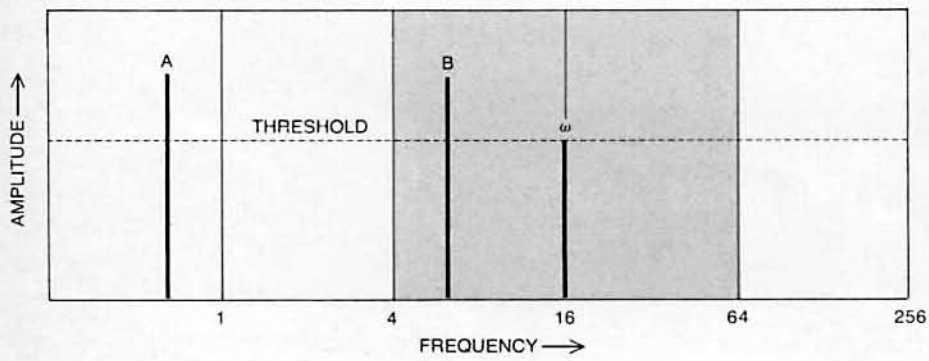
Why should recognition be improved by blurring? The explanation almost certainly lies in the "noise" that tends to obscure the image.

A picture, like a sound, can be described as the sum of simple component frequencies. In acoustical signals pressure varies with time; in the optical signals discussed here the frequencies are spatial and consist of variations of "density" (or darkness) with distance. Just as a musical note consists of a fundamental frequency and its harmonics, so an optical image consists of combinations of single frequencies, which make up its spatial spectrum. The spectral representation exists in two dimensions. This spectrum refers only to spatial frequencies; the color spectrum describes another aspect of the image.

When pictures are considered combinations of spatial frequencies, they can be manipulated in the same ways as other frequency-dependent signals are. For example, Fourier analysis can be used to determine the component frequencies of an image, or low-pass filtering can be used to remove the high frequencies that represent fine detail. Signal-frequency bands, the signal and noise spectrum and other terms usually associated with discussions of acoustical phenomena can be applied to the processing of visual images.

The description of a two-dimensional image as a signal of various spatial frequencies leads to a possible explanation of the enhancement of block portraits with blurring. Whenever a signal with a spectrum running from zero to some frequency designated ω is reduced by sampling to discrete frequency components, noise artifacts whose spectrum extends above ω are introduced. The noise is a product of the sampling procedure. In two-dimensional signals it appears as patterns not present in the original image.

Because the noise in these pictures is ordinarily of higher frequency than the signal it can be readily eliminated by a low-pass filter, that is, a filter that preserves only the low frequencies, eliminating the high frequencies that represent fine detail. This operation too is performed by the computer; all spectral components above ω are removed while the desired signal is retained [*see top illustration on this page*].



CRITICAL-BAND MASKING is known to occur in presentations of simple visual or audio signals, such as single sinusoidal waves. The test signal ω, at the threshold of perception, would be masked by signal *B*, within the band, but not by signal *A*. The critical band (*colored area*) extends for about two octaves above and below the test frequency. The author's investigations indicate that critical-band masking also affects two-dimensional signals.

In block portraits the most obvious noise is that introduced by the sharp edges of the squares. Although Fourier analysis shows that the energy content of these high frequencies is relatively small, one might speculate that because the eye is particularly sensitive to straight lines and regular geometric shapes such square-patterned noise masks particularly well. That is, such image-correlated noise might mask more effectively than randomly distributed noise of equal energy. If so, low-pass filtering should enhance perception. This explanation would seem to be confirmed by the fact that recognition is improved by progressive defocusing or distant viewing, since the effect of both of these actions is to filter out high frequencies.

This hypothesis, however, is not the only candidate; another possibility is called critical-band masking. In both hearing and vision the spectral proximity of noise to a signal drastically influences the detection threshold of the signal. For example, the threshold for detecting a single sinusoidal wave anywhere in the spectrum is elevated when a noise signal is introduced if the noise lies within about two octaves of the signal. If the noise lies outside this "critical band," masking does not occur [*see bottom illustration on this page*].

This phenomenon has been tested and confirmed by others for relatively simple visual presentations such as sine-wave and square-wave gratings in a single dimension. My colleague Bela Julesz and I reasoned that similar masking might occur in more complicated two-dimensional patterns.

If critical-band masking is the mechanism that hinders the recognition of block portraits, then those components of the noise that fall within about two octaves of the sampling frequency ω would be primarily responsible. The rest of the noise spectrum, including the high-frequency signals contributing to the sharp edges of the blocks, should cause little or no masking.

To resolve this question we prepared a series of block portraits that were spectrally manipulated by the computer. The original image was transformed to obtain its Fourier spectrum, filtered to specification, then transformed back and

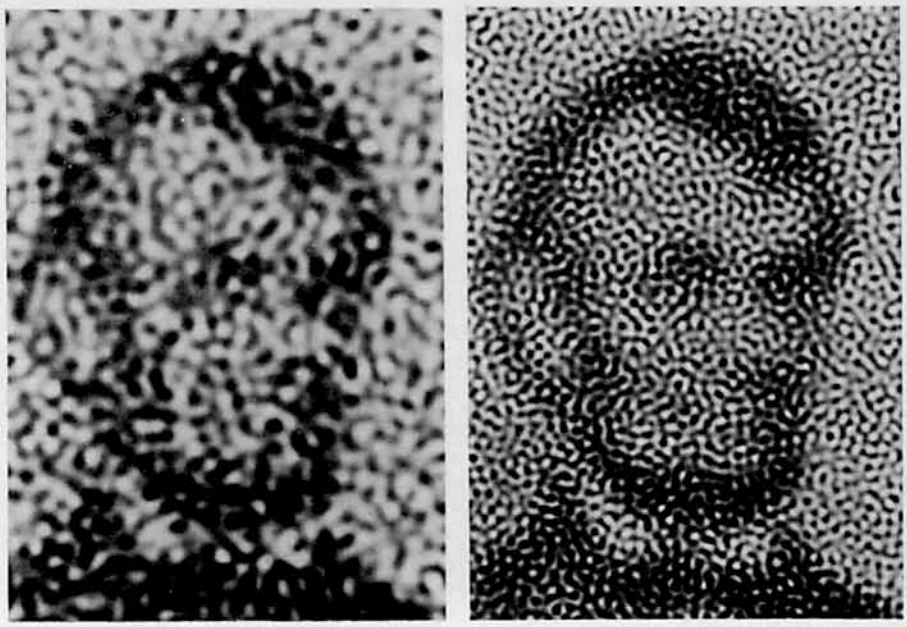
printed out. This technique provides precise control of spatial frequencies. We were able to remove all signals above a specified frequency, or to remove only a band of frequencies adjacent to ω.

In our first attempt to evaluate the relative importance of high-frequency and critical-band noise masking we prepared a series of filtered block portraits [*see illustration on page 75*]. The result looked promising: removal of the very high frequencies did little to change the block aspect, and some effort was still required to perceive the face. Removal of only the frequencies adjacent to the signal produced pictures that were much closer in appearance to the original photographs. That is, the very high frequencies did not seem to be the most important in masking.

Although the results of this experiment suggest that noise spectrally adjacent to the signal is most effective in masking recognition, the point is not proved. There are three reasons why the experiment is not conclusive. First, the noise generated by the block-sampling process is spatially periodic, at the block frequency and at higher harmonics. Second, the noise amplitudes are correlated with picture information: the magnitude of the noise in any block depends on the density of the image in that block. Finally, the energy of the noise spectrum is greatest at the block-sampling frequency, and it decreases with increasing frequency. Hence the adjacent band noise may mask more effectively simply because its amplitude is higher, not because of the critical-band effect.

There is a straightforward way to avoid these difficulties. We can simply add random noise of the proper frequency to a picture that is smoothly blurred rather than block-sampled. We added random noise of constant spectral energy to a portrait that had been low-pass filtered to the same bandwidth as that used in the first recognition studies. When such a picture containing adjacent-frequency noise is compared with one masked by remote-frequency noise, the result is unequivocal: critical-band masking is responsible for the suppression of recognition [*see illustration on this page*].

The discovery that critical-band masking affects complex pictures as well as simple sinusoidal presentations raises additional questions. How effective in masking are noises of equal energy and bandwidth but of various spectral shapes? When noise is added to a signal, is the shape of the noise signal or its location in the spectrum more important?



RANDOMLY DISTRIBUTED NOISE of uniform amplitude is added to smoothly blurred portraits of Lincoln. When the noise is in the band adjacent to the signal frequencies (*left*), it obscures the picture more effectively than when it is at least two octaves removed from the picture frequencies (*right*), confirming that critical-band masking is the most important mechanism limiting the recognition of degraded or blurred images such as block portraits.

What are the relative effects of spatial disposition and spectral disposition? That is, if equal amounts of noise energy are added to visual scenes, is the placement with respect to position or with respect to frequency more important for masking? These and related questions remain for future investigation. Their answers will provide new insights into the psychophysics of vision.

A more conventional means of blurring pictures is continuous smearing. In optical systems one can simply project the image out of focus; as I have noted, however, this operation cannot be precisely controlled. The analogous operation performed by a digital computer is intrinsically discrete, but by using sufficiently numerous sample points to represent an image, blurring can be made arbitrarily smooth and extremely precise.

Pictures made up of 256×256 elements (about one-fourth the resolution of television) produce fairly sharp portraits. Such pictures can be blurred by selecting for each point a brightness value computed by averaging the brightness of the points surrounding it. An "averaging window" of $n \times n$ points is used to compute a new value for each point in the 256×256 array; after a new point is written the window is moved over one element and a new average is made.

Through this process the computer can rapidly and accurately blur a picture to a specified degree. The size and shape of the averaging window and the relative weight given to each element in the array can be selected at will. For example, the average could be uniformly weighted or computed on a Gaussian, or bell-shaped, curve. In our experiments we used a square window of varying size with uniform weighting; each element contributed equally to the average value assigned the new point.

We have used portraits made in this way to study the limits of face recognition. Fourteen portraits were shown to subjects who were given a list of 28 names, including the names of the 14 "target" individuals. All 28 persons were known to the test subjects. Several degrees of blurring were tested [*see illustration on next page*]. Some subjects were shown the most blurred pictures first, some the least blurred, and so on, in order to simultaneously test for the effects of learning. The experiments were conducted by Ann B. Lesk, John Levinson and me.

Contrary to what we had expected, recognition scores were quite good. For photographs blurred by a 27×27-point averaging window the recognition was 84 percent. As the degree of blurring increased, the scores declined to about 65 percent for those portraits made with a 43×43-point window, which represents severe blurring. (The expected score for random guessing is 3.5 percent.)

Even more surprising were the results

of trials with photographs blurred with a 51 × 51-point window. Here the width of the window is 20 percent of the picture's width, and blurring is so extreme that facial features are entirely washed out. Nevertheless, the accuracy was almost 60 percent. (This level of recognition cannot continue with much more extensive blurring. When the averaging window includes all the picture elements, the field will be smeared to a uniform gray and pictures will differ only in the level of that gray.)

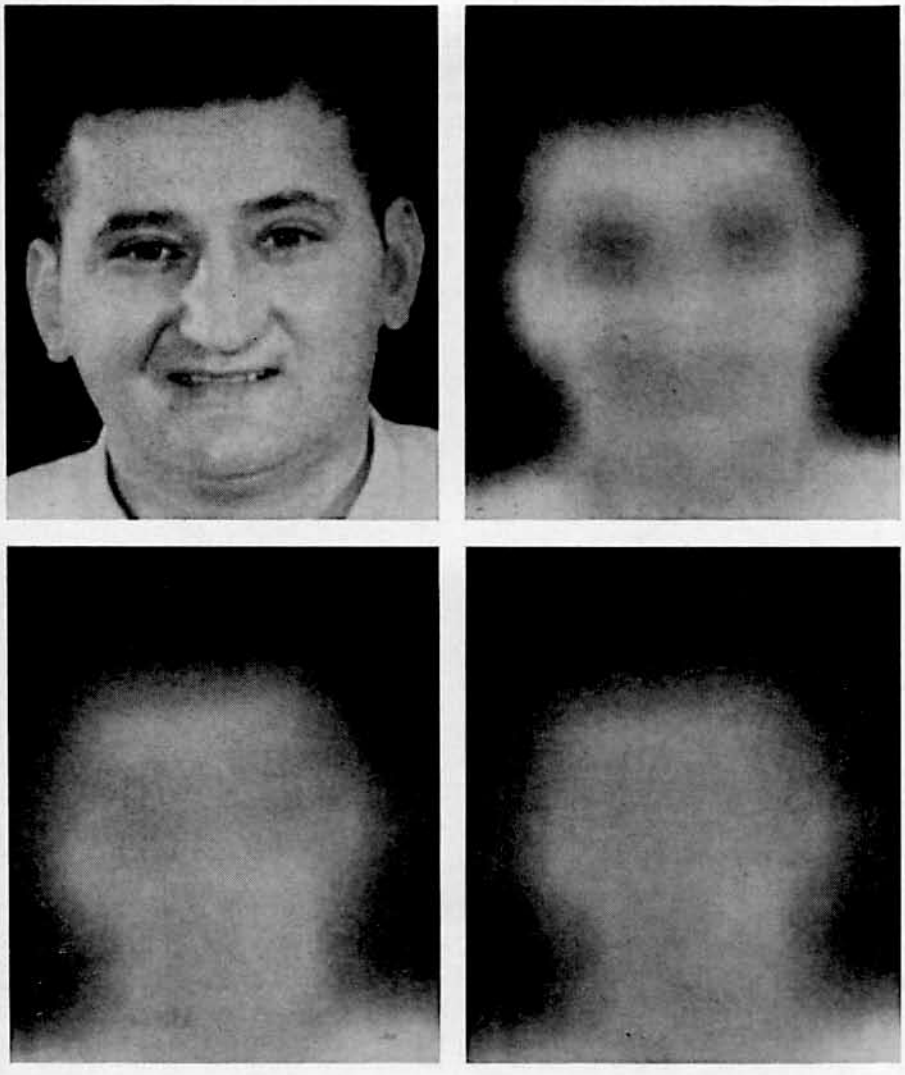
Recognition of the most strongly blurred of these portraits cannot depend on the identification of features. The high-frequency information required to represent the eyes, the ears and the mouth is lost. Although some intermediate frequencies remain, their representation of the chin, the cheeks and the hair is not clear. The low-frequency information that relates to head shape, neck-and-shoulder geometry and gross hairline is all that remains unimpaired, yet this alone seems to be adequate for rather good recognition among individuals in a restricted population.

Again, some faces were consistently well recognized. This time the responsible cues were easy to see. One portrait, for example, was distinguished by a round, bald head, and the picture was consistently recognized, even when it was badly blurred.

Some learning apparently took place in these experiments; it would appear that practice at struggling with the task improved performance.

Determining exactly how one recognizes a face is probably an intractable problem for the present. It is possible, however, to determine how well and with what cues identification can be achieved. Similarly, although machine recognition is not yet possible, search for and retrieval of faces by machine is a problem suitable for research. My colleagues and I approached these matters by investigating how effectively one can identify an individual face from a group of faces by using verbal descriptions.

It should be noted that successful identification of faces by feature descriptions does not suggest that the normal processes of recognition regularly detect and assess such features. All we can determine from experiments of this kind is how effectively people can perform a certain recognition task on the basis of certain assigned measures.



PRECISELY BLURRED PORTRAITS were constructed by a computer using an "averaging window." At top left is the original picture; it is not a continuous-tone photograph but an array of 256 × 256 dots. The averaging window determines a new value for each of the dots by averaging the values of those that surround it in some $n \times n$ field. When the window is set at 27 × 27 points (*top right*), basic facial features are still discernible. A 43 × 43-point window (*bottom left*) produces severe blurring and a 51 × 51-point window (*bottom right*) eliminates almost all information except gross forms. Accuracy of identification declined as blurring increased but even with the worst pictures approached 60 percent.

The problems of the automatic analysis of faces have received little attention. The work begun by W. W. Bledsoe and his colleagues is one of the few attempts I know of to automate the recognition of faces; the method uses a hybrid man-machine system in which a computer sorts and classifies faces on the basis of fiducial marks entered manually on photographs. The technique is called the Bertillon method, after Alphonse Bertillon, a French criminologist, and is better known for its application to fingerprint classification. A similar method has been developed by Makoto Nagao and his colleagues in Japan in an attempt to devise an automated system that would produce simple numerical descriptions of faces.

I was led to this line of inquiry by wondering if one could play a "20 questions" game with faces. (In games of this kind one player thinks of a person and the other asks him up to 20 questions, which must be answered yes or no, until the subject is guessed.) An informal, preliminary experiment began with 22 portraits; they were shown to subjects who were asked to list features they thought striking or extreme in order of decreasing extremeness. If a face displayed very wide-set eyes, for example, that statement was put at the top of the list. Or if the chin jutted extremely, that fact was listed first. A consensus list was compiled for each of the 22 faces, then new experimental subjects were selected.

Each subject was given the pile of pictures and a list of features, derived from the earlier work, describing one of the faces. He was asked to do a binary sorting, one feature at a time, starting with the most extreme and working down the