Second-Generation Image-Coding Techniques

MURAT KUNT, SENIOR MEMBER, IEEE, ATHANASSIOS IKONOMOPOULOS, AND MICHEL KOCHER

Invited Paper

The digital representation of an image requires a very large number of bits. The goal of image coding is to reduce this number, as much as possible, and reconstruct a faithful duplicate of the original picture. Early efforts in image coding, solely guided by information theory, led to a plethora of methods. The compression ratio, starting at 1 with the first digital picture in the early 1960s, reached a saturation level around 10:1 a couple of years ago. This certainly does not mean that the upper bound given by the entropy of the source has also been reached. First, this entropy is not known and depends heavily on the model used for the source, i.e., the digital image. Second, the information theory does not take into account what the human eye sees and how it sees.

Recent progress in the study of the brain mechanism of vision has opened new vistas in picture coding. Directional sensitivity of the neurones in the visual pathway combined with the separate processing of contours and textures has led to a new class of coding methods capable of achieving compression ratios as high as 70:1. Image quality, of course, remains as an important problem to be investigated. This class of methods, that we call second generation, is the subject of this paper. Two groups can be formed in this class: methods using local operators and combining their output in a suitable way and methods using contour–texture descriptions. Four methods, two in each class, are described in detail. They are applied to the same set of original pictures to allow a fair comparison of the quality in the decoded pictures. If more effort is devoted to this subject, a compression ratio of 100:1 is within reach.

I. Introduction

Every image acquisition system, be it high-resolution microdensitometer or TV camera, produces pictorial data by sampling in space and quantizing in brightness analog scenes. The sampling step size is usually chosen small enough to avoid interpolation before display and relies on the integration ability of the human visual system. A digital image is thus an N by N array of integer numbers or picture elements (pels) requiring N^2B bits for its representation where B is the number of bits per pel. This array is commonly referred to as the canonical form of a digitized picture, because it is assumed that the two-dimensional sampling theorem is respected. Generally, the canonical form requires a very large number of bits for its representation. For example, with a 512 by 512 raster and 8 bits per pel, 2×10^6 bits are needed, a rather large number! This paper is confined to grey-level still pictures.

A. The Problem

The goal of image coding is to reduce (to compress), as much as possible, the number of bits necessary to represent and reconstruct a faithful duplicate of the original picture. This is a very reasonable goal for two related reasons.

First, any data originated from what is called an image, are not random. Adjacent samples have similar grey values, exhibiting thus an important spatial correlation. If this correlation is exploited in an appropriate way, there is no doubt that the number of N^2B bits can be reduced. We call "first generation" the set of techniques based on this classical view of the image coding problem, although at least one of them (see Section II) was in advance for its time.

The second reason extends the first one and is as follows. Consider the canonical form of a natural image such as a portrait, a scene with trees and houses, or a building, etc., and let us ask the following question: Can we find a representation for such an image worse than its canonical form? (worse in the sense of not efficient or economical). Probably the answer is no. Following the famous sentence by Descartes stating that a picture is worth a thousand words, about 20 bits should be enough to represent it. Unfortunately, like Shannon, he did not tell us how to do it! The summary of this is that data compression is not only possible but has to be done by several orders of magnitude. How this can be achieved when the limits of compression have been reached within the framework of information theory and coding theory (see Fig. 1)? By simply going out of this framework. The reason for this is that often forgotten assumptions of ergodicity and stationarity are not satisfied for image data. The entropy of the image data source is not known and depends heavily on the model used. We call "second generation" the ensemble of techniques capable

Manuscript received July 29, 1984; revised November 2, 1984.

M. Kunt and A. Ikonomopoulos are with the Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, CH-1007 Lausanne, Switzerland.

M. Kocher was with AT & T Bell Laboratories, Holmdel, NJ 07733, USA. He is now with CERAC Institute, Ecublens, Switzerland.

0018-9219/85/0400-0549$01.00 ©1985 IEEE

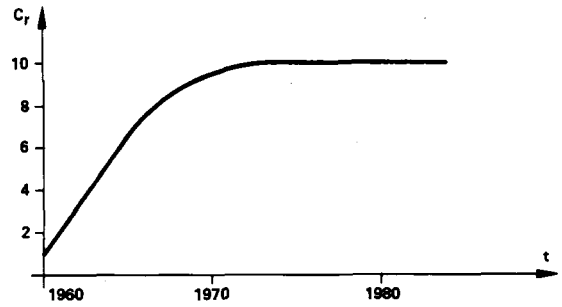


Fig. 1. Average compression ratio (sketch) of the first-generation image coding techniques.

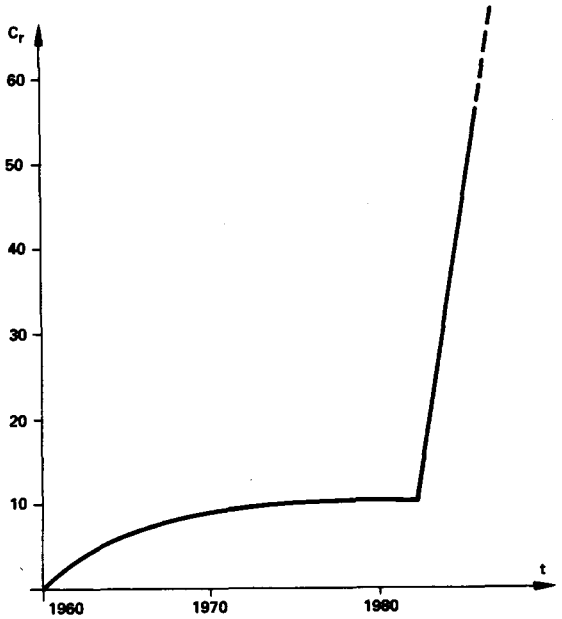


Fig. 2. Average compression ratio (sketch) of both generations.

of achieving high compressions beyond this saturation level (Fig. 2). The purpose of this paper is to report on the recent attempts made in this direction.

Another view of the difference between the techniques of the first and the second generation is as follows. Image coding is basically carried out in two steps: first, image data are converted into a sequence of messages and, second, code words are assigned to the messages. Methods of the first generation put the emphasis on the second step, whereas methods of the second generation put it on the first step and use available results for the second step.

B. A Way to a Possible Solution

The very end of almost every image-processing system is the human eye. Although our visual system is by far the best image-processing system one can think of, it is also far from being perfect. So, if the coding scheme is matched to the human visual system and attempts to imitate its functions, at least for the known part of it, high compressions can be expected. A few methods designed with these aims give encouraging compressions as high as 70:1 as shown in Fig. 2. Therefore, it does not seem unreasonable to preprocess the canonical form in this way before applying the classical coding theory to achieve high compression.

What are needed now are guidelines for preprocessing. An image can be described in terms of several possible entities such as pels of the canonical form, a group of pels in small blocks, Fourier or other transform coefficients, linearly predicted values, or derivatives. With the continuous progress in visual pattern recognition and scene analysis, another possibility is to describe an image in terms of contour and texture [1].

Contours are commonly referred to as abrupt changes in grey level [1], [2]. According to the abruptness, refinements such as strong or weak contours can be used. Texture, on the other hand, became a great center of interest in Pattern Recognition [3]–[6]. Although a large number of studies have been done on purely textured images [7], a clear-cut definition of texture does not exist. This is mainly because, in this context, texture is related to human visual perception and has a subjective component as well. The most suitable definition [8] relies on the shift invariance and shows the small amount of semantic information contained in it. In the context of coding, texture will be defined as anything else in images, beside contours.

The suggested way to a possible solution for high-compression image coding is therefore to represent an image in terms of textured regions surrounded by contours in such a way that the regions correspond, as faithfully as possible, to the objects of the scene. Efficient codes exist for contour coding. Because of the low information content, textured areas can also be coded efficiently.

C. Organization of the Paper

In Section II a brief overview of first-generation coding techniques is given. Its purpose is not to replace excellent well-known reviews by Netravali and Limb [9], by Jain [10], or the most recent by Musmann contained in this issue, but to summarize the tremendous amount of effort devoted to this particular problem over the last two decades. One particular technique that we consider as advanced for its time is described in some detail because of the insight it gave for the development of the second-generation techniques, providing thus a basis for their introduction.

Section III is devoted to the mechanism of vision and to the translation, into engineering language, of recent results from neurophysiology. Properties of the human visual system which can be used in image processing and coding are described. Successful applications, other than coding, are indicated.

Sections IV and V describe four techniques which can be considered as second generation. The first two are based on local operators whereas the last two use contour–texture models. Results showing decoded pictures are discussed in Section VI. Unless otherwise stated, all the pictures used in this paper are digitized with a 256 by 256 raster and quantized to 256 levels, thereby permitting an 8-bit representation. Output pictures are reproduced on a laser-receiver using dry silver paper with an optical density of 1.6.

II. Brief Review of First-Generation Coding Techniques

A. Preliminary Remarks

A digitized picture can be characterized by a sequence of messages. There are many ways of selecting the messages. The only requirement is to be able to reconstruct a faithful duplicate of the original picture from a sequence of mes-

sages. A source-coding technique is then applied to this sequence of messages to reduce its redundancy. The particular way of selecting the messages and assigning code words to them becomes a specific image-coding method. For example, the messages can be the brightness level of each pel (canonical form), or a group of pels, or values of a function computed from a group of pels. In this section, major image-coding techniques that we classed as "first generation" will be listed. For more detail, the reader may refer to [9], [10] or to the companion paper by Musmann in this issue.

A first classification of the coding methods can be made as information-lossless or information-lossy techniques. The former are able to reconstruct the original picture exactly, whereas the latter introduce some distortions which should be kept as unnoticeable as possible.

A second classification can be based on the space where the method is applied. For example a method combining the values of pels in an appropriate way is labeled as spatial method. In contrast, a method which uses a set of transform coefficients is called transform method. Finally, methods which are applied in both the spatial and the transform domain are referred to as hybrid methods.

Another classification is often made as fixed or adaptive methods in the sense that the parameters used are fixed or change as a function of the local data in the image.

B. Major Techniques of the First Generation

Spatial methods:

1) *Pulse-Code Modulation (PCM):* Acceptable quality pictures are obtained with 3 bits per pel. The compression ratio is then $C = 2.6$ to 1. Dithering, as introduced by Roberts [11], may be used to improve the quality. The properties of the human visual system are not used at all.

2) *Predictive Coding:* Predictive methods do not normally make use of the properties of the human visual system. With respect to the 8-bit canonical form, the compression ratio obtained by two-dimensional prediction is around 4 to 1. They can be made adaptive if the prediction parameters are adapted to the data in an appropriate way. For example, a local activity measure can be defined and prediction parameters can be updated at each noticeable change of activity. In this way compression ratio may be increased by about 10 to 20 percent.

Of interest are the particular cases of the prediction.

3) *Differential Pulse-Code Modulation (DPCM):* This technique allows compression ratios around 2.5 : 1. Adaptive DPCM schemes exist leading to compression ratios as high as 3.5 : 1.

4) *Delta Modulation:* Average compression ratio obtained with delta modulation is also not very high but the technique is rather simple.

5) *Interpolative Coding:* Most commonly used interpolators are zero-order and first-order interpolators, giving compression ratios around 4 : 1. Higher order polynomials or splines can also be used, but their computational complexities do not justify the results. Again, properties of the human visual system are not used.

6) *Bit-Plane Coding [12]:* With bit-plane coding an average compression around 4 : 1 can be obtained without making any use of the properties of the human visual system.

Transform methods: The basic motivation behind transform coding is to transform a set of data (pels) into another set of "less correlated" or "more independent" coefficients, before coding. The inverse transform recovers the original picture. Most commonly used transformations are linear transformations implemented with fast algorithms for computational efficiency.

1) Karhunen–Loéve transform: The Karhunen–Loéve transform is the best linear transformation in the sense that it leads to uncorrelated coefficients. It is, however, not often used in practice because of its computational load. It gives an indication about the upper bound of what other transformations, computationally more efficient, should attempt to reach in decorrelating data samples.

2) Fast transformations: There are several linear transforms which can be computed with $N \log_2 N$ operations as compared to $N^{**}2$. The most important ones are labeled as Fourier, Hadamard, Haar, sine, cosine, and slant. One important difference between these transformations and the Karhunen–Loéve transformation is that these transformations do not depend on the statistics of the input image.

3) Coding strategies: There are several coding strategies in transform coding. First the dimensionality of the transformation must be determined. A still picture can either be transformed by a two-dimensional transform or by a one-dimensional transform to be coded on a line-by-line basis. The next parameter to be fixed is the size of the transform. A commonly used strategy is to subdivide the image into subpictures of size M by M with M much smaller than the size N of the image (for example, $M = 32$ and $N = 512$) and to transform each subpicture separately. The important characteristic of these transformations is that all the "important" coefficients are packed into a specific area of the transform domain. Important compressions (up to 10 : 1) can be obtained depending on the number of coefficients retained in a given area. Another possibility is to put a threshold on the transform coefficient magnitude and set to zero all those below the threshold. Higher compression ratios (around 15 : 1) can be obtained with good-quality decoded pictures. Although there is some evidence that the human visual system is not a linear transformer, its basic properties may be included in the design of a transform coder.

Transform coding may also be adaptive by matching the parameters of the coder to the statistics of the subpicture being coded. Adaptation can be made at the level of the transform, of bit assignment, or of quantizer level assignment. Coding efficiency can be increased by about 25–30 percent compared to the nonadaptive case.

Hybrid methods [13]: In its common use, hybrid coding refers to methods combining predictive coding and transform coding. Both DPCM and transform coding techniques have some attractive characteristics and some limitations. Combining these two techniques leads to hybrid coding methods capable of achieving compressions around 8 : 1, having the advantages of hardware simplicity (DPCM) and robust performances (transform coding).

C. Synthetic High System [14]–[20]

Although presented in this section mainly for chronological reasons, it is our belief that this technique is the only one which was in advance for its time. It is surprising to see that it is even not referred to in general surveys. This technique has been known for more than two decades! In the early days of image coding it was considered a heavy

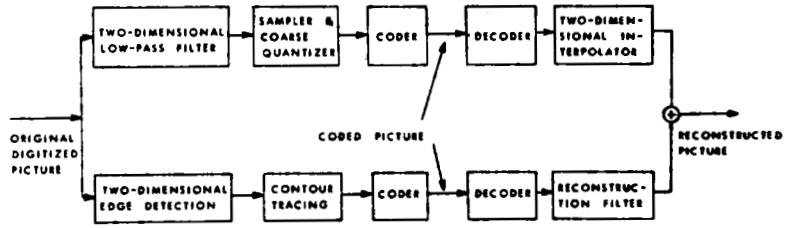


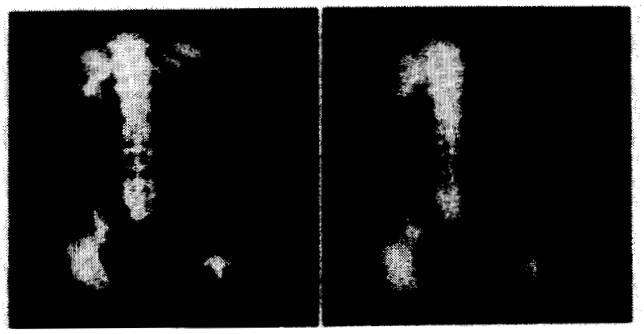
Fig. 3. Block diagram of the synthetic high system.

method because of its computational complexity, but today, it is as easy to implement as anything else with present technology. Compared to its contemporary methods, it gives very high compression ratios. Also, it gave a lot of insight to develop "second-generation" techniques. A number of studies [21]–[24] using the decomposition of a picture into low-frequency and high-frequency components are, directly or indirectly, based on the synthetic high system.

The sequence of messages for the synthetic high system is selected in the following manner. The original picture is split into two parts: the low-pass picture giving the general area brightness without sharp contours, and the high-pass picture containing sharp edge information. In its analog form, it is an information-lossless technique. According to the two-dimensional sampling theorem, the low-pass picture can be represented with very few samples. These samples are the messages characterizing the canonical form of the low-pass picture. The edge detection is performed either by a gradient or a Laplacian operator. A nonlinear operation, thresholding, is performed on the high-pass picture to determine whether an edge point is important. Then the method becomes information-lossy. Finally, the location and the magnitude of each selected edge point are coded. These variables are the messages characterizing the high-pass picture.

A two-dimensional reconstruction filter, whose properties are determined uniquely by the low-pass filter used for the low-pass picture is used to synthesize the high-frequency part from edge information. This "synthetic high" signal is then added to the low-pass picture to give the final output. A block diagram of this system is shown in Fig. 3. The reader may refer to [15]–[20] for more detail.

The synthetic high system exploits elegantly the properties of the visual system at its early processing levels by making use of the lateral inhibition phenomenon (see Section III-C). Thereby, it permits a considerable amount of redundancy reduction. Fig. 4 shows an original X-ray picture and its reconstructed version obtained with the synthetic high system with a compression of 8.3 : 1. The values of the thresholds used in this system are of primary importance for the compression ratios and for the quality of the reconstructed pictures. A high compression ratio is obtained but a large amount of texture and detail are lost if these thresholds are kept too high. On the other hand, a good picture quality but a poor compression ratio are obtained if they are kept too low. With a lack of theoretical method, the compromise between compression ratio and image quality can best be found with an empirical cut-and-try procedure. The directional decomposition based method, discussed in Section V-B, can be viewed as a refinement of the synthetic high system, where the ability to extract and code edges is improved with directional filters.



(a) (b)

Fig. 4. An original X-ray picture (a) and its reconstructed version obtained with the synthetic high system (b). Compression ratio: $C = 8.37$.

III. Mechanism of Vision

As discussed earlier, the very end of almost every image processing system is the human eye. Therefore, it is useful to know how and what the eye sees. The answers to these questions lie far outside from our usual engineering context. That is why, a summary of the brain mechanism of vision is given here from [25] and [26] along with our engineering interpretations.

A. General Description

The human visual system is a part of the nervous system. The latter is doubtlessly the most complicated communication network. It is managed by the most powerful computer: the brain. The communication in this network is carried out through nerve cells called neurons. The brain contains about 10^{11} neurons, roughly the same number as that of stars in our galaxy.

A neuron has a body of size varying between 5 and 100 μm. A main fiber called the axon and a number of fiber branches called dendrites are attached to this body. Fig. 5 depicts some typical neurons.

The information transfer from one neuron to another is made electrochemically. The junction between two neurons is called the synapse. The transmitting and the receiving neurons are called presynaptic and postsynaptic, respectively. The information generated in a presynaptic neuron travels its axon like an electrical signal in a cable. Terminal branches of the axon transmit this signal to the dendrites of the postsynaptic neuron. During this transmission, the electrical signal generates chemicals at the end of the axon which are deposited on the postsynaptic neuron. In turn, these chemicals generate a new electrical signal in the postsynaptic neuron. A neuron can receive signals from thousands of presynaptic neurons and can transmit to thousands of postsynaptic neurons. A given neuron can handle up to 200 000 synapses. This shows how capable a

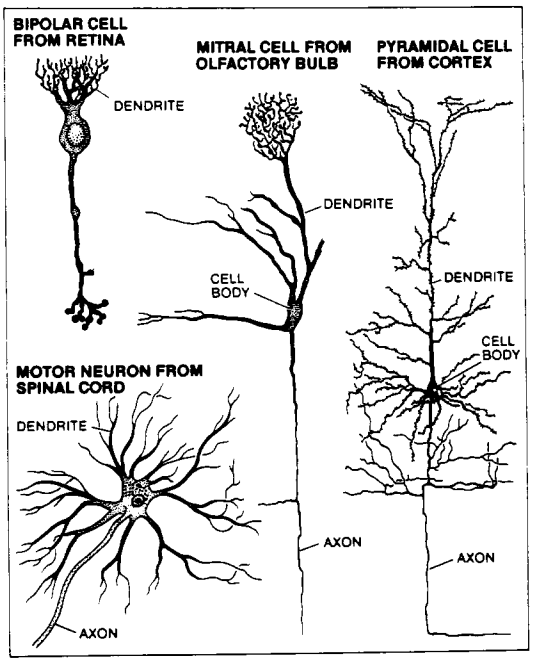


Fig. 5. Typical neurons (after [26]).

neuron is compared to an integrated circuit whose fanouts cannot exceed a few dozen. The action of a neuron can be of two types: excitatory or inhibitory. The first one generates pulses in the postsynaptic neuron whereas the second inhibits the existing pulses. This basic behavior is at the origin of a variety of phenomena such as Mach bands, band-pass characteristic of the visual frequency response, and the edge and line detection mechanism of the eye. Moreover, it provides the justification for the existing so-called local operators for edge detection.

To have an idea on the complexity of the nervous system, just imagine a network of 10^{12} neurons connected in cascade, in parallel, and with feedback! A schematic view of this organization is depicted in Fig. 6. A successful application of this structure is the architecture used in the WISARD system developed by Aleksander and his team [27] to recognize faces. The analysis of such a complex network of neurons seems, *a priori*, quite difficult if not impossible. There are however a number of features which simplify the study. These features make it possible, nowadays, to analyze the nervous system on a cell-by-cell basis. The first characteristic is that there are only two types of signals in the nervous system: one for long distances and the other for short distances. The second characteristics is that these signals are almost identical in all neurons regardless of the information they carry: visual, tactile, audible, etc. Moreover, their shape does not vary from species to species. A signal recorded from a cat is similar to that recorded from a human being. The signals received and processed by the brain are thus symbols representing external events. The nervous system is analyzed with three basic tools of the neurobiologist: the microscope, the selective stain, and the microelectrode. The signal recorded at a given neuron is a pulse train. Each pulse has a magnitude of about 100 mV and a duration of about 1 ms. The repetition rate (frequency) of these pulses is proportional to the intensity of the stimulus. The nervous system communicates through frequency modulation. What allows the brain to distinguish between two identical signals is the pathway used by each of the signals, in other words the wiring. There is thus a specific ensemble of neurons corresponding to each type of excitation. From a mathematical point of view, there is a one-to-one mapping between different parts of the body and the brain.

B. The Eye

The eye is the sensor of visual signals. It focuses them to form the image on the retina. The latter analyzes the image and sends the message to the brain through the optical nerve and optical paths in the head. Very roughly stated, the eye can be viewed as a photo camera (see Fig. 7). The light passes through the cornea and aqueous humor and enters the inner eye through the pupil. The amount of light allowed to enter is regulated by the pupil as a diaphragm. The lens focuses the light on the photoreceptive cells of the retina. The volume of the eye is not large (6.5 cm^3). Its diameter is about 24 mm and it weighs 7 g. The lens of the eye is not perfect even for persons with no weakness of

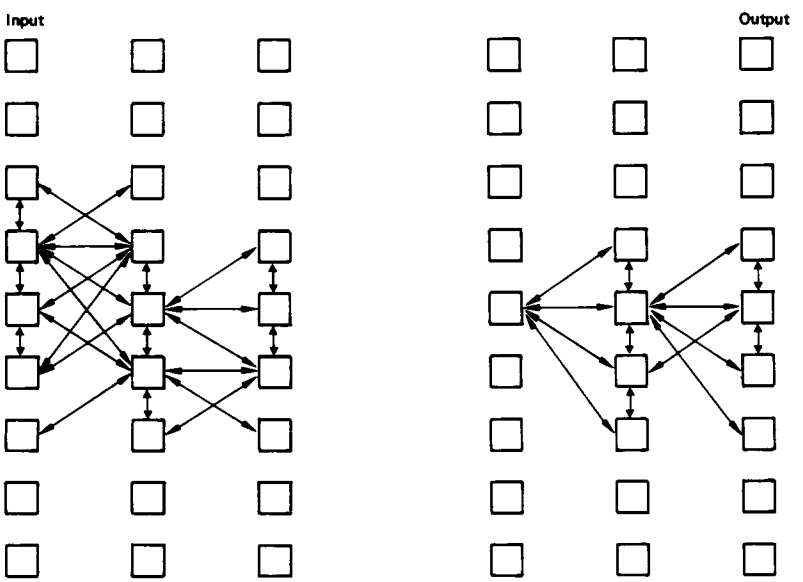


Fig. 6. Block diagram of the nervous system.

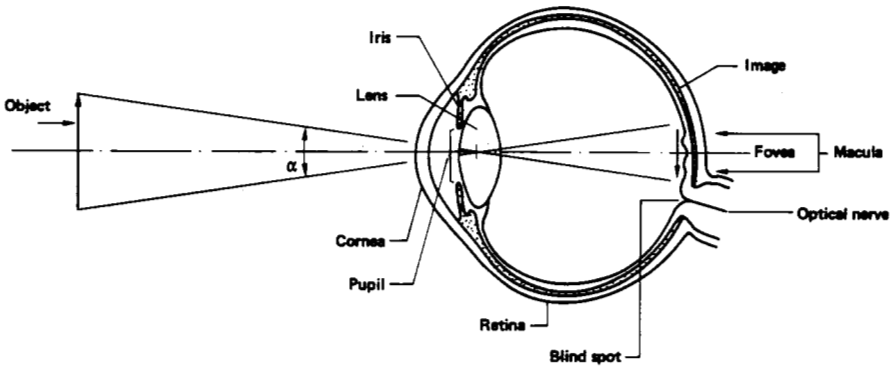


Fig. 7. Schematic view of the eye.

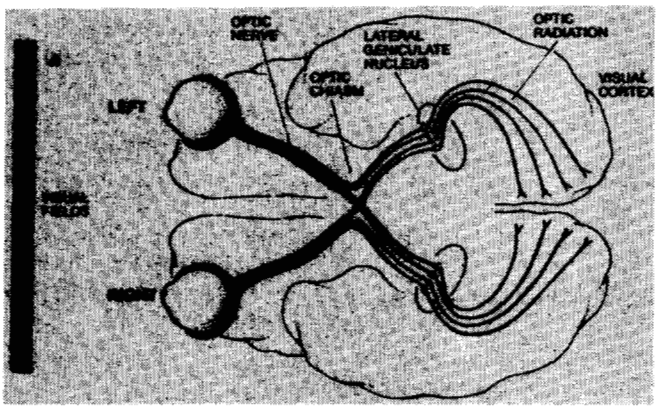


Fig. 8. Pathways in the human visual system.

vision. This imperfection is the source of the spherical aberration which appears as a blur in the focal plane. Such a blur can be modeled as a two-dimensional low-pass filter. The pupil's diameter varies between 2 and 9 mm. This aperture can also be modeled as a low-pass filter. The highest cutoff frequency corresponds to 2 mm. Continuous enlargement of the pupil's diameter decreases the cutoff frequency.

C. The Retina

The retina is the neurosensorial layer of the eye and its area is about 12.5 cm^2. It transforms the incoming light into electrical signals that are transmitted to the visual cortex through the optic nerve. Optical pathways in the visual system are depicted in Fig. 8. It should be noticed that at the optic chiasm, the output of each eye is divided into two. Consequently, the information content of the left half of the visual field is processed by the right side of the brain and the information content of the right half of the visual field is processed by the left side of the brain.

The anatomy of the retina shows five types of cells organized in layers as schematized in Fig. 9. Notice here an illustration of the general wiring principles of Fig. 6. The furthest layer from the incoming light is that of photocells. There are two types of photocells: rods and cones. A normal eye contains about 130 million rods and 6.5 million cones. Rods and cones are different enough to be examined separately. Rods are sensitive to shapes and need low luminance (scotopic vision). In contrast, cones need daylight (photopic vision). They detect color and distinguish details.

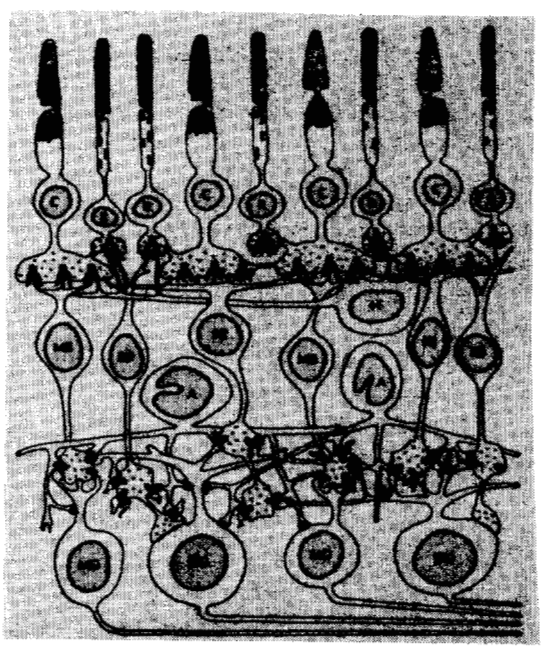


Fig. 9. Cells of the retina (after [29]).

Their distribution in the retina is highest in the vicinity of the optical axis of the eye. That is why a precise detail vision is obtained only when the eye "fix" them, in other words, when their image is formed at the fovea. In this region there are about 120 cones per degree which fix the visual resolution to 1 min of arc.

Photoreceptors are responsible for transforming the in-

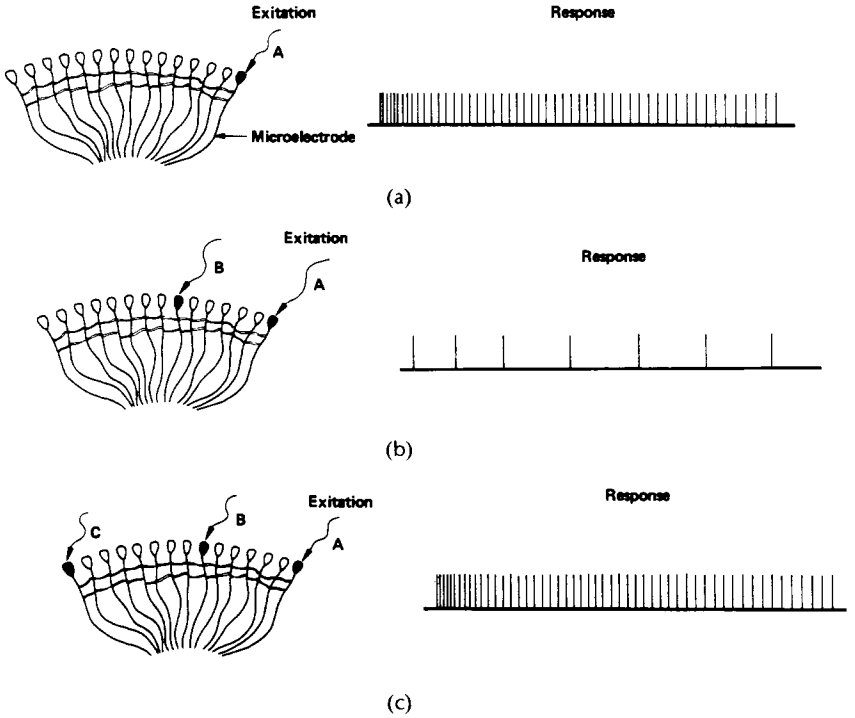


Fig. 10. Illustration of the lateral inhibition. An active cell A (a). Activation of B cells inhibit cell A (b). Disinhibition of A (c).

coming light into electric signals, while compressing its dynamic range. This compression is made according to a nonlinear law whose classical interpretation leads to a logarithmic curve (Weber–Fechner law). It is obtained by the following experiment. If a visual field of luminance L is divided into two parts along a straight line and the luminance of one part is increased by ΔL until a difference is noticed, the ratio $\Delta L/L$ remains constant for very large variations of L. If this result is interpreted as a small increase ΔB in brightness, $\Delta B = \alpha \Delta L/L$ and by integration we obtain $B = \alpha \log L + \beta$. More recent experiments, done with complex scenes, led to power laws $B = aL^{\gamma}$, with $\gamma = 1/2$ or $1/3$ depending on bright or dark surrounding, respectively, as nicely reported in [28].

In serial connections of the retina (Fig. 9), photoreceptors are connected to bipolar cells which, in turn, are connected to ganglion cells, whose axons make up the optical nerve. In parallel connections, horizontal cells receive synapses from photocells and may act on bipolar cells and photocells. Amacrin cells receive synapses from bipolar cells and may act on ganglion cells and bipolar cells. Some of these actions indicate feedback loops in parallel connections.

These parallel and feedback connections are responsible for the so-called lateral inhibition [29]. This phenomenon can be summarized as follows (Fig. 10). When a given cell A is excited, it produces a pulse train whose frequency is proportional to the intensity of light (Fig. 10(a)). If other neighboring cells B are excited as well, they inhibit the pulse train generated by A (Fig. 10(b)). A disinhibition of cell A may occur when the inhibition exerted on it by B cells is partially released by exciting other cells C, in the close neighborhood of B, that inhibit the action of B but are too far away from the cell A to act on it (Fig. 10(c)). If the responses of these cells are recorded spatially, a point-spread function or a spatial impulse response is obtained as depicted in Fig. 11. Clearly, this corresponds to a high-pass filter. It is interesting to note that the same effect, due to totally different mechanisms, occurs in xerographic copying which does not reproduce faithfully large black areas, unless a halftone screen is used to brake up the original black pattern into small dots. Also noticeable is the fact that the synthetic high system is the first image-coding method which makes an elegant use of this property.

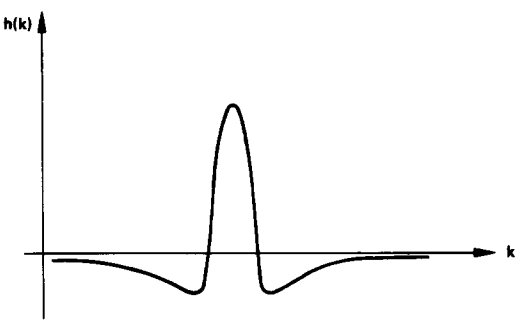


Fig. 11. Spatial impulse response of the retinal cells.

As one may suspect, the high-pass behavior of the lateral inhibition does not extend infinitely. Beyond a radial spatial frequency of about 10 cycles per degree of solid angle, an integration takes place due to previously described low-pass filters and to the limited number of cells per unit retina length. If two excitations are closer than 1 min of arc to each other, their projections on the retina fall on the same cell and, therefore, are not distinguishable. As a summary, at this level of the visual system, a linear model with a bandpass frequency response can be assumed, as shown in Fig. 12.

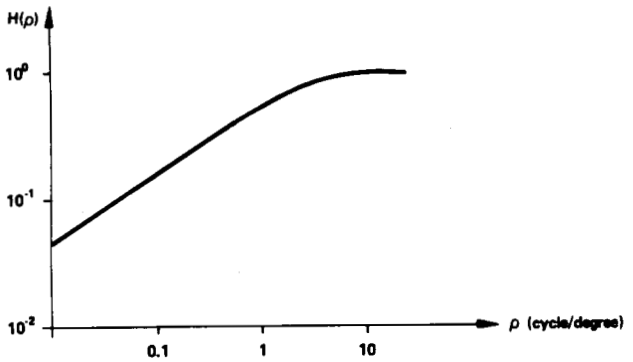


Fig. 12. Frequency response (radial) of the retinal cells.

D. Ganglion Cells and the Lateral Geniculate Nucleus

To study specialized cells, such as ganglion cells or the other cells in the visual cortex, two concepts need to be introduced. The first one is that of the receptive field of a cell. It is simply the area of the retina which can influence the behavior of that cell. The second is that of the most effective excitation. It is the stimulus which produces a pulse train of highest frequency. Because of parallel and feedback connections, a receptive field is divided into regions. Some of them are excitatory whereas some others are inhibitory. They are often called "on areas" and "off areas," respectively. The effective excitation of a ganglion cell is a circular spot whose diameter is around 0.2 mm. The receptive fields are also circular and there are two types: those which are "on" in the center and "off" at the border and those which are "off" in the center and "on" on the border (Fig. 13). At this level in the visual system, the information is processed independently from spatial orientation. Notice also that the absolute light intensity is ignored by ganglion cells which measure only differences in their receptive fields. A systematic analysis shows that even a very small spot (0.1 mm in diameter) can cover several overlapping receptive fields, some of the ganglions cells being excited and some inhibited. These cells are spatially grouped. More generally, cells in the visual system processing information coming from a given area of the retina are grouped. This is thus an economical arrangement of the wiring. The receptive fields in the vicinity of the fovea are smaller than the others (high visual resolution). Their area increases progressively as we go away from the fovea. With this last observation, the previous high-pass filter can be said to be shift-variant because of nonconstant spatial resolution. Simultaneous contrast effect and edge detection are mainly due to ganglion cells. Their effect can be modeled as a two-dimensional high-pass filter. Lateral excitations and inhibitions are also the source of the so-called Mach phenomenon which is the subjective enhancement of sharp luminance changes between relatively uniform regions.

The information processed by the retina reaches the lateral geniculate nucleus after division at the optic chiasma. The cellular analysis of this nucleus shows also a layered organization of cells. Cells on each layer receive information only from one eye. This is called ocular dominance. Moreover, here also, neurons receiving information from a given area of the retina are grouped, independently from the specific layer. Even if the exciting eye is changed, the position in the visual field remains stable. The cells in the lateral geniculate nucleus function in a way very similar to that of ganglion cells. Independence of orientation is maintained.

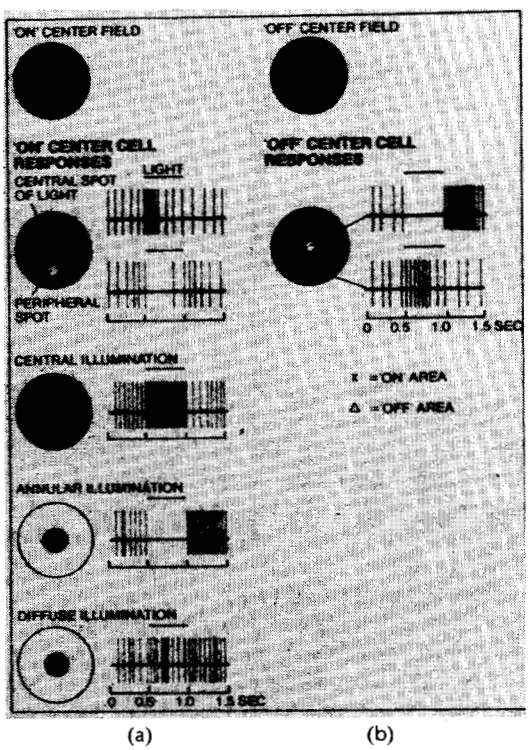


Fig. 13. Receptive fields of ganglion cells. Center on surround off (a) and center off surround on (b) (after [26]).

E. The Visual Cortex

Located at the back of the brain, the visual cortex is a folded layer of neurons of about 2 mm in thickness. The information transmitted from the lateral geniculate nucleus is received at the "area 17" of the cortex which is connected to areas 18 and 19. Cellular analysis of the cortex indicates also that 10^{10} neurons contained in it are hierachically organized in layers with only a few types of neurons. They are classified as simple cells, complex cells, hypercomplex cells, and higher order hypercomplex cells.

The receptive field of a simple cell is elliptical. The center is bar shaped and surrounded by two opposed areas (on and off). The effective excitation is not a circular spot but a slit oriented in the same direction as that of the central bar of the receptive field. If the slit is rotated, other simple cells start reacting depending on the angle. It is at this level of the visual system that specific processing for a given orientation is introduced. For a fixed orientation, if the excitation is slid in the receptive field, the pulse train vanishes progressively. Fig. 14 shows the receptive field of a simple cell. The receptive field of some simple cells may have only two antagonistic areas. In this case the most effective excitation is a border between the zones. It is interesting to note here the similarity between this receptive field (also the one in Fig. 13) and various masks used for edge and line detection by local operators such as Roberts, Sobel, Kirsch, Hueckel, and Mero-Vasy [2]. One may wonder if these operators were designed on this basis!

Complex cells are also sensitive to the orientation of the

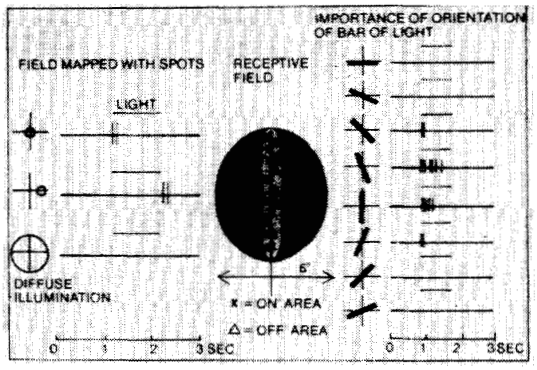


Fig. 14. Receptive field of a simple cell (after [26]).

excitation. In contrast with simple cells, however, they are not sensitive to the position of the excitation in the receptive field. Complex cells indicate orientation independently from the position.

The effective excitation for hypercomplex cells also requires a specific orientation, but in addition, it needs a discontinuity such as a corner or the end of a line. Fig. 15 summarizes the functions of these various cells in the perception of a white rectangle on a dark background.

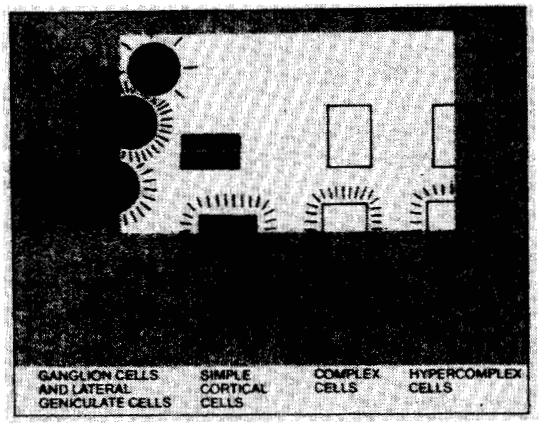


Fig. 15. Perception of a white rectangle on a dark background by cortical cells (after [26]).

F. Visual System Model

A systematic, cell-by-cell analysis of the cortex indicates a columnar organization in both ocular dominance and direction dominance. If a microelectrode penetrates perpendicularly to the cortex, all the cells it encounters respond to excitation coming from one eye only. A neighboring penetration shows cells responding to the other eye. Finally, if the penetration is slanted from one layer to the next, the ocular dominance alternates from one eye to the other.

It is interesting to note that if the same experiment is repeated by looking, not for the ocular dominance, but the response to a given orientation, the same columnar organization can be observed. A perpendicular penetration shows that all the cells encountered respond to the same orientation. In the neighboring column, the orientation is slightly different. There are roughly thirty quantized directions.

Combining the results of these two experiments, a columnar model of the cortex is obtained, as shown in Fig. 16, where the bars in the columns indicate the preferred direction.

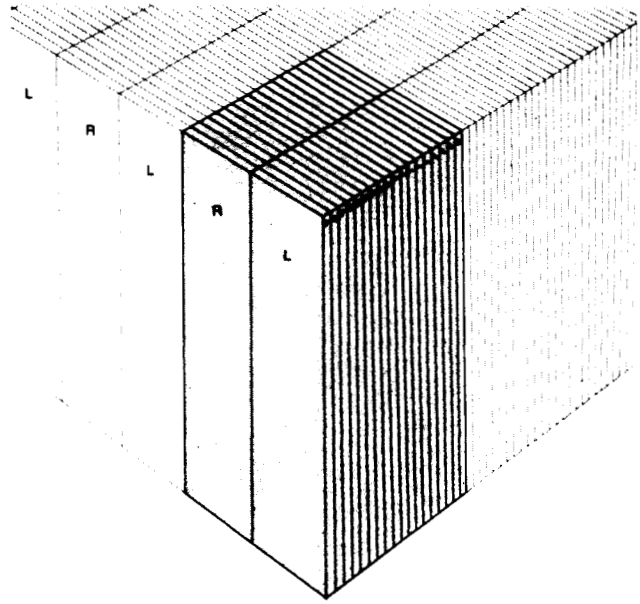


Fig. 16. Hubel and Wiesel's columnar model.

All the properties described in this section can be summarized in a block diagram (see Fig. 17) where parts related to the lens, the retina, and the cortex are indicated. Bars in the boxes indicate the directional filters followed by another filter bank for detecting the intensity of the stimulus. The first block is a spatial, isotropic, low-pass filter. It represents the spherical aberration of the lens, the effect of the pupil, and the frequency limitation by the finite number of photoreceptors. It is followed by the nonlinear characteristic of the photoreceptors. Here, a logarithmic curve for simplicity or a law of the type L^{γ} for more accuracy, can be used. At the level of the retina, this nonlinear and very likely memoryless transformation is followed by an isotropic high-pass filter corresponding to the lateral inhibition phenomenon of the ganglion cells. The processing done by the cells in the lateral geniculate nucleus can be included in this high-pass filter.

Finally, there is a directional filter bank that represents the processing performed by the cortex' cells. It should be noted that the whole system is shift-variant because of the decrease in resolution away from the fovea. This block diagram is the basis of the GOP system which will be described in Section IV.

IV. Local Operator Based Techniques

A. Pyramidal Coding [30]

1) General Remarks: Because it combines features of predictive and transform coding methods and because the compressions obtained are not too high (around 10:1), this pyramidal coding technique could have been described as a hybrid method of the first generation. It is included in this section, however, since its hierarchical structure is similar to that of the nervous system, and it uses functions close to

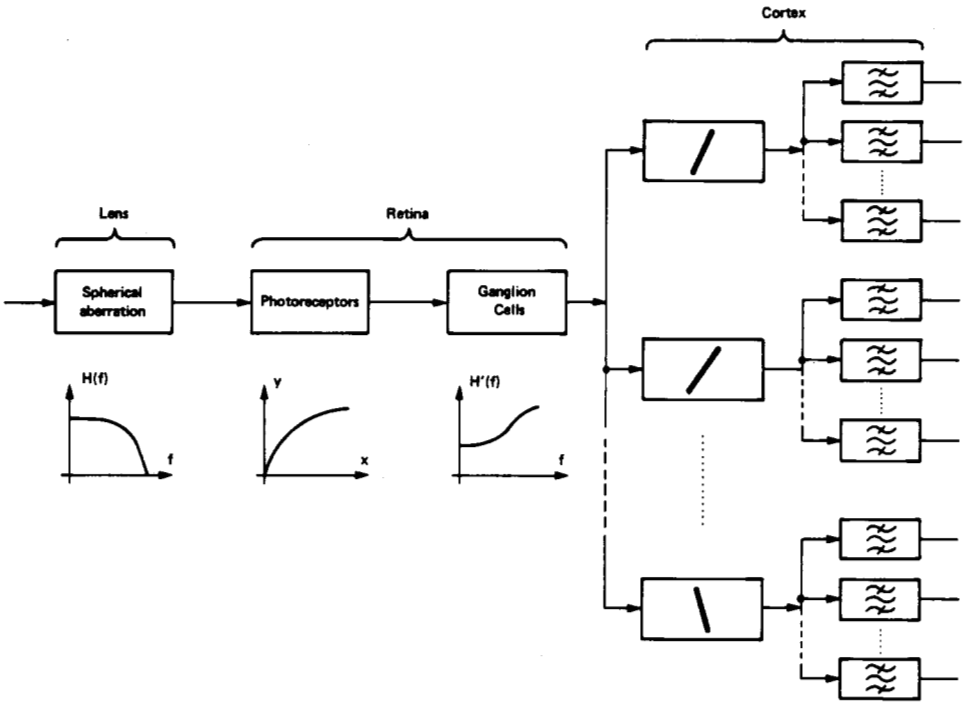


Fig. 17. Block diagram of the human visual system.

those of the human visual system. It has also elegant capabilities for progressive transmission or reconstruction. The reader may notice the role played by the synthetic high system in the design of this method.

2) Basic Method: Starting from the original picture $x(k,\ell)$ to be coded, a low-pass version $x_1(k,\ell)$ of it is computed using local averaging with a unimodal Gaussian-like two-dimensional impulse response. The low-pass image, with a cutoff frequency of f_1, can be viewed as a prediction of $x(k,\ell)$. The prediction error is then

$$e_1(k,\ell) = x(k,\ell) - x_1(k,\ell). \tag{1}$$

Clearly, coding $e_1(k,\ell)$ and $x_1(k,\ell)$ is equivalent to directly coding the picture itself. The compression that can be expected is due to two facts. a) $e_1(k,\ell)$, by its nature, is a high-pass image. Because of the low sensitivity of the eye at these frequencies, fewer bits per sample than those used for the original picture could be sufficient. b) $x_1(k,\ell)$ is a low-pass image. It can be represented by fewer samples than those necessary to represent the original image, because of the two-dimensional sampling theorem.

At this level, beside particular interpretations and specific functions used, this method is conceptually identical to the synthetic high system. The refinement is that the basic operations described above are iterated. More specifically, $x_1(k,\ell)$ is again low-pass filtered, say at a cutoff frequency f_2, and the result, $x_2(k,\ell)$ is used as a prediction of $x_1(k,\ell)$. The error for this prediction is then

$$e_2(k,\ell) = x_1(k,\ell) - x_2(k,\ell). \tag{2}$$

Note that this second error image is smaller than $e_1(k,\ell)$ by a factor which is the ratio of the two cutoff frequencies f_1/f_2. After n iterations, a series of prediction error images $e_1(k,\ell)$, $e_2(k,\ell), \cdots, e_n(k,\ell)$ are obtained. At each iteration, the dimensions of these images are reduced by a factor $f_i/f_{(i+1)}$. For a simple implementation, a factor of two can be used at each iteration. If these images are viewed as stacked one above another, the result is a pyramidal data structure. The image at a given level of this structure is the difference of two Gaussian-like functions convolved with the original image. The difference of two Gaussian-like two-dimensional functions is a good approximation of the impulse response representing the lateral inhibition phenomenon of the human visual system. These functions are extensively used to detect intensity changes [31], [32].

A fast algorithm is proposed in [30] to obtain a series of prediction error images. A 5 by 5 separable Gaussian kernel is used to determine equivalent two-dimensional impulse responses $h_i(k,\ell)$ for each level i of the pyramid. Convolving the original image with these impulse responses leads directly to the error images. These images are then quantized and coded. To reconstruct the decoded image, interpolation filters are designed and used to compensate the decimation performed at each level. When all the error images are decoded and interpolated to reach the original resolution, their pel-by-pel sum gives the decoded image. Fig. 18 shows a block diagram of this system. A nice feature of this system is that the quality of the decoded picture can be improved as desired at the expense of a lower compression ratio. Good-quality pictures with an average compression ratio around 10:1 are obtained as shown in [30] and in Section VI.

B. Anisotropic Nonstationary Predictive Coding [33]

1) General Remarks: Like the pyramidal coding procedure described previously, this method can also be classified as a hybrid method for combining prediction and transform coding. We consider it as a "second-generation" method because of the extensive use of the properties of the human visual system and because of the high compressions achieved. It is a good example showing the emphasis put on the selection of messages before coding, an im-

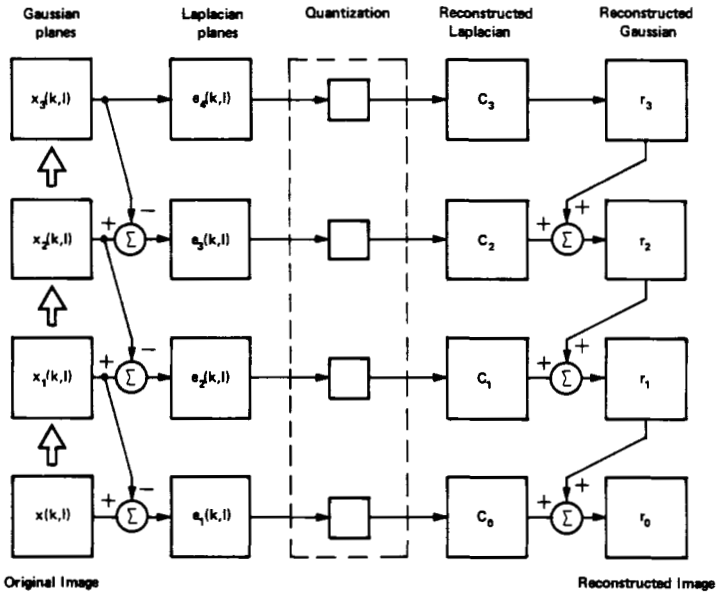


Fig. 18. Block diagram of the pyramid coding method (after [30]).

portant feature of the second-generation methods. Impacts of the synthetic high system in the design of this method are also quite visible.

2) Principles of the Method: The principles of the anisotropic, nonstationary predictive coding can be viewed as a by-product of a scheme for image restoration [34]. Starting with the classical estimation problem of recovering the estimate $\hat{x}(k,\ell)$ of an image $x(k,\ell)$ corrupted with additive, stationary, and independent white noise, refinements for the well-known Wiener filter are derived. The resulting restoration filter has three components.

The first component is the classical isotropic Wiener filter expressed as

$$H_1(\rho) = \frac{\Phi_x(\rho)}{\Phi_x(\rho) + \sigma_n^2} \tag{3}$$

where ρ is the radial frequency in the Fourier domain, $\Phi_x(\rho)$ the power spectrum of the original image, and σ_n^2 the noise power. This filter has a low-pass characteristic since the image has more low-frequency energy than the noise.

The second component is also an isotropic filter given by

$$H_2(\rho) = \frac{\sigma_n^2}{\Phi_x(\rho) + \sigma_n^2}. \tag{4}$$

In contrast with the first component, it has a high-pass characteristic. The combination of these two components leads to a bandpass behavior such as that of the early stages of the human visual system (see Fig. 12).

The third component is an anisotropic filter given by

$$H_3[\rho,\theta,\varphi(k,\ell)] = \frac{\cos^2[\varphi(k,\ell) - \theta]\sigma_n^2}{\Phi_x(\rho) + \sigma_n^2} \tag{5}$$

where θ is the polar angular frequency in the Fourier domain. This filter is used to weight local noise according to its direction with respect to that of a local anisotropy indicated by $\varphi(k,\ell)$. Local noise components aligned with $\varphi(k,\ell)$ are enhanced, whereas those which are orthogonal to $\varphi(k,\ell)$ are attenuated for better restoration.

The final restoration filter is a weighted sum of these three components

$$H(\rho,\theta,k,\ell) = H_1(\rho) + \beta(k,\ell)\tau(k,\ell)H_2(\rho) + \beta(k,\ell)[1 - \tau(k,\ell)]H_3[\rho,\theta,\varphi(k,\ell)]. \tag{6}$$

Weight function $\beta(k,\ell)$ is related to the magnitude of locally rectilinear features (edges and lines) in the vicinity of the point (k,ℓ) whereas $\tau(k,\ell)$ represents the variability in angle of the same feature. The weight $\beta(k,\ell)$ can be viewed as direction-insensitive estimation of a local nonstationarity. Thus together with the directional information $\varphi(k,\ell)$, it forms a vector image which indicates the magnitude and the direction of locally rectilinear features of the image. All the unknown functions in (6), such as $\beta(k,\ell)$, $\varphi(k,\ell)$, $\tau(k,\ell)$, and the frequency responses of the filters must be estimated from the available data. To alleviate the computational burden on these estimations, spatially limited filters (local operators) are suggested in [34] as arrays of at most 15 elements. Furthermore, the restoration can be performed iteratively by several passes over the data. Good restoration results are obtained by using a high-speed, parallel GOP processor [36].

3) Application to Coding: The logical way to use the restoration filter for picture coding is to transform it into a prediction filter and to use the prediction error as the messages to be coded. In this scheme, the discrete cosine transform is first applied to the prediction error before quantization and coding. Weighting functions $\beta(k,\ell)$, $\varphi(k,\ell)$, and $\tau(k,\ell)$, estimated from the original picture, must be coded and transmitted (or stored) as well.

In general, the restoration filter need not be recursive. The prediction filter, however, must be recursive. This requires the truncation of the corresponding impulse responses involved in (6) to obtain a recursive two-dimensional differential equation to implement the prediction. When the prediction and its error are computed, a one-dimensional discrete cosine transform is applied to each line of the prediction error image. The transform coefficients are then quantized in a two-level process: adaptive coding of the stationary part and threshold coding of the nonstationary part.

Weighting functions are coded in a rather simple way.

These functions have low bandwidth. Accordingly, they are coded with 1:6 undersampling in both directions. Magnitudes, angles, and the variations in angle are coded with 2, 4, and 1 bits, respectively. This leads to a very high compression, higher than 70:1, for representing the weighting functions. Taking into account the number of bits used to code the prediction error, global compressions as high as 35:1 are obtained with good-quality decoded pictures.

V. Contour–Texture Oriented Techniques

In contrast with local operator based methods, contour–texture oriented techniques attempt to segment the image into textured regions surrounded by contours such that the contours correspond, as much as possible, to those of the objects in the image. Contour and texture informations are then coded separately. Contours may be extracted in two ways: by region growing or by using the by now well-known contour extraction—or edge detection—techniques [1], [2]. In the first case, closed contours are obtained which makes it simple to list regions and their properties. The segmented image looks like a puzzle image. In the second case, contours that are obtained are not necessarily closed. Combination of this information with texture information becomes thus more problematical. Two techniques using these approaches are described below, leading to compressions as high as 70:1.

A. Region Growing Based Coding [37]–[40]

1) Segmentation: In the first stage of this method, the image is segmented to classify its pels into contour pels and texture pels. This procedure partitions the image into a set of adjacent regions under the constraint that the variation of the grey level within the region does not contain any sharp discontinuities, i.e., contours. A similar idea was suggested in [20]. Segmentation is carried out in three steps: preprocessing, region growing, and elimination of artifacts.

The preprocessing is intended to reduce the local granularity of the original image without affecting its contours, so that not too many small regions are obtained after region growing. Furthermore, these small regions do not correspond, in general, to real objects of the original image and thus become false contours. The key problem in this preprocessing is the reconciliation of apparently two contradictory goals; namely, granularity removal and edge preservation. Most of the granularity removal filters have low-pass characteristics and therefore smooth the edges as well. The inverse gradient filter [41] is chosen because of its ability to adapt its coefficients according to local contrast. Accordingly, this filter behaves like a low-pass filter in areas free of contours and like an all-pass filter in highly contrasted areas. Since its frequency characteristics are not perfectly low-pass or all-pass, it is not very efficient in removing granularity, at least in one pass. That is why it has to be applied iteratively until the granularity has been sufficiently removed. The result of applying this filter to a one-dimensional signal is shown in Fig. 19.

The mechanism of region growing is the following. Regions to be extracted must be characterized with some property in the first step. The property might be, for example, the grey level of a pel, the variation of the grey level, or the energy within a given frequency band. The selection of this property plays a very important role in the complexity

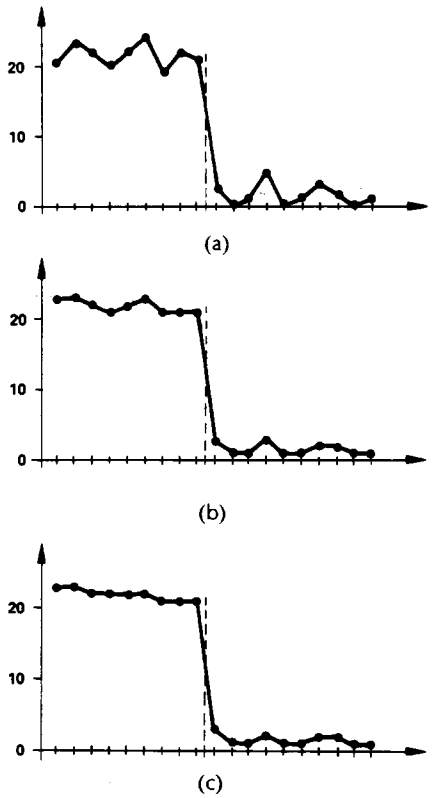


Fig. 19. Inverse gradient filtering of a one-dimensional signal. The original signal (a), result after one iteration (b), result after two iterations (c).

of the method and in the exactness of the contours obtained after segmentation. Then, starting with a given pel in the picture, its neighboring pels are examined to see whether they share the same property. If this is the case, that pel is included in the region, and in turn, its neighboring pels are examined, and so on. When there are no more pels left, connected to the region and sharing the same property, the procedure stops and restarts at any other pel which is not included in the first region. The segmentation is complete when all the pels of the picture are assigned to some region. The property used in the example below is very simple: it is a fixed grey-level interval. Although it has a constant width, this interval is made adaptive by moving it up and down on the grey-level scale in order to intercept the maximum number of pels. This displacement is constrained, however, so that previously intercepted pels always remain in the region. After this region growing, two types of artifacts are obtained: Contours which do not completely separate two regions and contours that are two pels wide. Fig. 20 shows these artifacts. They are easily eliminated with an appropriate procedure [38]. Two original pictures and the results of the region growing are shown in Fig. 21. The number of regions are 1001 and 1111 for the cameraman picture and the building picture, respectively.

At this level, the resulting image can be viewed as that of a puzzle with one-point thick closed contours. Unfortunately, because of the simple property used, the number of these contours is much higher than that of the objects in the original image. Two procedures are available to overcome this problem: introduction of some distortions by eliminating insignificant regions and their contours, or the use of a more refined property. The first alternative is

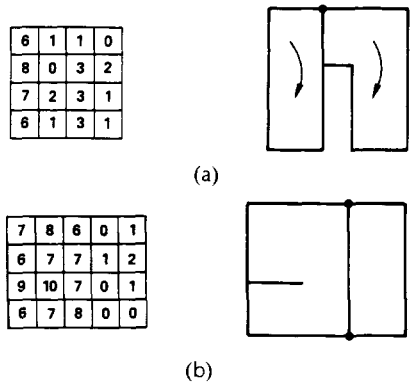


Fig. 20. Artifacts resulting from region growing. Two-point thick contours (a) and incompletely separated regions (b).

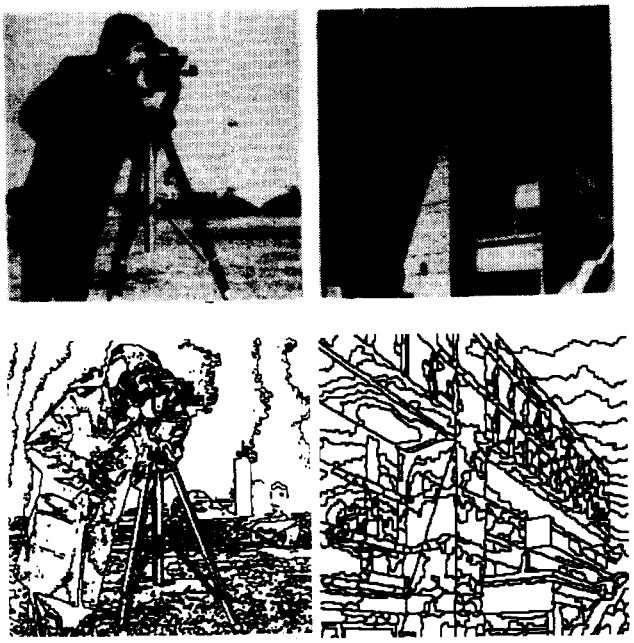


Fig. 21. Two original images and the result of region growing with a grey-level interval of 10 in both cases.

described below whereas the second is under investigation at present.

There are two heuristics to decrease the number of regions obtained by region growing: elimination of the small regions and merging weakly contrasted adjacent regions.

Fig. 22 shows the histogram and the cumulative histogram of the number of regions as a function of their size expressed in terms of the number of inner points resulting from the image of the building in Fig. 21. These curves indicate that roughly 70 percent of the regions have less than 15 pels. There are two reasons for this result. First, the granularity is not completely eliminated during preprocessing, and second, small areas of high gradient in the original image are segmented into several small regions because of the property used. If it is assumed that regions containing a number of pels less than a threshold are not significant, their elimination decreases drastically the number of regions. To avoid the creation of holes in the image, these regions are included in one of their adjacent regions. To minimize the corresponding distortion, the enclosing region is chosen as the adjacent region whose mean grey level is closest to that of the small region to be included.

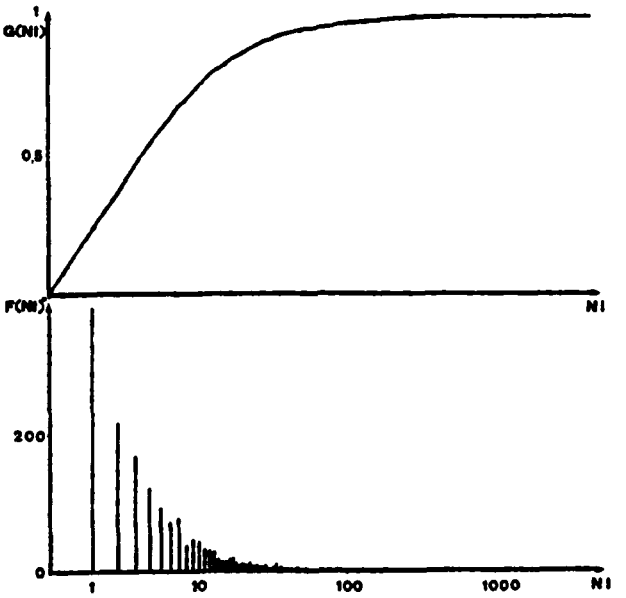


Fig. 22. Histogram and the cumulative histogram of the number of regions in the picture of the building from Fig. 25 as a function of their size.

By observing areas of constant luminance gradient in the pictures of Fig. 21 (the sky for example), it can be noticed that they are subdivided into regions even though there is no real contour. This is due to the property used in region growing which divides the image into regions of fixed grey-level dynamic range. The second possibility to decrease the number of regions is thus to merge adjacent regions whose contrast is below a certain level. The contrast between adjacent regions is defined as the mean grey level difference calculated along their common border. Notice that this procedure does not introduce any distortion. Fig. 23 shows the histogram and the cumulative histogram of the number of adjacent regions in the image of the building as a function of their common contrast. These

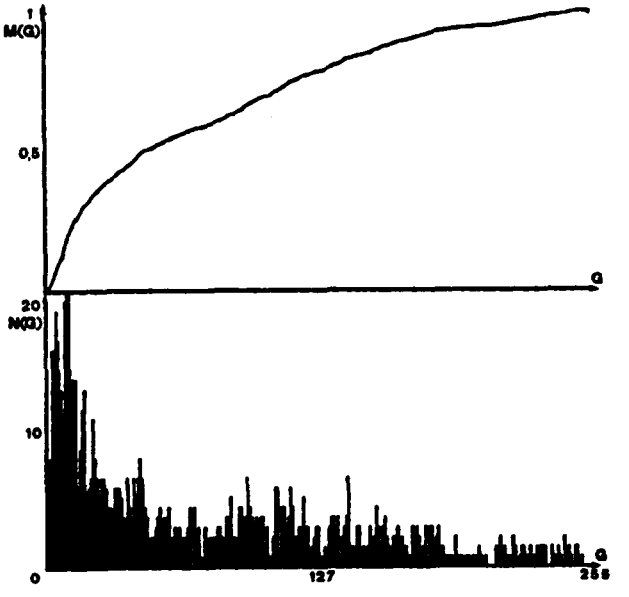


Fig. 23. Histogram and the cumulative histogram of the number of adjacent regions in the picture of the building from Fig. 25 as a function of their common contrast.

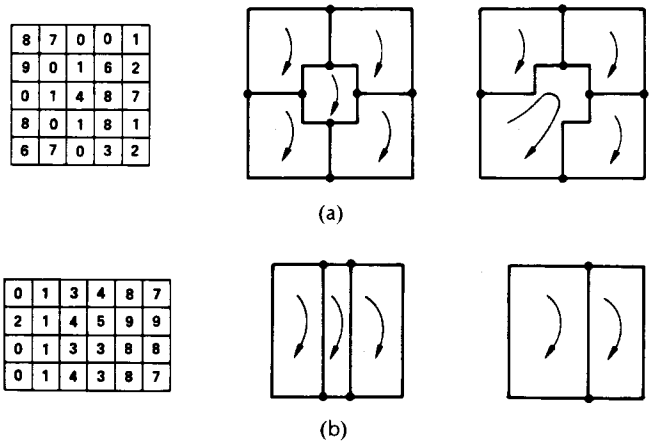


Fig. 24. Procedure for the elimination of small regions (a) and for merging adjacent regions whose local contrast is small (b).

curves indicate that roughly 35 percent of the region borders in the building image have a local contrast of less than 15 grey levels. If it is assumed that adjacent regions whose local contrast is lower than a threshold can be merged without introducing discontinuities, the number of regions is reduced further. The schematic views of these two elimination procedures are depicted in Fig. 24. The result of their application to segmented images of Fig. 21 is shown in Fig. 25. The number of regions are now 195 and 164 for the cameraman picture and the building picture, respectively.

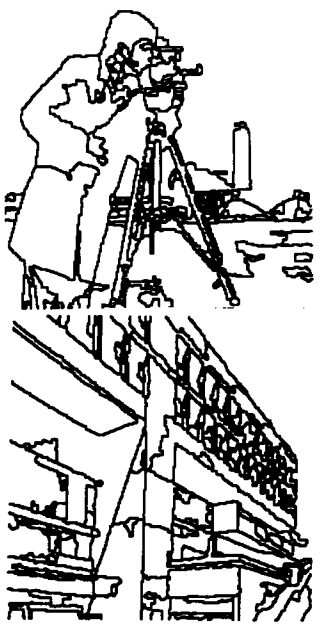


Fig. 25. Final results of the segmentation procedure.

The entire segmentation procedure has two interesting features: it is parametric in the sense that the number of regions can be fixed *a priori*, and there is no discontinuity in grey level within the regions which allows the use of global analytic functions (splines or polynomials) for their description. We are now investigating a region growing procedure using polynomial approximations over larger regions with better fidelity.

2) Contour Coding: Efficient description of the contours obtained after segmentation is a part of the messages to be coded. As discussed in Section III, a precise description of contours is essential for the human visual system. In this technique contour coding is carried out in two steps.

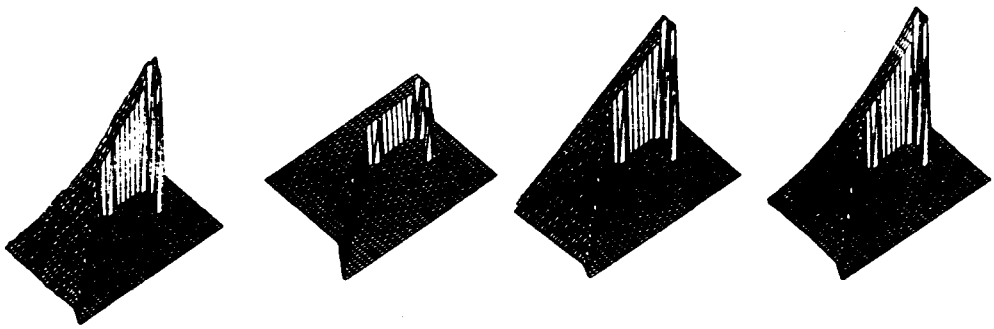
Since regions are closed, contour points along the border of two adjacent regions are described twice, once for each region. In the first step, these points are removed from one of the regions to be described and coded only once.

In the second step, remaining contour segments are described in a three-mode procedure [42]: 1) approximation by line segments, 2) approximation by circle segments, and 3) without approximation. Starting from the first contour point, the longest straight-line segment and circle segment are drawn under the constraint that the maximum error between the original data and their approximation does not exceed a given bound. The cost, associated with each mode, in terms of number of bits for coding, is evaluated. The "cheapest" mode is chosen. This procedure leads to an average of 1.6 bits per contour point for natural pictures if the error bound is 1.1 pels. Recently, a more refined code [43] without approximation has been proposed, using about 1.2 bits per contour point.

3) Texture Coding: The missing part of the messages after contour coding is texture coding. Note that within each region there is no longer any sharp discontinuity and hence the variation of the grey level within a region can be described with smooth two-dimensional polynomial functions. Texture coding is also carried out in two steps.

In the first step, the general shape of the grey level in each region is approximated by a two-dimensional polynomial function. The order of the polynomial is determined as a function of the approximation error and of the cost involved in coding polynomial coefficients. The approximation criterion used is the mean-square error which is minimized over each region for polynomials of order 0, 1, and 2. A three-dimensional view of these approximations is shown in Fig. 26. In this particular case, the best ("cheapest") approximation is obtained with a first-order polynomial function.

In the second step, the granularity removed with preprocessing is added back in the form of a pseudo-random noise to render the image more natural and less "painted by numbers." The mean-square error between the original image and the image reconstructed with polynomial functions is computed in each region. This error is used to control the variance of zero-mean Gaussian pseudo-random signal added as microtexture or "salt-and-pepper." Fig. 27



(a) (b) (c) (d)

Fig. 26. Approximation of the grey level of a region with two-dimensional polynomial functions. Original data (a), zero-order approximation (b), first-order approximation (c), and second-order approximation (d).

Fig. 27. Decoded pictures with compression ratio 50:1.

shows the final state of the decoded pictures with compressions around 50:1. Additional results of this method are discussed in the next section.

4) Remarks: The region growing based technique seems to be a powerful approach for high-compression image coding. Its weak aspect is the property used in region growing which leads to a rather large number of regions whose borders are not necessarily the contours of the objects in the image. We are currently investigating another property for region growing; namely, two-dimensional polynomial approximation, for obtaining a smaller number of larger regions with better quality. A number of special procedures were designed to combat unnecessary artifacts resulting from region growing. Also under investigation are better ways to analyze and synthesize texture.

B. Directional Decomposition Based Coding

1) General Remarks: In this method, heavy emphasis is placed on edge detection to preserve edge information in the best possible way. According to the properties of the human visual system discussed in Section III with reference to coding, features required for an edge detector are precision of the edge position, economy in the sense that the smallest possible set of representative edge points are detected, and the ability to reconstruct the original edge. It is not obvious that most of these edge detection operators are simultaneously optimal with respect to all these features. The commonly used point-based definition of edge elements leads to the detection of redundant information. The weakness of this kind of operator is visible in the enormous amount of work that has been done (edge thinning, edge tracking, etc.) to improve these results. That is why it is preferable to define an edge element (EE) to be a two-dimensional step function of a given width and given direction. For notational ease it will be represented by a vector $\vec{e}$ with:

$$|\vec{e}| = \ell \quad \text{and } a_1 < \arg(\vec{e}) < a_2. \tag{7}$$

The goal is to obtain a representation of the edges of an image by means of such edge elements. The set of edge elements that provides this representation constitute the edge element set (ES). An edge detection scheme which uses this representation and produces the necessary information for reconstructing the edges, satisfies all the requirements mentioned previously. An edge element subset (EES) is defined to be a subset of the edge element set containing edge elements confined to limited directions, bounded by two angles a_1 and a_2

$$S_k = \{\vec{e}_k \mid \|\vec{e}_k\| = \ell,\ \arg(\vec{e}_k) \in (a_1, a_2)\}. \tag{8}$$

With each EES we associate a principal direction corresponding to the bisector of the angle (a_1, a_2) and an orthogonal direction which is perpendicular to the principal direction.

Using these definitions, a directional image is defined to be an image containing edge elements belonging to only one edge element subset.

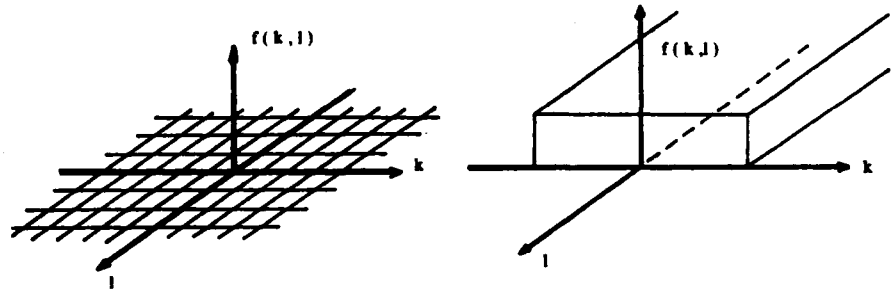
2) Directional Filtering: Directional filtering is based on the relationship between the presence of an edge in an image and its contribution to the image spectrum. It is largely motivated by the existence of direction-sensitive neurones in the human visual system. The discrete Fourier transform of an image $x(k, \ell)$ is given by

$$X(m,n) = \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} x(k,\ell) \exp[-j2\pi(mk + n\ell)/N]. \tag{9}$$

If the content of the spectrum along a line passing through the origin of the Fourier domain is examined, choosing (for computational ease) the line $X(0, n)$ we have

$$X(0,n) = \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} x(k,\ell) \exp[-j2\pi n\ell/N] = \sum_{\ell=0}^{N-1} \exp[-j2\pi n\ell/N] \sum_{k=0}^{N-1} x(k,\ell). \tag{10}$$

This equation can be viewed as the Fourier transform of a one-dimensional signal which represents the mean values of the image $x(k, \ell)$ along the k axis (see Fig. 28(a)). If there is an edge parallel to this axis (Fig. 28(b)), the signal corresponding to the second part of (10) is a step-like



(a) (b)

Fig. 28. Coordinate system to project an image along the k axis (a) and an ideal edge parallel to it (b).

function. Therefore, the contribution of the edge along the m axis of the Fourier domain may be deduced from the Fourier transform of this signal. In a similar way, the result for other directions can be found by turning the k axis to the left and to the right up to 90° and computing the projection of the image on the ℓ axis. The series of signals that is obtained is shown in Fig. 29. A theoretical aspect of

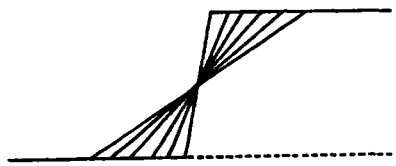


Fig. 29. Projections of an ideal edge for various directions.

the Fourier transforms of these signals may be obtained by approximating them by means of the following series of error functions:

$$g_n(z) = \frac{2\pi}{\sigma_n} A \int_{-z}^{z} \exp\left(-x^2/2\sigma_n^2\right) dx. \qquad (11)$$

They are shown in Fig. 30. The Fourier transforms of these

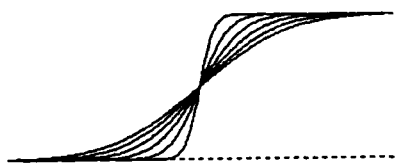
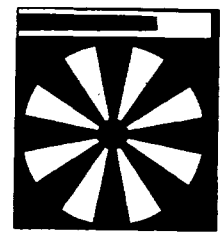


Fig. 30. Approximation for the signals of Fig. 33.

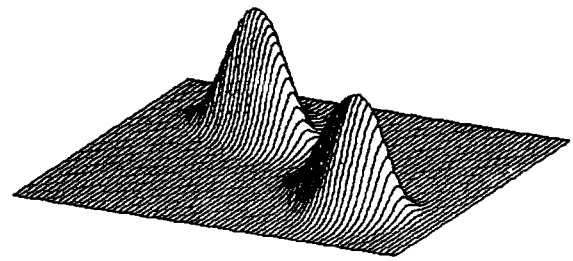
functions are given by

$$G_n(f) = \frac{A}{j2\pi f} \exp\left(-\sigma_n^2 f^2\right). \qquad (12)$$

It can be deduced from this equation that as σ_n^2 decreases, getting closer to the step function, the high-frequency content of the Fourier transform increases. The contribution of an edge is distributed all over the spectrum; the highest frequency component lies in the direction orthogonal to that of the edge, and as we turn away from this direction, the frequency of the contributions decreases, to vanish at 90°. Therefore, to detect an edge element subset, the sector of the spectrum that corresponds to the interval of directions of the EES should be retained. A filter whose frequency response covers a sector or a part of a sector in



(a)



(b)

Fig. 31. Sectors covered by directional filters (a) and a low-pass filter in the Fourier domain designed with a Gaussian window of large variance (b).

the frequency domain is called a directional filter. To make edge detection with these filters easier, high-pass filtering along the principal direction is introduced. Areas of the Fourier domain corresponding to these filters are shown in Fig. 31. The entire frequency plane is thus covered with n directional filters and one low-pass filter. The ideal frequency response of the ith directional filter is given by

$$H_i(f,g) = \begin{cases} 1, & \text{if } \left[\theta(i) < \tan^{-1}(g/f) < \theta(i+1)\right] \\ & \text{and if } \left(f^2 + g^2\right) > \rho_c \\ 0, & \text{otherwise} \end{cases} \qquad (13)$$

with

$$\theta(i) = (i-1)\pi/2n$$

$$\theta(i+1) = (i+1)\pi/2n$$

and

$$|f|, |g| < 0.5$$

where f and g are spatial frequencies, ρ_c is the cutoff frequency of the low-pass filter, and where a unity sampling step size is assumed.

Accordingly, a directional filter is a high-pass filter along its principal direction and a low-pass filter along the orthogonal direction. The bandwidth along this direction is

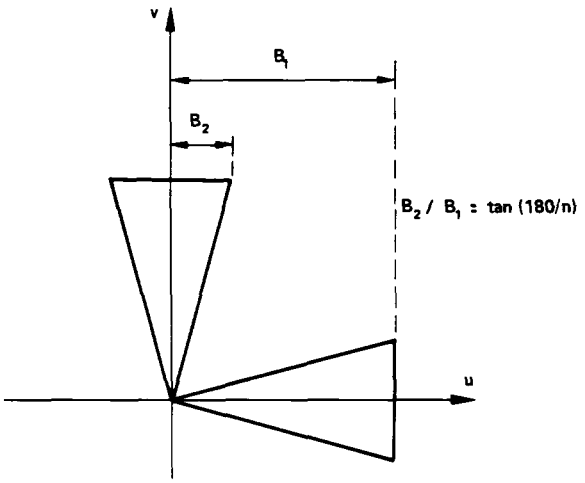


Fig. 32. Bandwidth of a directional filter.

$g\tan(\pi/2n)$ (see Fig. 32). Because of the Gibbs phenomenon, the ideal frequency response of the directional filters should be modified by an appropriate window function. The purpose is to avoid oscillation around zero crossings corresponding to real edges. One of the most appropriate window functions for this purpose is the Gaussian window given by

$$w(k,\ell) = \exp\left[-(k^2 + \ell^2)/s^2\right]. \qquad (14)$$

It can be shown [45] that, after windowing the filters and filtering, the superposition of all the directional images and the low-pass image leads to the original image. Thus the directional filtering, as defined, is an information preserving transformation. An example of directional decomposition with eight directional filters is shown in Fig. 33. There is,

Fig. 33. Directional decomposition in eight components and the resulting sum.

however, a compromise to be made. If the variance s^2 of the window function is too large, directional filters are more selective in their sector but quite poor as low-pass filters along the orthogonal direction. This introduces undesirable oscillations. In contrast, if this parameter is too small, these oscillations are considerably attenuated to the detriment of the directional selectivity. The same edge may appear in several directional images. A solution to this compromise may be found by using the following nonlinear transformation. It consists of first computing the sum of the values of the image points lying on the same position in all the directional images and retaining the point of the maximum value. Then, all the other points are set to 0 except the point of maximum value, which now takes the value of the sum. An illustration of this transformation is given in Fig. 34. This transformation considerably improves the directional selectivity and keeps the undesirable oscillations at an acceptable level.

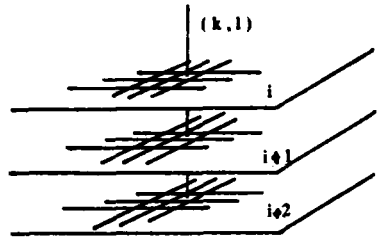


Fig. 34. Illustrating the nonlinear transformation to improve the directional selectivity.

There are two other parameters involved in the directional filters: their number and the cutoff frequency of the low-pass filter. The number of filters is directly related to the minimum width of edge elements that is accepted *a priori* in the image. Therefore, a direct way to define the number of filters (directions) is obtained by fixing the minimum length of accepted edge elements. The segmentation of curved edges (curvature quantization) resulting from the definition of edge elements is another fact that must also be taken into account. From a physiological point of view [25], it seems that the quantization of the directions is made by 20 to 30 different groups of cells, each one specialized to a limited interval of directions. However, to better define the range of acceptable quantization, experimental work is required. Since the choice of the cutoff frequency influences only the compression ratio and the quality of the decoded image, it will be discussed in the next section.

3) Coding: The messages to be coded are the directional images and the low-pass image. Note that the following scheme is not information lossless and that a certain quality degradation is assumed when coding. The main objective is to achieve the highest compression for a given degradation. The low-frequency component is, by its nature, adapted to transform coding. Criticism of transform coding was based on its weakness in coding edges. Since there are no edges in the low-frequency component, transform coding is the preeminently adequate method. As it will be shown, there is no degradation resulting from low-frequency component coding.

High-frequency images will be used for detecting and coding edges. A loss of information and hence a quality degradation is associated with this coding procedure. The loss of information comes from the inevitable choice between weak and strong edges. If the compression ratio is set to high rates, very weak edges must be eliminated. On the other hand, the indirect approximation of the edges by line segments, as assumed by the definition of the edge elements, introduces some degradations at the locations of high curvature. Edge detection in the directional images is based on the high-pass character of the directional filters along their principal direction. Filtering a signal with a high-pass filter gives zero crossings at the locations of abrupt changes (edges). Accordingly, edge detection in the

directional images is performed by searching the zero crossings along the principal direction of each image.

For ease of implementation, a normalization of all the principal directions on the horizontal direction is introduced. It is based on interpolation, rotation, and resampling. The details of this algorithm can be found in [48]. Edge detection in the normalized images is done by searching the zero crossings along the columns. The following parameters are used:

$$p_1 = x_i(k,\ell) \cdot x_i(k+1,\ell)$$
$$p_2 = x_i(k,\ell) \cdot x_i(k+2,\ell) \qquad (15)$$

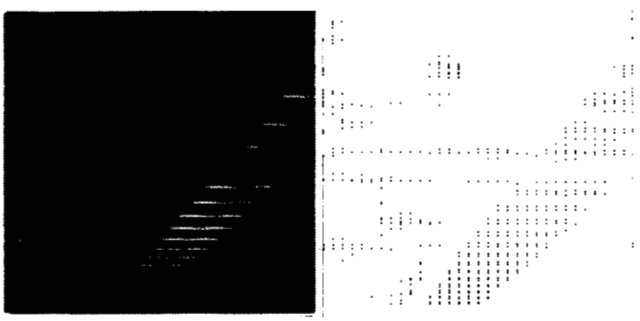
where $x_i(k,\ell)$ represents the ith directional image.

The existence of a zero crossing is established by the condition $p_1 < 0$ or $p_1 = 0$ and $p_2 < 0$. Another variable which must be used after this stage is the magnitude of the zero crossings, defined as

$$p_3 = |x_i(k,\ell) - x_i(k+k',\ell)| \qquad (16)$$

where k' equals 1 or 2 if $p_1 < 0$ or $p_1 = 0$ and $p_2 < 0$, respectively. By setting a threshold on this variable, it is possible to control the strength of the edges to be retained.

Because of the bandwidth reduction along the lines (orthogonal direction) and the subsampling, edge detection is applied to smaller directional images. Accordingly, each zero crossing detected corresponds to an edge element as defined by (7). Its direction belongs to the sector determined by the directional filter ($180/2n$) and its length in pels is equal to the subsampling rate. An example of zero crossing edge detection is given in Fig. 35. The first directional image of the set given in Fig. 33 is shown again in Fig. 35(a) whereas the zero crossings detected from this image



(a) (b)

Fig. 35. Zero crossing edge detection (b) in one directional image (a).

by applying the detection algorithm to every fifth column (subsampling rate 5:1), are plotted in Fig. 35(b). When edges in all the directional images are detected, the global result is as shown in Fig. 36. This image was obtained by rotating and superposing each directional image after edge detection. Fig. 37 shows the edge image obtained by interpolation between the zero crossings of each directional image before rotation and superposition. Notice that the representation of Fig. 36, which contains 1107 points, permits the reconstruction of the image of Fig. 37 which contains 5381 points. This result shows that the requirements of precise and economical detection are largely satisfied.

Fig. 36. Global result of zero crossing edge detection in eight directional images.

Fig. 37. Edges obtained after interpolation.

Each directional image is represented by the positions and the magnitudes of the zero crossings. The positions are coded with run-length coding using the Huffman code, requiring an average of 4.5 bits per position. The magnitudes of zero crossings are coded by trying different number of quantization levels. Exploiting the low sensitivity to contrast at high frequencies, a 3-bit code word was finally retained for coding the magnitudes of the zero crossings. The improvements obtained by using larger code words are not enough to justify the price payed. Smaller code words lead to objectionable distortions.

The low-frequency image can be coded in two equivalent ways. Since the maximum frequency of this component is much lower, it can be resampled using the two-dimensional sampling theorem and the resulting pixels can be coded by a standard procedure. The alternative is transform coding. The choice of the transform technique is directly dictated by the filtering that was used. The locations of the Fourier coefficients are known from the characteristics of the filter, and the importance of all these coefficients excludes any elimination by thresholding. This falls, therefore, in the category of zonal coding. After experimenting with several possibilities such as logarithmic quantization, bit-allocation plane, etc., the coefficients are quantized linearly. Fixed-length words are used to code the

phase and variable-length words, as attributed by the Huffman code, are used to code the magnitudes.

4) Decoding: In order to reconstruct the original image, all the components have to be decoded and added. The low-frequency component is obtained by inversely transforming the coded coefficients. The high-frequency component is obtained by synthesizing the directional images from the zero crossings. The synthesis of edge profiles from the zero crossing information and the interpolation between the columns of the normalized directional images, are the most critical procedures for the quality of the decoded image. The edge model, that was introduced in Section V-B1, offers the theoretical basis for the synthesis of the one-dimensional signals along the columns. Equation (11) requires two parameters: the magnitude A of zero crossings representing the contrast of the edge and the standard deviation related to the steepness of the slope of the edge. As the magnitude of zero crossings is coded, the only unknown parameter is the standard deviation. Two possible solutions were studied. The first consists of assuming a constant value for the standard deviation. This solution gives a standard, more-or-less step-like edge. The second consists of assuming a dependence of the standard deviation on the contrast. Experimental results indicate that a linear variation of the standard deviation with the contrast gives more realistic edges. The prototype wavelet which was adopted for approximating the profiles of zero crossings is the following:

$$g(u) = (u/Ak)\exp\left[-u^2/Ak\right] \qquad (17)$$

where u is the distance from the zero crossing at $u = 0$, A the magnitude, and k a constant. The family of these functions for different values of A is illustrated in Fig. 38.

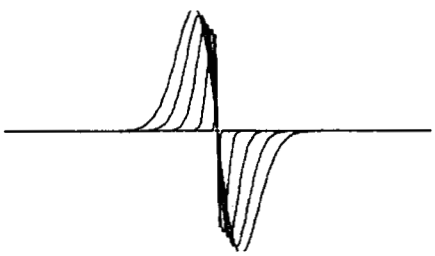


Fig. 38. Synthetic zero crossing profiles.

Note that the prototype wavelet of (17) is closely related to the derivative of (11).

Once the synthesis of zero crossing profiles is carried out at coded locations, the whole directional image is reconstructed by interpolation between the columns of the subsampled images. For a perfect interpolation between the columns of these images, the fact that the edge elements assumed by the presence of each zero crossing may have any direction within this interval must be taken into account. The interpolation algorithm looks for a neighboring point not only on the same line but also on the two preceding or following lines.

An example of coding the low- and high-frequency components is given in Fig. 39. Low- and high-frequency images

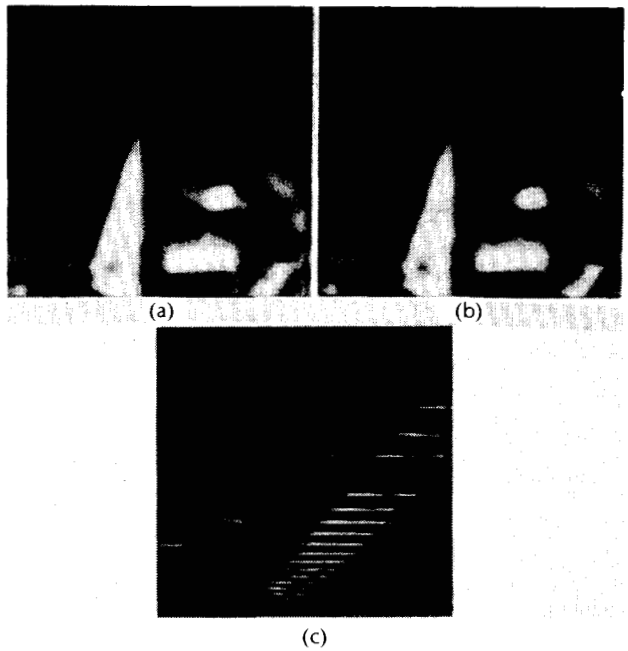


Fig. 39. The low-pass image before coding (a) and after coding (b). Decoded one directional image (c).

before coding are shown in Figs. 39(a) and 35(a), respectively. Their reconstructed versions after coding and decoding are shown in Fig. 39(b) and (c). The final result of decoding is given in Fig. 40 with a compression ratio of 41 : 1.

Fig. 40. An example of coded and decoded image with a compression ratio of 41 : 1.

VI. Simulation Results

All the decoded pictures which will be shown and discussed in this section are obtained with information-lossy methods. It would have been useful to derive a criterion to evaluate the distortions. As is well known, commonly used mean-square error criterion does not correspond to the subjective judgement of human observers. Although several attempts have been made to combine objective and subjec-

tive components into one mathematically tractable criterion, the results are not convincing, at least for the authors. For this reason, quality evaluation of the decoded pictures are left to the judgement of the reader. Coding methods have been applied to three original images shown in Fig. 41.

Fig. 41. Three original digitized pictures.

As mentioned earlier, these images are digitized with a 256 by 256 raster and each pel is quantized to 256 levels. Accordingly, their canonical form requires 524 288 bits for each of them. Compression ratios given in this section are the ratios of this number to the number of bits required for the coded versions. Results presented and discussed in this section are related to two first-generation methods and four second-generation methods described in Sections IV and V. Last but not least, errorless channel or storage is assumed for all the results.

A. First-Generation Methods

The purpose for giving here indicative results obtained with the first-generation techniques is not to provide comparison grounds and to disgrace these techniques, but to show, as a reminder, the type of results obtained in the early 1970s. Two examples of such results are given: one with DPCM coding and the other with transform coding. These two methods have been applied in the simplest possible way and all the refinements to improve the quality or the compression ratio by a few percent are omitted. These little improvements do not play an important role in this context.

Fig. 42 shows the image of the building after transform coding and decoding. Fourier transform and threshold coding are used to retain 2521 Fourier coefficients out of 32 768. The magnitude and the phase of each of them are coded with code words of average length of 5.44 and 5 bits, respectively. Addresses of the retained coefficients are coded with run-length coding requiring 3.08 bits per address. The compression ratio is therefore 15.4 : 1. Although

Fig. 42. An example of transform coded and decoded picture with a compression ratio of 15.4 : 1.

Fig. 43. An example of DPCM coded and decoded picture with a compression ratio of 3.7 : 1.

distortions are visible, the quality of this picture can be considered as very good.

An example of simple DPCM coded and decoded picture is shown in Fig. 43, for which the difference signal is quantized to 32 levels and coded with Huffman code. The compression ratio is 3.7 : 1. The price paid for the excellent quality of the result is the low compression ratio. Further increase of the compression leads to rapid deterioration of the image quality.

B. Pyramidal Coding

To generate the series of images $x_i(k,\ell)$ and $e_i(k,\ell)$, the low-pass filters and their cutoff frequencies must be fixed. A 5 by 5 separable weighting function is used for low-pass filtering. In one dimension, these weights are $w(0) = 0.7$, $w(-1) = w(1) = 0.25$, and $w(-2) = w(2) = 0.1$. The pyramid generated with these weights is shown in Fig. 44. The ratio of two cutoff frequencies $f(i)/f(i+1)$ in each level of the pyramid is fixed at 2. In this structure, each picture has four times fewer pels than its predecessor. Fig. 45 shows the results obtained with this technique with a compression ratio of about 6 : 1. A good quality is observed in all the results when compared to the original images (Fig. 41). If the compression is set at a higher ratio (16 : 1 or 32 : 1) degradations become visible due to successive decimations and interpolations, as shown in Figs. 46 and 47. These degradations may be eliminated at the expense of a higher computational cost.

C. Anisotropic Nonstationary Predictive Coding

This method, as described in Section IV-B, was also applied to our original images. Fig. 48 shows the results. The average compression ratio is around 20 : 1 with good-

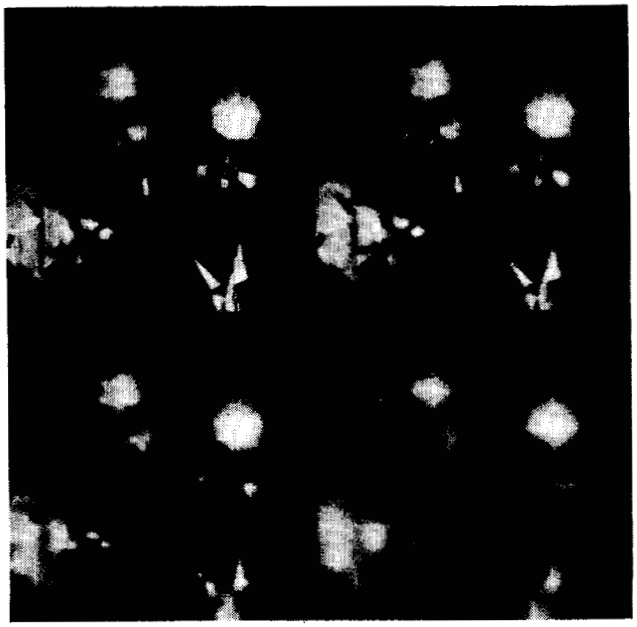


Fig. 44. First four levels of the Gaussian pyramid.

Fig. 45. Pyramidal coding results. The compression ratio is 6 : 1.

Fig. 46. Pyramidal coding results. The compression ratio is 16 : 1.

Fig. 47. Pyramidal coding results. The compression ratio is 32 : 1.

quality decoded images. Other examples with higher compression ratios but on larger images (512 by 512) can be found in [33].

D. Region Growing Based Coding

This method is a succession of parametric procedure as described in Section V-A. As in any parametric method, the ideal value of each parameter is data and goal dependent. The first step of this method is inverse gradient filtering to attenuate local granularity in the image. Its parameters are the size of the mask used in the convolution and the number of iterations. Experiments indicate that a 3 by 3 mask and 10 iterations are appropriate for most of the natural images. The second step is that of region growing, requiring a grey-level interval as the unique parameter. If its value is too small, a large number of regions is obtained decreasing the compression ratio. In contrast, if it is too large, the number of regions is small, leading to a high compression ratio. The price paid for this is the unacceptable distortions in the decoded image. In the third step, small regions are eliminated and strongly connected regions are merged using two threshold parameters. These parameters are the most important parameters of the method. The value of the first threshold on the region size depends on

Fig. 48. Anisotropic nonstationary predictive coding results.

the distortions accepted or tolerated. Experimental results show that regions containing more than 20 inner points should not be eliminated. The second threshold is on the gradient along the common border of two adjacent regions. These regions are merged if the average gradient is less than 15–20 grey levels compared to the full dynamic range of 256 levels. For high compression, the number of regions after segmentation should not exceed 100 or 110. The contours of these regions are coded using the new procedure described in [43], using 1.173 bits per contour point. Finally, texture components, deterministic as well as random, are coded depending on the order of the two-dimensional polynomial function used. The compression ratio for this method is given by

$$C = \frac{524\,288}{0.634P + 10N0 + 30N1 + 60N2} \qquad (18)$$

where P is the number of contour points, $N0$ the number of regions approximated with a zero-order polynomial, $N1$ the number of regions approximated with a first-order polynomial, and $N2$ the number of regions approximated with a second-order polynomial function. Three series of coded and decoded images are shown in Figs. 49–51 along with the corresponding compression ratios. For high compression, several defects appear. Many of them are due to the elimination of the small regions (lips of the cameraman, an ash tray on the wall of the building, etc.). Accordingly, this method is suitable for images which can be described by a small number of large regions. This is not the case of the couple image as reflected by the low compression ratios and poor quality of the results. There is no doubt that the quality of the decoded pictures can be improved if the property used for region growing is replaced by a more powerful one.

E. Directional Decomposition Based Coding

The compression ratio and the quality of an image coded with this scheme depend on the following parameters: the cutoff frequency of the filters, the subsampling rate, the zero crossing detection threshold, and the quantization of the low-frequency components. The value of the cutoff frequency determines the amount of detail preserved in the low-frequency component: the larger the cutoff frequency,

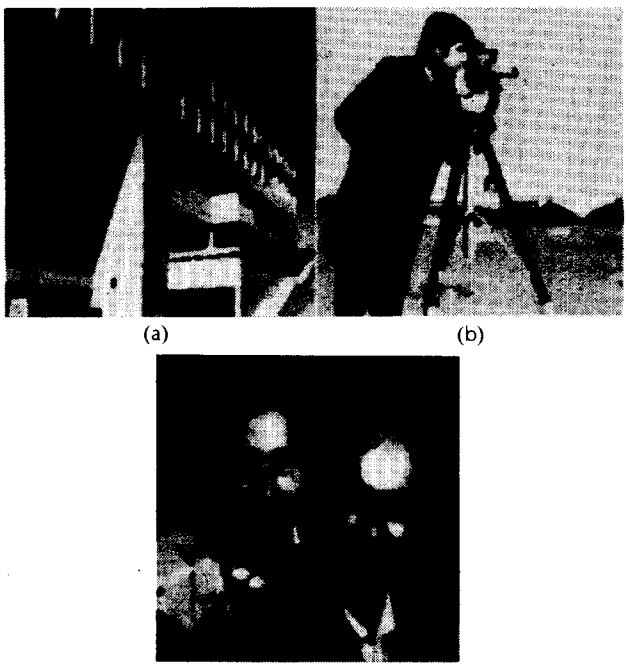


Fig. 49. Region growing based coding results. The parameters used are: (a) $P = 21\,210$, $N0 = 182$, $N1 = 59$, $N2 = 21$, and $C = 29:1$; (b) $P = 16\,448$, $N0 = 160$, $N1 = 38$, $N2 = 8$, and $C = 38:1$; (c) $P = 25\,352$, $N0 = 249$, $N1 = 91$, $N2 = 41$, and $C = 22:1$.

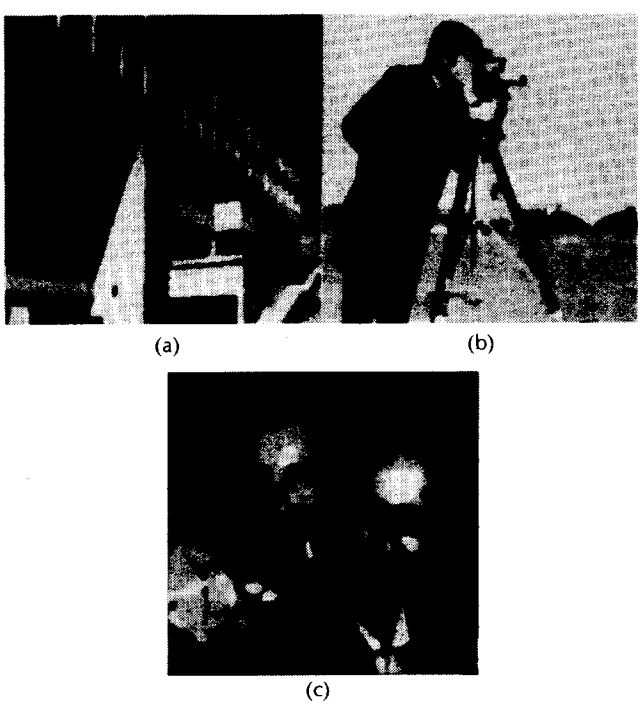


Fig. 50. Region growing based coding results. The parameters used are: (a) $P = 18\,046$, $N0 = 129$, $N1 = 40$, $N2 = 10$, and $C = 38:1$; (b) $P = 14\,590$, $N0 = 122$, $N1 = 35$, $N2 = 5$, and $C = 44:1$; (c) $P = 22\,350$, $N0 = 179$, $N1 = 72$, $N2 = 31$, and $C = 26:1$.

(a) (b)

(c)

Fig. 51. Region growing based coding results. The parameters used are: (a) $P = 15188$, $N0 = 96$, $N1 = 25$, $N2 = 5$, and $C = 45:1$; (b) $P = 11952$, $N0 = 98$, $N1 = 31$, $N2 = 3$, and $C = 60:1$; (c) $P = 19140$, $N0 = 131$, $N1 = 50$, $N2 = 24$, and $C = 32:1$.

the richer in detail the low-frequency component, the better the quality of the decoded image, but lower the compression ratio. The subsampling rate is directly related to the number of directional filters and to the length of the assumed edge elements. For the results presented here, the number of directional images is fixed at eight. The subsampling rate may vary from three to five resulting in more or less visible distortions in edge reconstruction. For a larger number of directional filters, the directional selectivity is improved at the level of zero crossing profile synthesis but the short edge elements are eliminated causing objectionable distortions.

The zero crossing detection threshold controls the importance of edges to be detected and reconstructed. Fig. 52 is a plot of the number of zero crossings as a function of the threshold, as computed in the picture of the building. The compression ratio for a given set of parameters is indicated for each point of this curve. For a threshold value larger than 30, important contours are eliminated causing visible distortions.

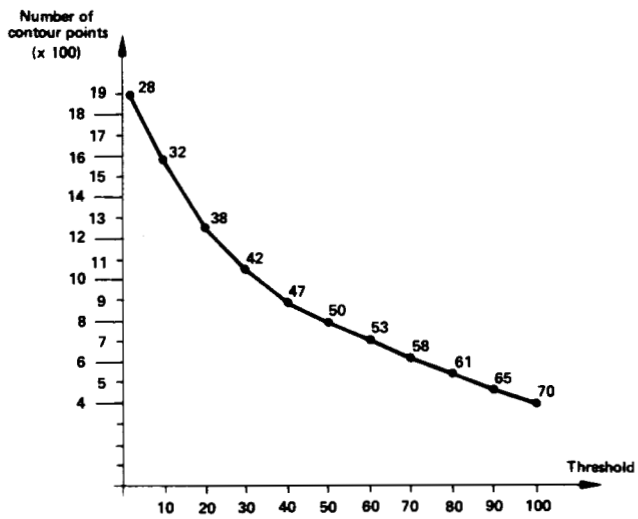


Fig. 52. The number of zero crossings as a function of the threshold.

The quality of the low-frequency component depends on the quantization of the magnitude and the phase of the coefficients to be coded. In Fig. 53, the root mean-square error between a low-pass image and its coded version is plotted as a function of the number of bits used to code the magnitude. As this plot indicates, in contrast with regular images, the subjective differences between the corresponding images are not as important. The low-pass image coded with 10 bits has no visible distortion and the one coded with 8 bits is acceptable. If the number of bits for the magnitude is kept constant (10 bits), and that for the phase is varied, a similar result is obtained, as depicted in Fig. 54.

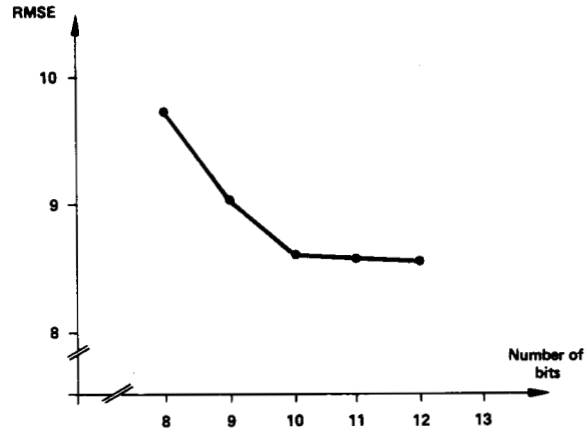


Fig. 53. Root mean-square error as a function of the number of bits used for coding the magnitudes of the Fourier coefficients.

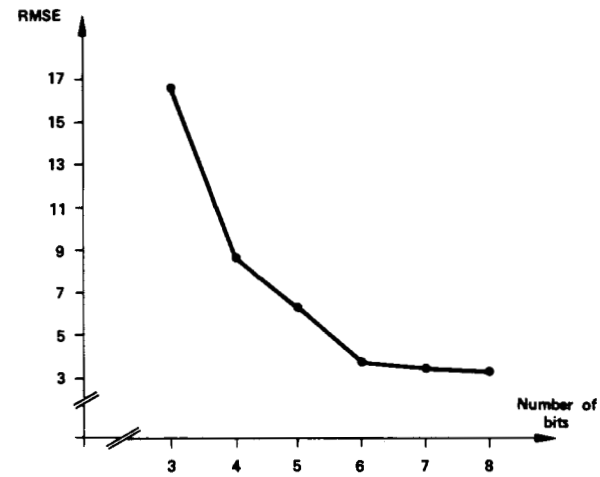


Fig. 54. Root mean-square error as a function of the number of bits used for coding the phase of the Fourier coefficients.

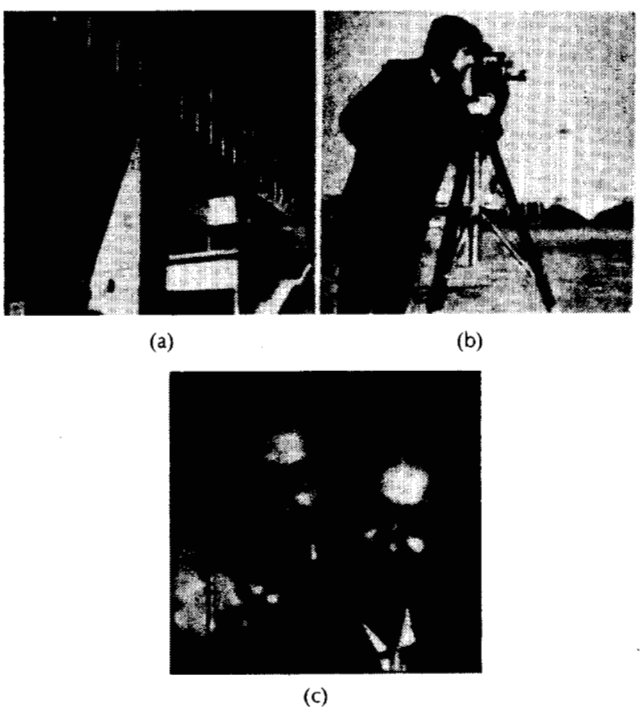
Using 5 bits for the phase, a perfect low-pass image is obtained. The result obtained with 4 bits is acceptable. A lower number of bits introduces important distortions.

The first series of results of decoded images obtained with a low compression ratio is shown in Fig. 55. The average compression ratio is around 20:1 and the quality of the picture is quite high. Comparing these results to those of Fig. 49, one can notice the difference of quality especially in the image of the couple. By decreasing the cutoff frequency and increasing the zero crossing detection threshold, a second series of results is obtained with higher compression ratios, as shown in Fig. 56. They illustrate a typical performance of this method. In this case, the compression ratio varies from 33:1 for the image of the building to 51:1 for the image of the couple.

One may wonder what an image looks like after compression with a very high compresssion ratio close to 100:1. By setting the important parameters beyond their acceptable limits, very high compression ratios can be obtained with rather poor quality results, as shown in Fig. 57. The elimination of important edges and distortions in the edge and low-frequency component reconstruction are quite visible. These results can be improved by the introduction of post-processing procedure after decoding, especially by better matching low- and high-frequency components.

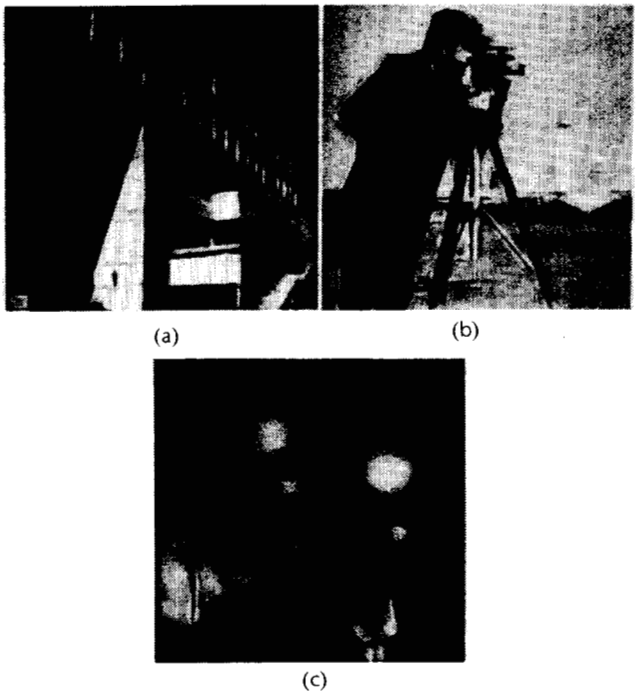
VII. Conclusions

In this paper a brief overview is first given of image-coding techniques using only (or mainly) statistical properties of the information source. The compression ratio achievable with these techniques has reached a saturation limit which does not correspond to the entropy of the source. This entropy, so far, is model dependent. Furthermore, information theory and coding theory do not take into account the properties of the receiver, i.e., the human visual system. These techniques are classified as first-generation techniques. They are characterized by the use of a



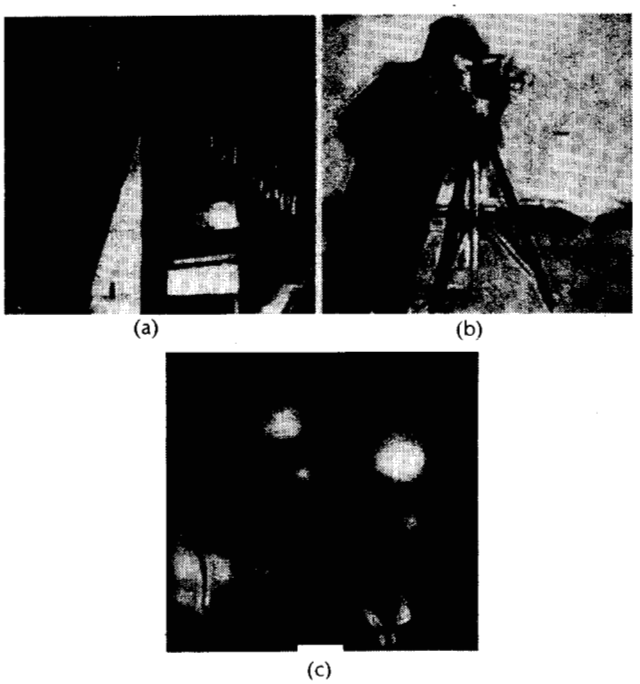
(a) (b) (c)

Fig. 55. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.19$, subsampling rate = 3, zero crossing detection threshold = 15, magnitude and phase quantized to 1024 and 32 levels, respectively. The compression ratios are 17:1 (a), 20:1 (b), 24:1 (c).



(a) (b) (c)

Fig. 56. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.12$, subsampling rate = 4, zero crossing detection threshold = 25, magnitude and phase quantized to 1024 and 32 levels, respectively. The compression ratios are 33:1 (a), 40:1 (b), and 51:1 (c).



(a) (b) (c)

Fig. 57. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.12$, subsampling rate = 5, zero crossing detection threshold = 50, magnitude and phase quantized to 256 and 16 levels, respectively. The compression ratios are 56:1 (a), 65:1 (b), and 92:1 (c).

signal processing approach in the selection of the messages to be coded and the application of the coding theory to code them.

Given the importance of the human visual system, recent neurophysiological results are discussed and a block diagram of this system is presented. These results suggest strongly the use of a general contour–texture model for image processing and coding. Techniques attempting to use this model are classified as second-generation techniques. It is interesting to note that the synthetic high system, designed 25 years ago, is the starting point for many of the second-generation techniques. These methods put heavy emphasis on the selection of the messages to be coded. In this context, signal processing approaches are not as successful as pattern recognition or artificial intelligence approaches. The coding is done in the classical way.

Four methods were described as second-generation techniques. The first two, namely pyramidal coding and anisotropic nonstationary predictive coding, as based on local operators designed in agreement with the processing done in the human visual system. They produce good-quality results at high compression ratios. The other two methods are based on contour–texture models. They need several improvements to produce better quality images at the same compression ratio or to reach higher compression ratios for the same quality. The goal to be reached is to segment the image into regions corresponding to the real objects of the scene, without missing small ones and without introducing false objects and hence false contours. Powerful representations should be designed to describe the grey level evolution within each region. Recent efforts in texture analysis and synthesis will be of great value to image coding to render the natural look when added to the representation of regions. It is hoped that image coding will remain a center of interest for researchers and that, one day, Descartes' limit will be reached.

ACKNOWLEDGMENT

Several individuals have contributed to our experience in picture coding and hence to this paper through various kinds of interactions. We would like to acknowledge contributions particularly from Prof. F. de Coulon, Prof. W. K. Schreiber, Prof. M. Eden, Prof. I. T. Young, Prof. G. H. Granlund, Prof. W. Frei, and Dr. A. N. Netravali, Dr. C. Rubinstein, Dr. R. Forcheimer, and Dr. P. Stucki.

Special thanks are due to Dr. P. J. Burt, Dr. E. H. Adelson, Dr. R. Wilson, Dr. H. Knutson, and Prof. G. H. Granlund for making their technique available to us to code our images.

REFERENCES

[1] M. Kunt, "Edge detection: A tutorial review," in *Proc. ICASSP 82* (Paris, France, May 2–5, 1982), pp. 1172–1176.

[2] S. Levialdi, "Edge extraction techniques," in *Fundamentals in Computer Vision*, O. D. Feugeras, Ed. Cambridge, England: Cambridge Univ. Press, 1983, pp. 117–144.

[3] R. M. Haralick, "Statistical and structural approaches to textures," *Proc. IEEE*, vol. 67, pp. 786–804, May 1979.

[4] A. Rosenfeld, "Survey, picture processing 1982," *Comput. Vis. Graph. Image Process.*, vol. 22, pp. 339–387, June 1983.

[5] ———, "Survey, picture processing 1981," *Comput. Vis. Graph. Image Process.*, vol. 19, pp. 35–76, May 1982.

[6] H. Wechler, "Texture analysis—A survey," *Signal Process.*, vol. 3, pp. 271–282, July 1980.

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

[8] M. Unser, "Caractérisation statistique de texture: application à l'inspection automatique," Ph.D. dissertation, Dep. Elec. Eng., Swiss Federal Inst. Technol., Lausanne, Switzerland, June 1984.

[9] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE*, vol. 63, pp. 366–406, Mar. 1980.

[10] A. K. Jain, "Image data compression: A review," *Proc. IEEE*, vol. 69, pp. 349–389, Mar. 1981.

[11] L. G. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Informat. Theory*, vol. IT-8, pp. 145–154, Jan. 1962.

[12] *Proc. IEEE* (Special Issue on Digital Encoding of Graphics), A. N. Netravali, Guest Ed., vol. 68, pp. 757–944, July 1980.

[13] A. Habibi, "Hybrid coding of pictorial data," *IEEE Trans. Commun.*, vol. COM-22, pp. 614–624, May 1974.

[14] W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs, an experimental TV bandwidth reduction system," *J. SMPTE*, vol. 68, pp. 525–537, Aug. 1959.

[15] W. F. Schreiber, "The mathematical foundation of the synthetic highs system," *MIT, RLE Quart. Progr. Rep.*, no. 68, p. 140, Jan. 1963.

[16] J. W. Pan, "Reduction of information redundancy in pictures," Sc.D. dissertation, MIT, Dep. Elec. Eng., 1966.

[17] D. N. Graham, "Image transmission by two-dimensional contour coding," Ph.D. dissertation, MIT, Dep. Elec. Eng., 1966.

[18] G. Walpert, "Image bandwidth compression by detection and coding of contours," Ph.D. dissertation, MIT, Dep. Elec. Eng., 1970.

[19] W. F. Schreiber, "Picture coding," *Proc. IEEE*, vol. 55, pp. 320–330, Mar. 1967.

[20] W. F. Schreiber, T. S. Huang, and O. J. Tretiak, "Contour coding of images," in *Picture Bandwidth Compression*, T. S. Huang and O. J. Tretiak, Eds. New York: Gordon and Breach, 1972.

[21] J. K. Tan and D. J. Sakrison, "Encoding of images based on a two-component source model," *IEEE Trans. Commun.*, vol. COM-25, pp. 1315–1322, Nov. 1977.

[22] B. R. Hunt, "Nonstationary statistical image models (and their application to image data compression)," *Comput. Graph. Image Process.*, vol. 12, pp. 173–186, Feb. 1980.

[23] D. E. Troxel *et al.*, "A two-channel picture coding system: I—Real-time implementation," *IEEE Trans. Commun.*, vol. COM-29, pp. 1841–1849, Dec. 1981.

[24] W. F. Schreiber and R. R. Buckley, "A two-channel picture coding system: II—Adaptive companding and color coding," *IEEE Trans. Commun.*, vol. COM-29, pp. 1849–1858, Dec. 1981.

[25] D. H. Hubel and T. N. Wiesel, "Brain mechanisms of vision," *Sci. Amer.*, vol. 241, pp. 150–162, Sept. 1979.

[26] S. W. Kuffler and J. G. Nicholls, *From Neuron to Brain*. Sunderland, MA: Sinauer Assoc., 1976.

[27] I. Aleksander, T. J. Stonham, and R. A. Wilkie, "Computer vision systems for industry," *Digital Syst. Indust. Automat.*, vol. 1, no. 4, pp. 305–320, 1982.

[28] D. E. Pearson, *Transmission and Display of Pictorial Information*. New York: Wiley, 1975, pp. 38–42.

[29] T. N. Cornsweet, *Visual Perception*. New York, Academic Press, 1970, pp. 270–310.

[30] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, Apr. 1983.

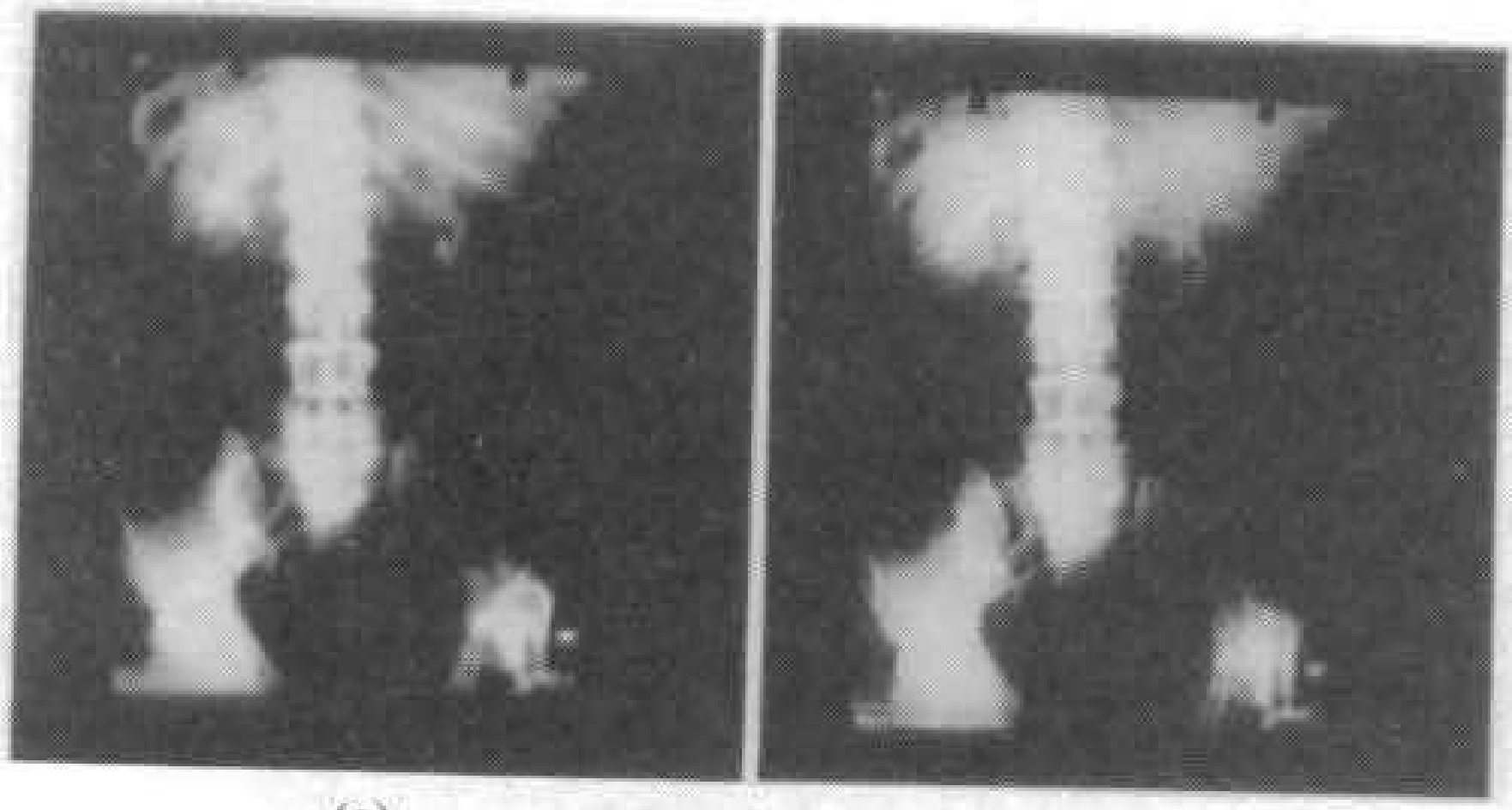
[31] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, B 207, pp. 187–217, Mar. 1980.

[32] D. Marr, *Vision*. San Fransisco, CA: H. Freeman and Co., 1982.

[33] R. Wilson, H. E. Knutsson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its applications: Part II—Predictive image coding," *IEEE Trans. Commun.*, vol. COM-31, pp. 398–406, Mar. 1983.

[34] H. E. Knutsson, R. Wilson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its applications: Part I—Restoration of noisy images," *IEEE Trans. Commun.*, vol.

COM-31, pp. 388–397, Mar. 1983.
[35] G. H. Granlund, "In search of a general picture processing operator," *Comput. Graph. Image Process.*, vol. 8, pp. 155–173, Oct. 1978.
[36] ——, "GOP—A fast flexible processor for image analysis," in *Proc. IEEE Conf. Pattern Recognition* (Miami, FL, 1980), pp. 489–492.
[37] M. Kocher and M. Kunt," A contour-texture approach to picture coding," in *Proc. ICASSP-82* (Paris, France, May 1982), pp. 436–440.
[38] M. Kocher, "Codage d'images à haute Compression basé sur un modèle Contour-texture," Ph.D. dissertation, no. 476, Dep. Elec. Eng., Swiss Federal Inst. Technol., Lausanne, Switzerland, Mar. 1983.
[39] M. Kocher and M. Kunt, "Image data compression by contour-texture modelling," in *SPIE Int. Conf. on the Applications of Digital Image Processing* (Geneva, Switzerland, Apr. 1983), pp. 131–139.
[40] ——, "A contour-texture approach to picture coding," in *Proc. Melecon-83* (Athens, Greece, May 24–26), 1983, paper C2.03.
[41] D. C. C. Wang and A. H. Wagnucci, "Gradient inverse filtering weighted smoothing scheme and the evaluation of its performance," *Comput. Graph. Image Process.*, vol. 15, pp. 167–181, Feb. 1981.
[42] P. Faeh and M. Kunt, "Efficient coding of high resolution typographic characters," in *Proc. ICASSP-82* (Paris, France, May 1982), pp. 440–443.
[43] M. Eden and M. Kocher, "On the performance of a contour coding algorithm in the context of image coding. Part II: Coding and contour graphs," *Signal Process.*, vol. 8, May 1985, to be published.
[44] A. Ikonomopoulos and M. Kunt, "Image coding based on a directional decomposition," in *Proc. Int. Picture Coding Symp.* (Davis, CA, Mar. 28–30, 1983), pp. 13–15.
[45] ——, "Directional filtering, zero crossing, edge detection and image coding," in *Proc. EUSIPCO-83* (Erlangen, RFA, Sept. 1983), pp. 203–206.
[46] A. Ikonomopoulos, M. Kocher, and M. Kunt, "Image coding based on human visual system properties for optimal reduction of redundancy," in *Proc. 3rd Scandinavian Conf. on Image Analysis* (Copenhagen, Denmark, July 12–14, 1983), pp. 216–222.
[47] M. Kunt, A. Ikonomopoulos, and M. Kocher, presented at the "Compression d'images: Méthodes de la deuxième génération" (invited lecture), Premièr Colloque GRETSI–CESTA, Biarritz, France, May 21–25, 1983.
[48] A. Ikonomopoulos and M. Kunt, "High compression image coding via directional filtering," *Signal Process.*, vol. 8, no. 3, May 1985, to be published.



(a) (b)

Fig. 4. An original X-ray picture (a) and its reconstructed version obtained with the synthetic high system (b). Compression ratio: $C = 8.37$.

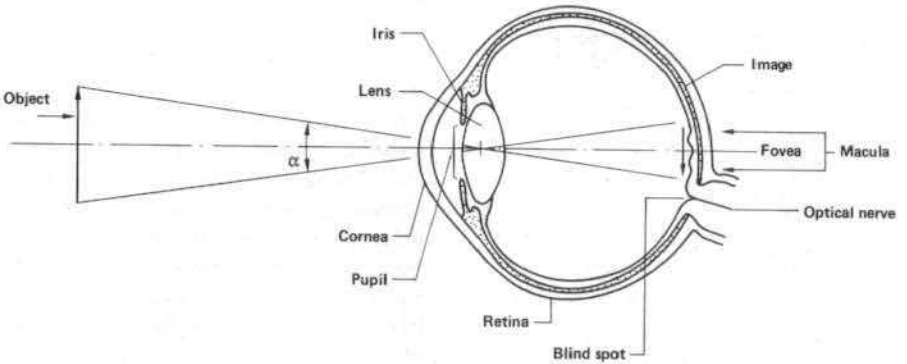
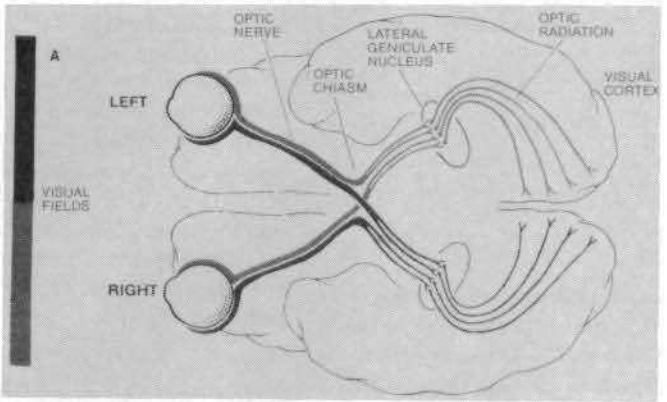


Fig. 7. Schematic view of the eye.



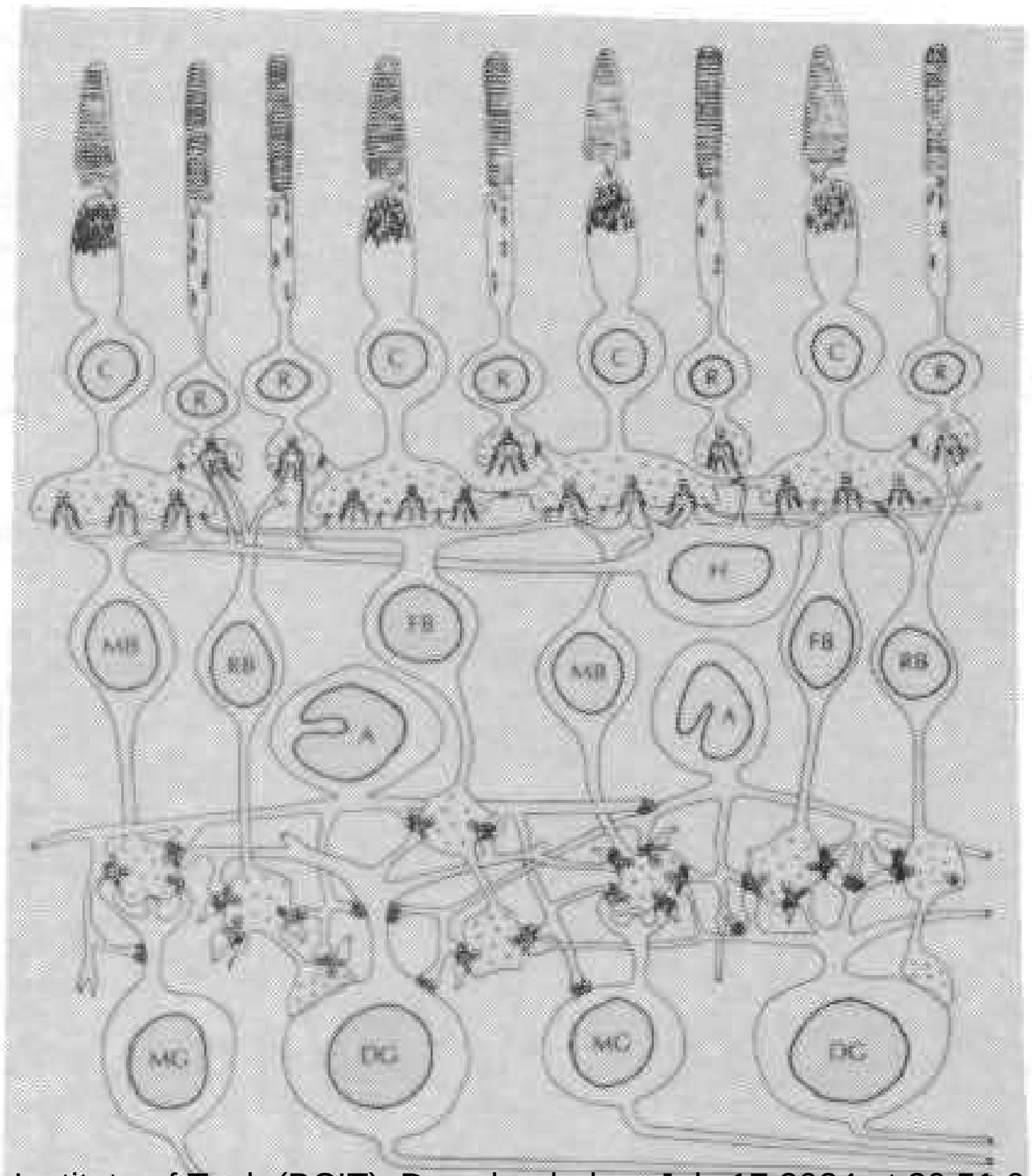


Fig. 9. Cells of the retina (after [29]).

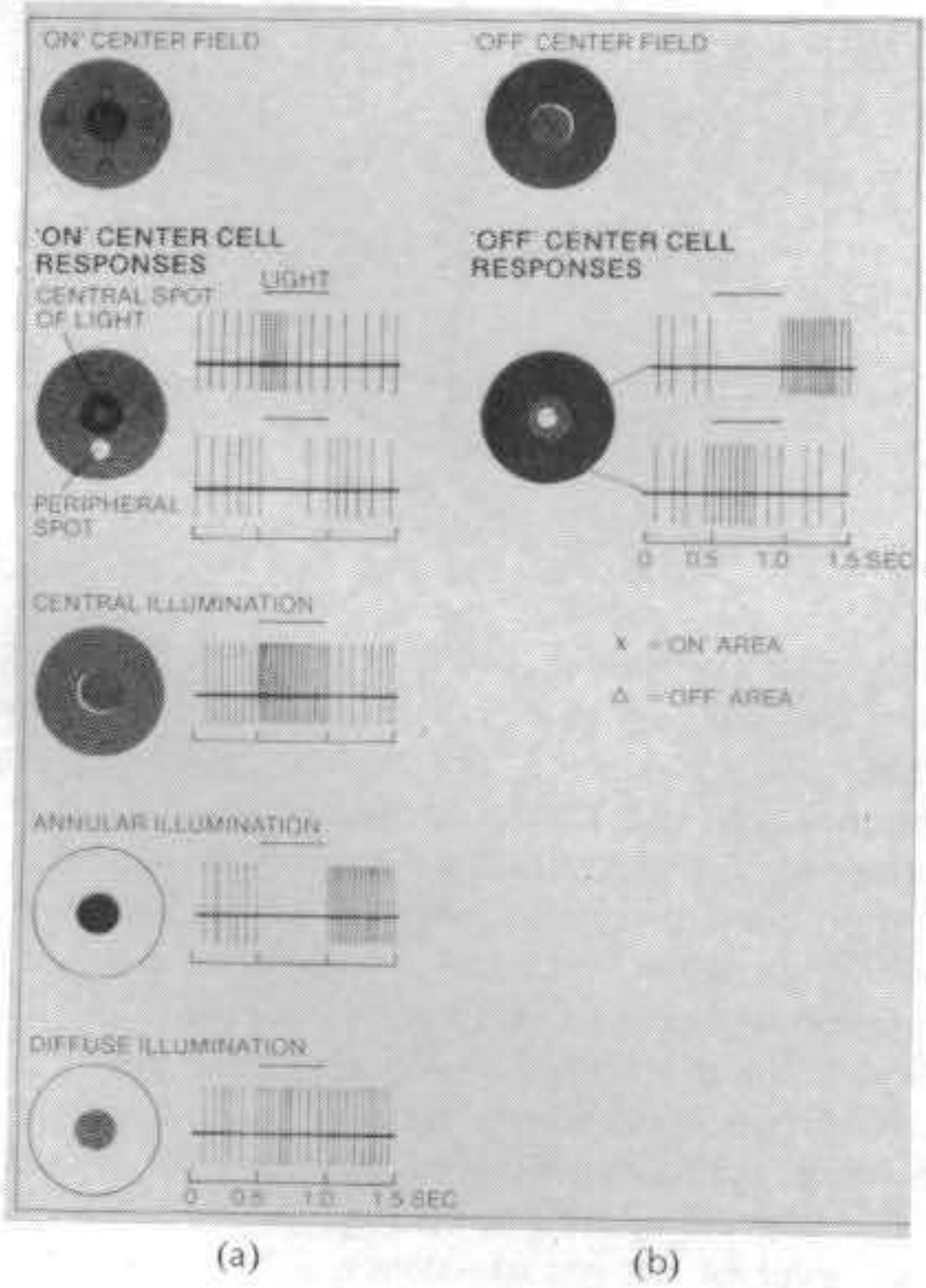


Fig. 13. Receptive fields of ganglion cells. Center on surround off (a) and center off surround on (b) (after [26]).

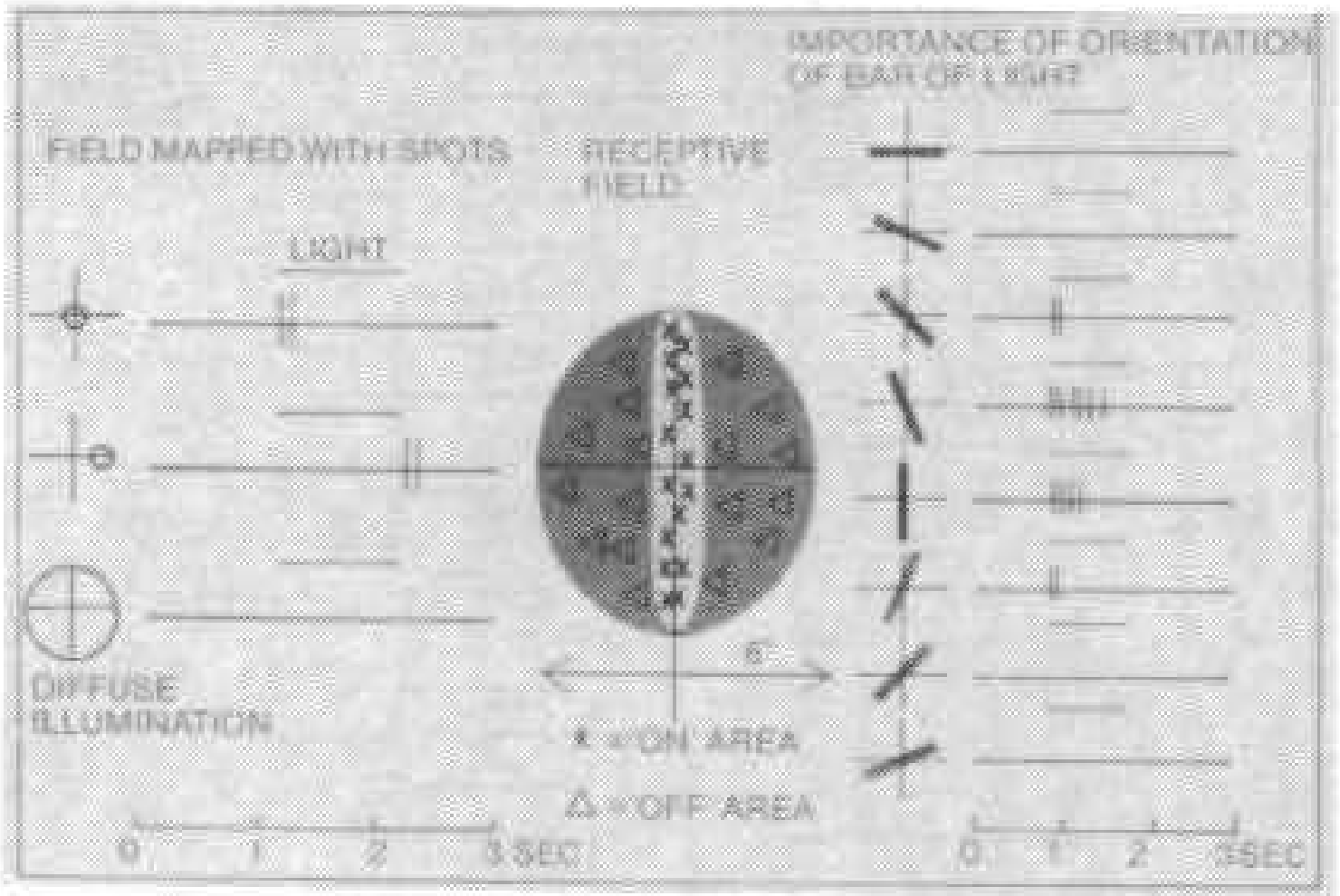
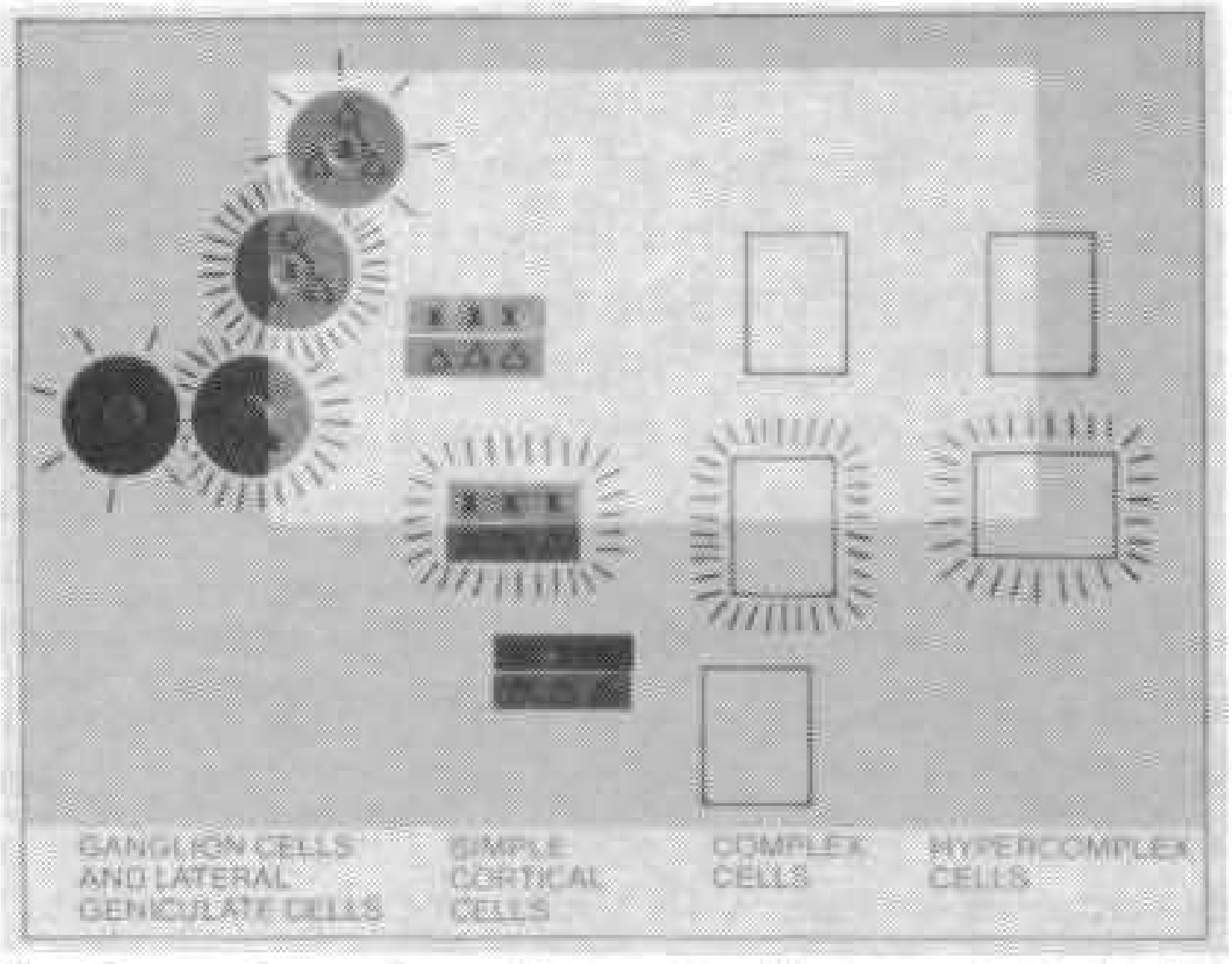


Fig. 14. Receptive field of a simple cell (after [26]).



ground by cortical cells (after [26]).

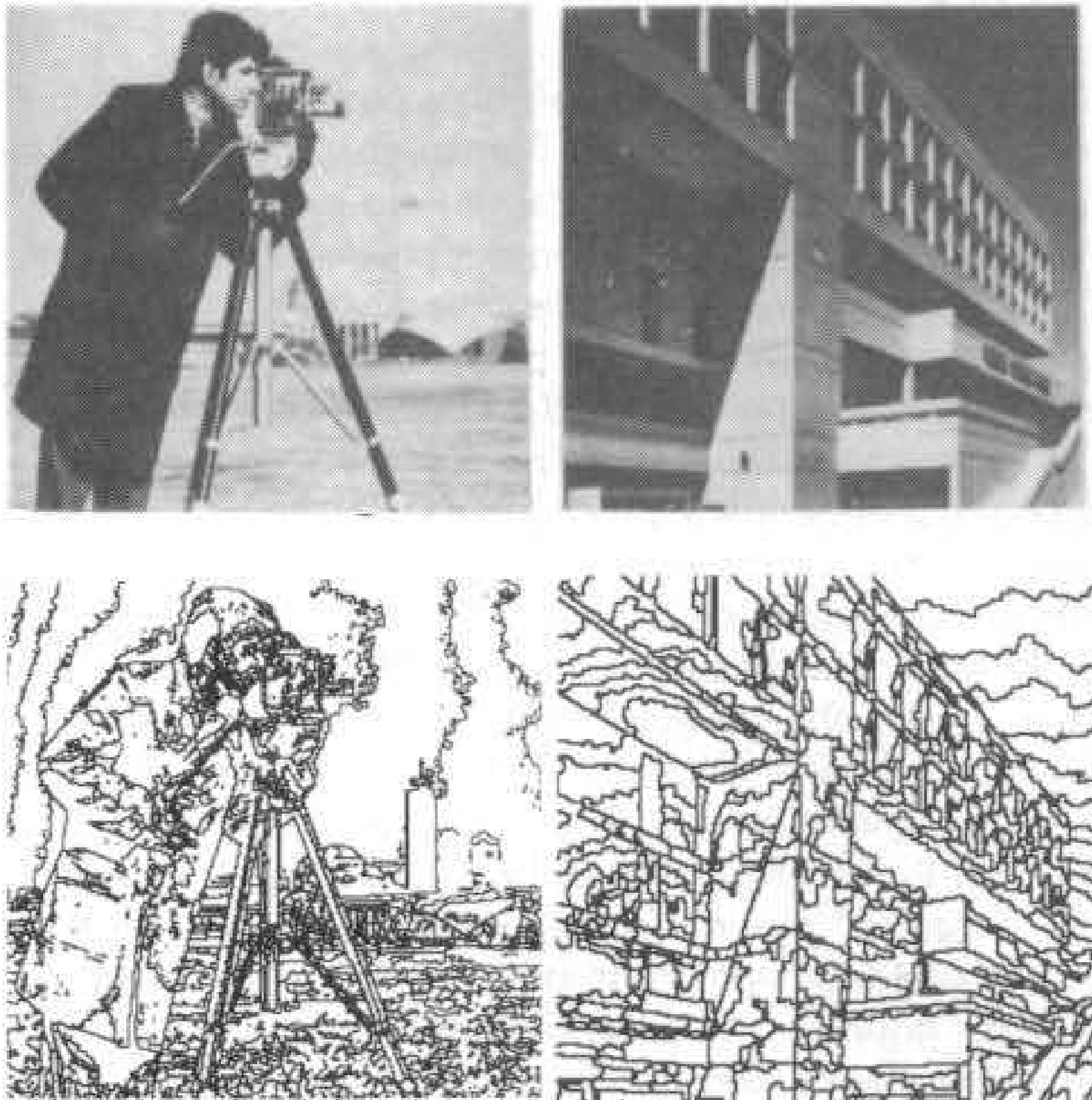


Fig. 21. Two original images and the result of region growing with a grey-level interval of 10 in both cases.

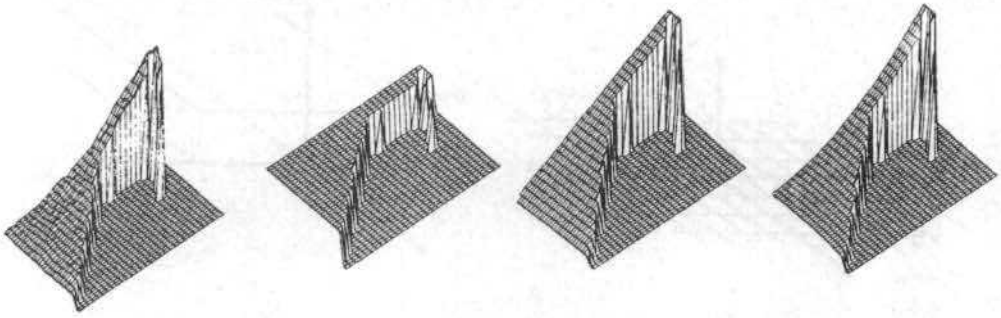
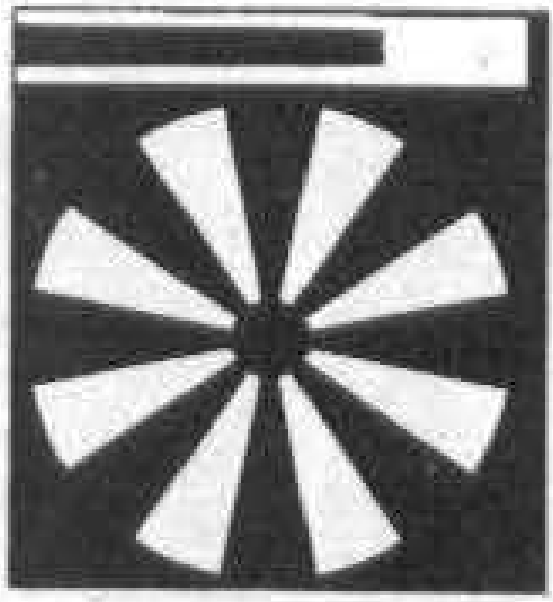
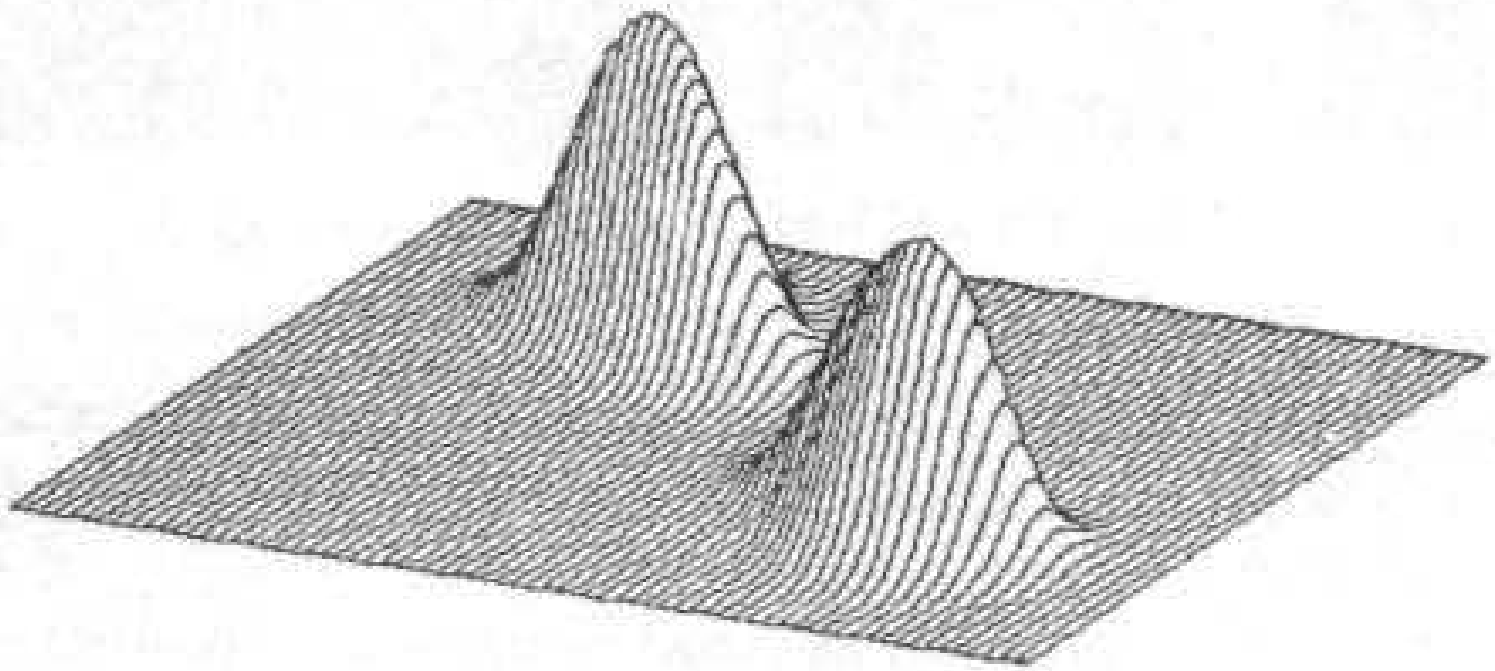


Fig. 26. Approximation of the grey level of a region with two-dimensional polynomial functions. Original data (a), zero-order approximation (b), first-order approximation (c), and second-order approximation (d).

Fig. 27. Decoded pictures with compression ratio 50 : 1.



(a)



(b)

Fig. 31. Sectors covered by directional filters (a) and a low-pass filter in the Fourier domain designed with a Gaussian window of large variance (b).

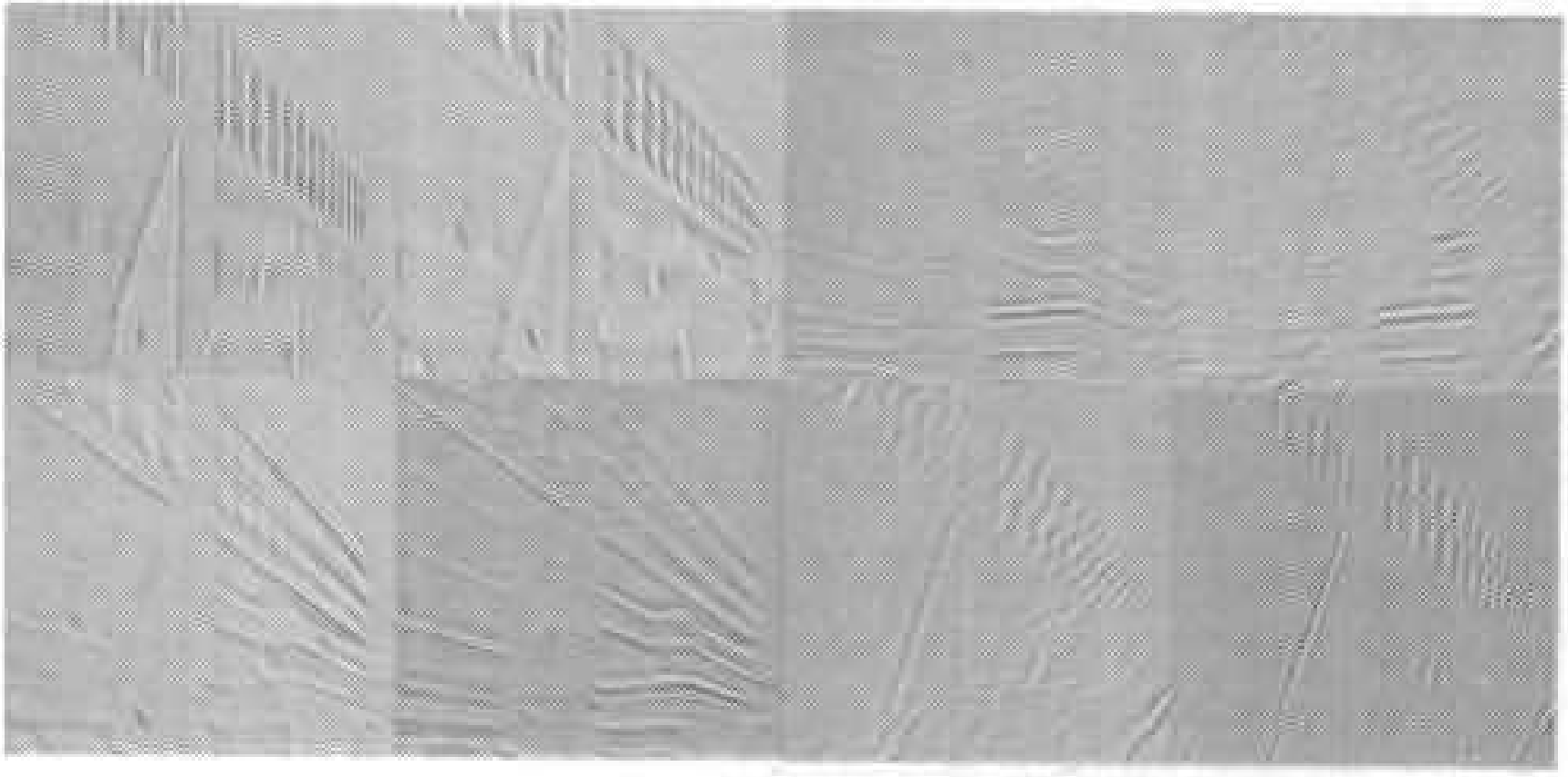
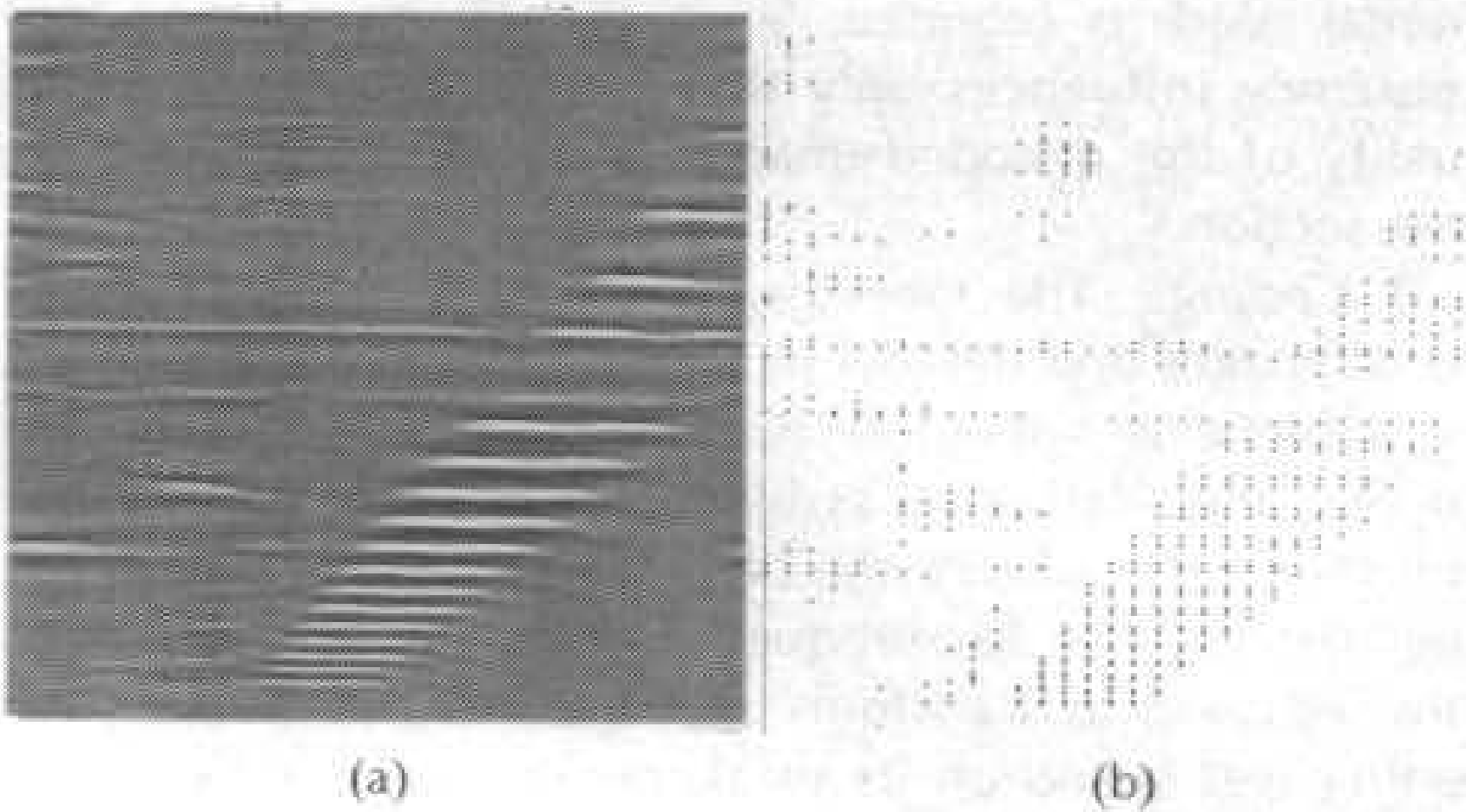


Fig. 33. Directional decomposition in eight components and the resulting sum.



(a) (b)

Fig. 35. Zero crossing edge detection (b) in one directional image (a).

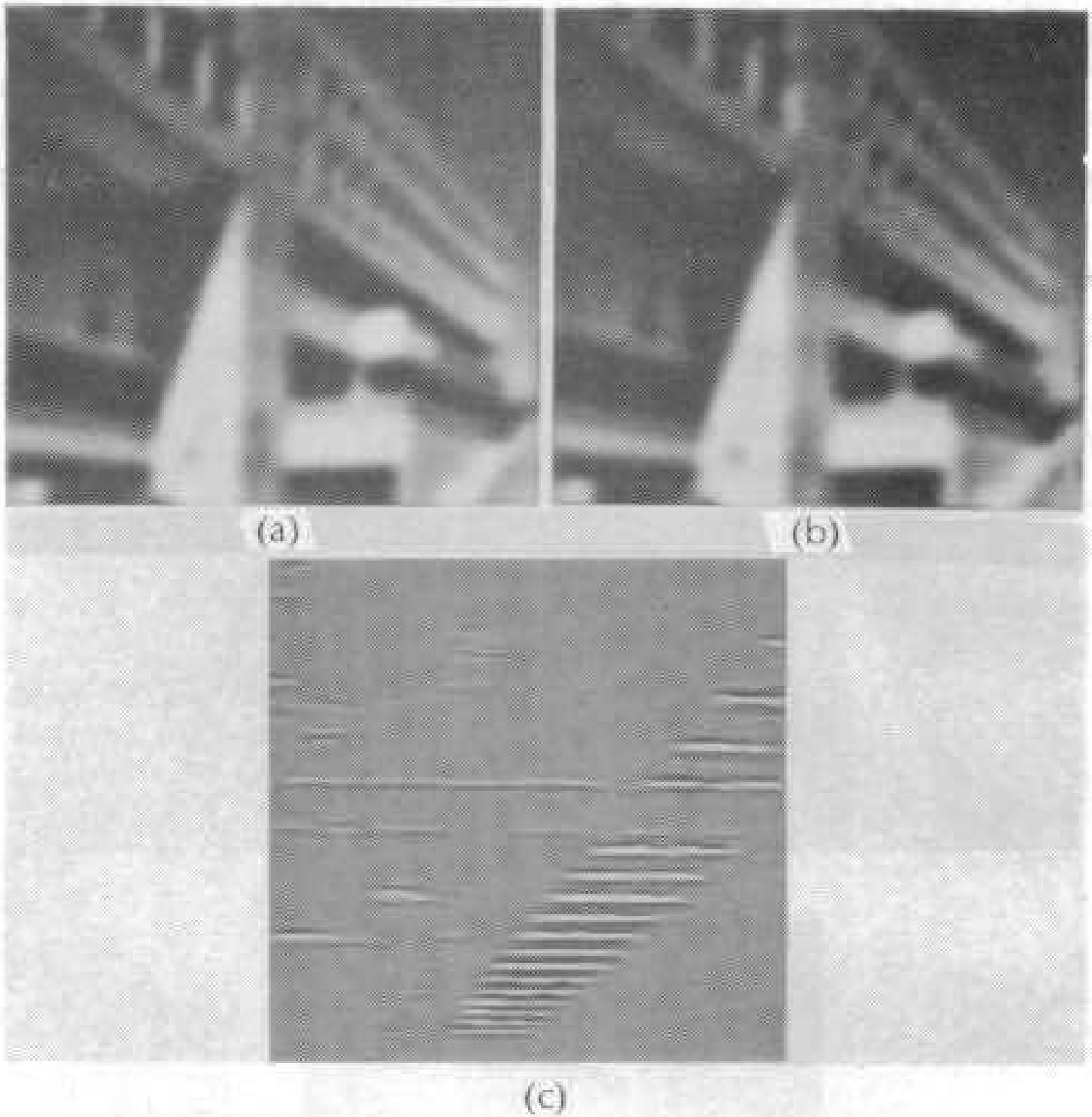


Fig. 39. The low-pass image before coding (a) and after coding (b). Decoded one directional image (c).

Fig. 41. Three original digitized pictures.

Fig. 42. An example of transform coded and decoded picture with a compression ratio of 15.4 : 1.

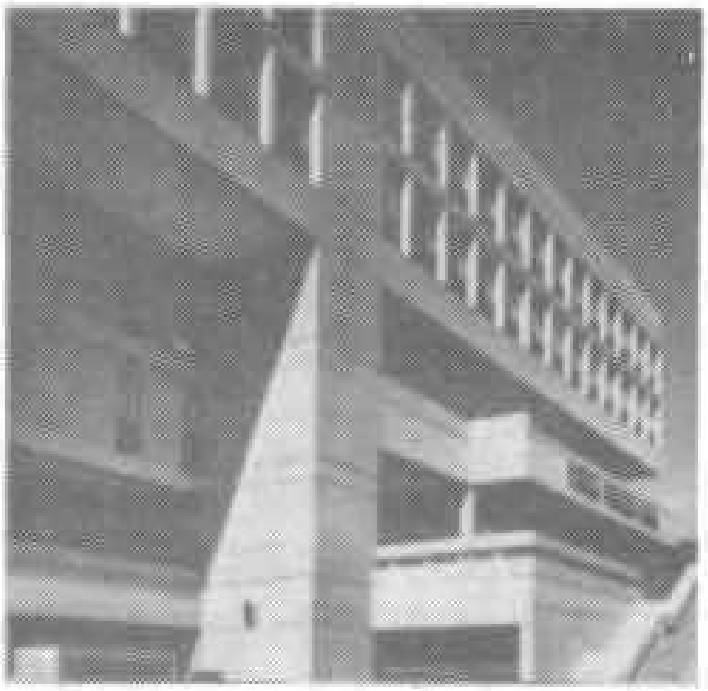


Fig. 43. An example of DPCM coded and decoded picture with a compression ratio of 3.7 : 1.

Fig. 44. First four levels of the Gaussian pyramid.

Fig. 45. Pyramidal coding results. The compression ratio is 6:1.

Fig. 46. Pyramidal coding results. The compression ratio is 16:1.

Fig. 47. Pyramidal coding results. The compression ratio is 32:1.

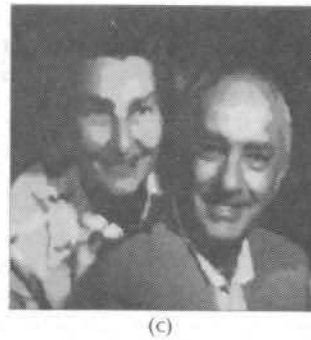
Fig. 48. Anisotropic nonstationary predictive coding results.

(a) (b)

(c)

Fig. 49. Region growing based coding results. The parameters used are: (a) $P = 21\,210$, $N0 = 182$, $N1 = 59$, $N2 = 21$, and $C = 29:1$; (b) $P = 16\,448$, $N0 = 160$, $N1 = 38$, $N2 = 8$, and $C = 38:1$; (c) $P = 25\,352$, $N0 = 249$, $N1 = 91$, $N2 = 41$, and $C = 22:1$.

(a) (b)



(c)

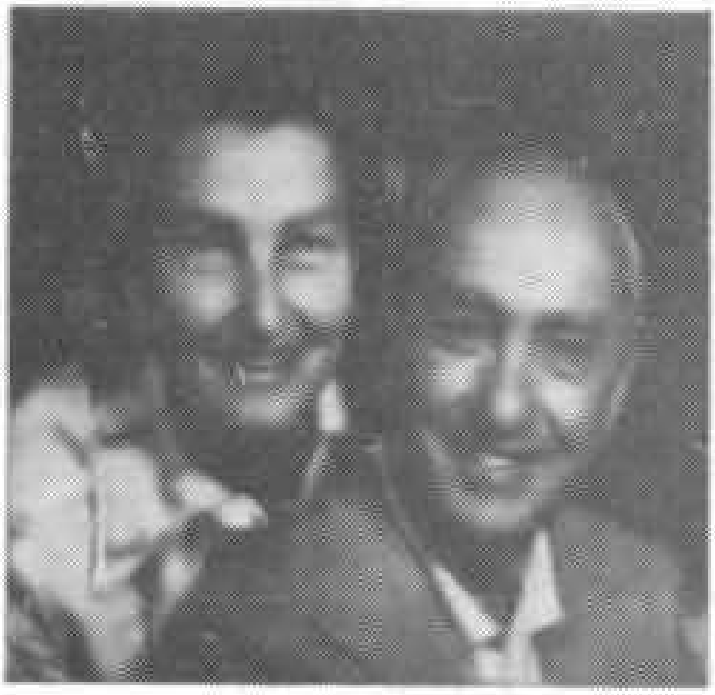
Fig. 50. Region growing based coding results. The parameters used are: (a) $P = 18\,046$, $N0 = 129$, $N1 = 40$, $N2 = 10$, and $C = 38:1$; (b) $P = 14\,590$, $N0 = 122$, $N1 = 35$, $N2 = 5$, and $C = 44:1$; (c) $P = 22\,350$, $N0 = 179$, $N1 = 73$, $N2 = 24$, and $C = 26:1$.

(a) (b)

(c)

Fig. 51. Region growing based coding results. The parameters used are: (a) $P = 15\,188$, $N0 = 96$, $N1 = 25$, $N2 = 5$, and $C = 45:1$; (b) $P = 11\,952$, $N0 = 98$, $N1 = 31$, $N2 = 3$, and $C = 60:1$; (c) $P = 19\,140$, $N0 = 131$, $N1 = 50$, $N2 = 24$, and $C = 32:1$.

(a) (b)



(c)

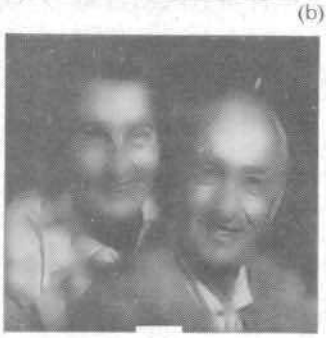
Fig. 55. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.19$, subsampling rate = 3, zero crossing detection threshold = 15, magnitude and phase quantized to 1024 and 32 levels, respectively. The compression ratios are 17:1 (a), 23:1 (b), 24:1 (c).

(a) (b)

(c)

Fig. 56. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.12$, subsampling rate = 4, zero crossing detection threshold = 25, magnitude and phase quantized to 1024 and 32 levels, respectively. The compression ratios are 33 : 1 (a), 40 : 1 (b), and 51 : 1 (c).

(a) (b)



(c)

Fig. 57. Directional decomposition based coding results. The parameters used are: $\rho_c = 0.12$, subsampling rate = 5, zero crossing detection threshold = 50, magnitude and phase quantized to 256 and 16 levels, respectively. The compression ratios are 56 : 1 (a), 65 : 1 (b), and 92 : 1 (c).

account the properties of the receiver, i.e., the human visual system. These techniques are classified as first-generation techniques. They are characterized by the use of a

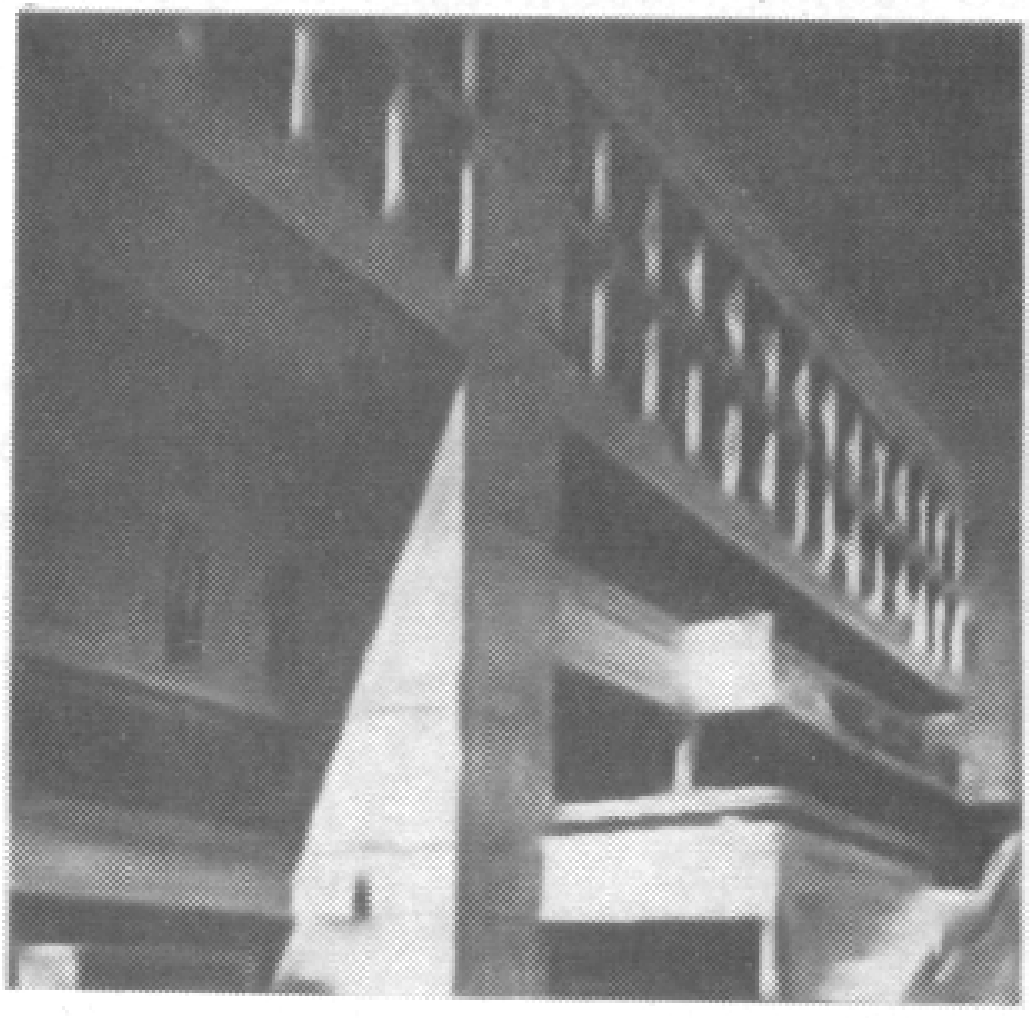


IMAGE CODING BASED ON A DIRECTIONAL DECOMPOSITION

number of directions (filters) : 8
cutoff frequency : 0.12
subsampling rate : 5
zero-cros. detection threshold : 25
total number of zero cros. :1107
phase mean coding length : 4
compression ratio :41.27

Fig. 40. An example of coded and decoded image with a compression ratio of 41 : 1.