COMPUTER GRAPHICS AND IMAGE PROCESSING **16**, 210–239 (1981)

Entropic Thresholding, A New Approach

T. PUN

Laboratoire de Traitement des Signaux, Swiss Federal Institute of Technology, 16, ch. de Bellerive, CH-1007 Lausanne, Switzerland

Received August 29, 1980

This paper describes an automatic threshold selection method for picture segmentation. The basic concept is the definition of an anisotropy coefficient, which is related to the asymmetry of the gray-level histogram. Its use permits the derivation of the entropic threshold, which has been successfully applied to images having various kinds of histograms. Several experimental results are presented. An extension to multithresholding is also suggested.

1. ENTROPIC THRESHOLDING, THE PREVIOUS APPROACH

1.1 Nature of the Problem

In the ideal case of a black object superimposed on a white background, the *gray-level density histogram* presents two peaks that are easily distinguishable. Thresholding that gray-level image into a black and white one is obvious: the threshold is selected in the valley of the histogram.

In the general case, however, for three major reasons, this operation is not so straightforward. First, nobody can say how many black picture elements (*pels*) are necessary and sufficient for a good interpretation. Second, the histogram gives only first-order statistical information, disregarding the semantic content of the image. Third, which criterion has to be applied when the histogram is not clearly bimodal?

The answer to the first question has not yet been found. Given more black points, more details will appear but also more noise. The limit has to be chosen depending on the desired use of the bilevel picture. The second problem has the consequence that varying the threshold will only lead to accepting more or fewer black pels. We meet then the previous problem. The use of higher-order statistics has been attempted [1–7], but this area seems to be still open. The third question is the one we are trying to solve here. We will see how another approach has been previously derived from considerations about the entropy of the histogram [8], and what kind of problems its use presents. We will then present a new approach to this concept of "entropic thresholding."

1.2 Entropic Thresholding, The Old Method

Let us consider the first-order probability histogram of a picture. Assuming that all symbols are statistically independent, its entropy (in the Shannon sense) is

$$H = -\sum_{i=0}^{n} p[i]\cdot \mathrm{lb}(p[i]) \qquad \text{shannon/symbol,} \tag{1}$$

where $n + 1$ is the number of gray levels, $p[i]$ the probability of occurrence of level i, and lb the log in base 2.

0146-664X/81/070210-30$02.00/0

After thresholding, the picture has two levels: white (w) and black (b). Its entropy becomes

$$H' = -p'[w]\cdot \mathrm{lb}(p'[w]) - p'[b]\cdot \mathrm{lb}(p'[b]) = H'[w] + H'[b], \tag{2}$$

where the ′ denotes the two-level picture.

Let us denote by s the value of the threshold. We can define two partial entropies

$$H[w] = -\sum_{0}^{s} p[i]\cdot \mathrm{lb}(p[i]) \qquad H[b] = -\sum_{s+1}^{n} p[i]\cdot \mathrm{lb}(p[i]). \tag{3}$$

$H[w]$ and $H[b]$ are an objective measure of the *a priori* quantity of information associated with white and black points, while $H'[w]$ and $H'[b]$ measure the *a posteriori* information.

In [8] we present a method for maximizing the *a posteriori* entropy H' using its intrinsic relationships with H. Achieving this, the aim was to have an adaptive unsupervised global criterion for selecting a threshold disregarding the geometrical shape of the histogram. The purpose of the use of H in the maximization process of H' is to avoid the trivial optimal (in the Shannon sense) case $H' = 1$. We would then always obtain $p'[w] = p'[b]$, and the threshold would not be adaptive.

1.3 Remarks

The experimental results show that this method leads to a thresholded picture with a number of black pels close to the number of white pels. That is to say that, despite the proposed method, the *a posteriori* entropy is close to its absolute maximal value. And the more the picture decreases in size, the more strong this effect becomes [8].

Thus, instead of using the entropy itself for deriving the value of the threshold, we will use it as a tool for the classification of the histogram. Taking this class into account, we will then select a threshold.

2. CLASSIFICATION OF HISTOGRAMS

2.1. Definition of the Anisotropy Coefficient

Let us consider a probability histogram with $n + 1$ possible gray levels. The probability of occurrence of each of these levels is denoted by $p[i]$. We obviously have

$$p[0] + p[1] + \cdots + p[n-1] + p[n] = 1. \tag{4}$$

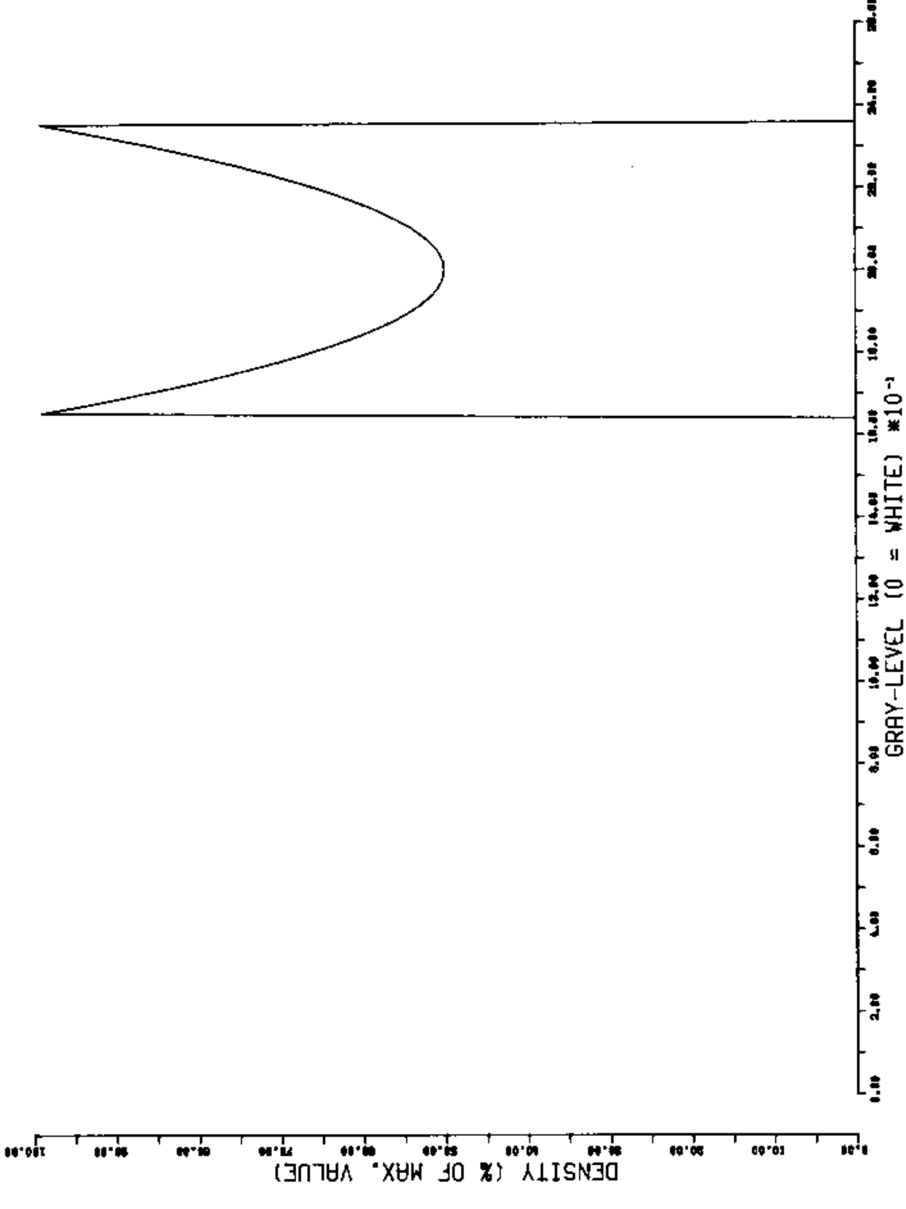
Let us denote by na the first level which satisfies

$$p[0] + p[1] + \cdots + p[na] \geq \tfrac{1}{2}. \tag{5}$$

If equality holds, na divides the histogram into two parts containing the same number of points. Otherwise, na separates it into two parts that are as equal as possible.

The *anisotropy coefficient* α is defined by

$$\alpha = \left\{\sum_{0}^{na} p[i]\cdot \mathrm{lb}(p[i])\right\} \Big/ \left\{\sum_{0}^{n} p[i]\cdot \mathrm{lb}(p[i])\right\}. \tag{6}$$


GRAY-LEVEL (0 = WHITE) *10⁻¹
DENSITY (% OF MAX. VALUE)

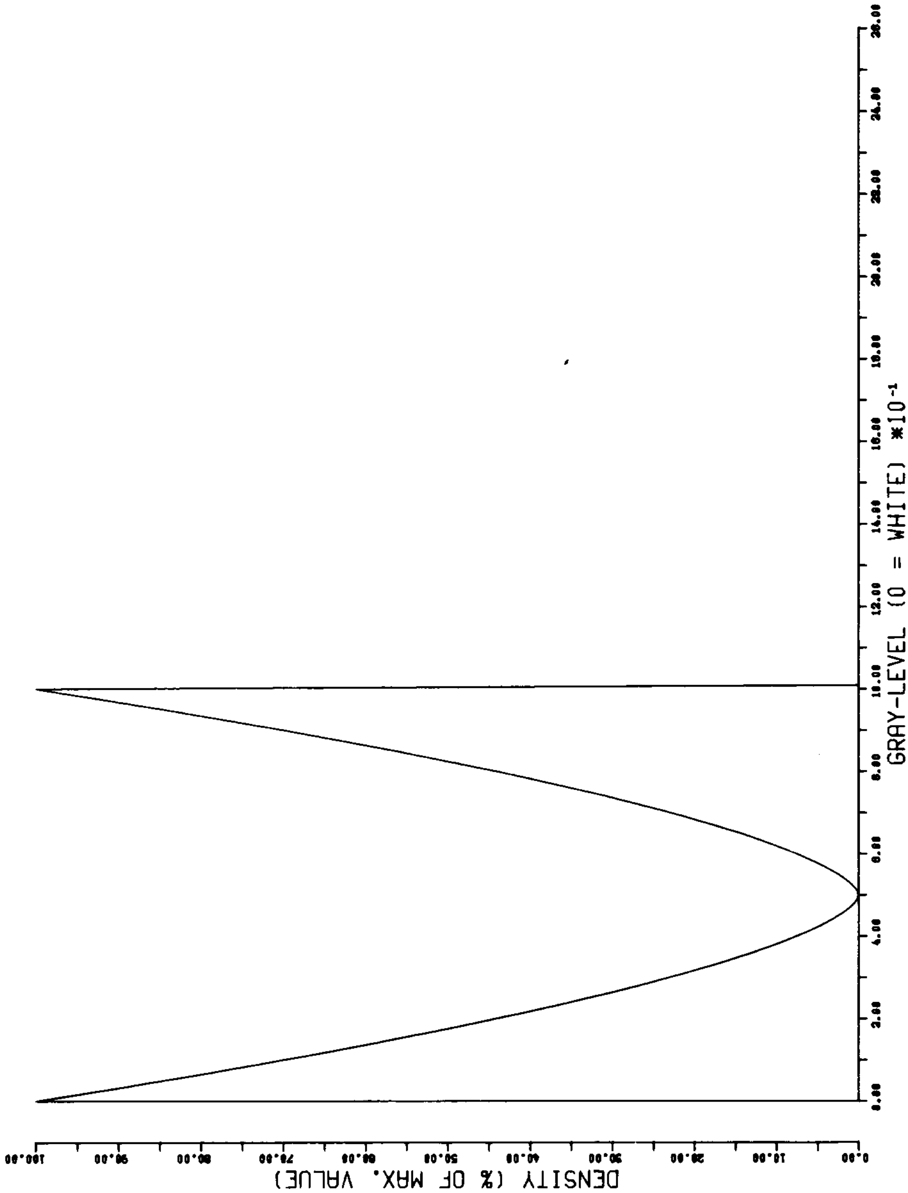


FIG. 1. Two histograms with central symmetry.

If na were the threshold, we would have

$$p'[w] = p'[b], \qquad H' = 1, \qquad \text{and} \qquad \alpha = H[w]/H. \tag{7}$$

In the following sections, we show how α can describe the geometric shape of the histogram.

2.2 Histograms with Central Symmetry

Let us consider an *equalized histogram*, that is to say, a histogram satisfying $p[i] = 1/(n + 1)$ for every level i. If $n + 1$ is even, it is obvious that $\alpha = \frac{1}{2}$. If $n + 1$ is odd, α is not exactly $\frac{1}{2}$, but all the following considerations will remain the same. In the remaining part of the paper, we will assume $n + 1$ even. For numerical applications, $n + 1$ will be 256 (8-bit pels).

$\alpha = \frac{1}{2}$ remains true for every histogram having a symmetric density function on each side of a central value (see Fig. 1). This is due to the fact that na will be the level closest to the central value. The proof is trivial and will be omitted.

This shows the *basic property* of the anisotropic coefficient: it is equal to $\frac{1}{2}$ for all histograms with central symmetry. Here it is also shift invariant; moving the histogram along the gray-level axis keeps α the same.

2.3. The "Left–Right" Effect

We will now analyze an ideal case, a two-peak histogram (see Fig. 2a) with probabilities of occurence p and $1 - p$. Let us assume that na will always be selected between the two peaks. This is not true if p is lower than 0.5, but this assumption is only made to demonstrate the '*left–right" effect*. We easily obtain from (5) and (6)

$$\alpha(p) = \{p \cdot \mathrm{lb}(p)\} / \{p \cdot \mathrm{lb}(p) + (1 - p) \cdot \mathrm{lb}(1 - p)\}. \tag{8}$$

(Without the above assumption, α would be 1 for p in [0.,0.5], and would be given by (8) for p in]0.5, 1.]).

The function $\alpha(p)$ is plotted in Fig. 2b. One can see that, while balancing the weight p from the left peak to the right peak, α varies symmetrically from 0 to 1. This is the "left–right" effect. For a *left-histogram* (with more weight on the left), α is between 0 and 0.5; for a *right histogram*, it is between 0.5 and 1.

Let us analyze what happens when we have two symmetric histograms, whose probability laws are denoted by p[i] and $p''[i]$. We have

$$p''[i] = p[n - i] \tag{9}$$

and na and na'' are given by

$$p[0] + p[1] + \cdots + p[na] = \tfrac{1}{2} \qquad p''[0] + p[1] + \cdots + p''[na''] = \tfrac{1}{2}. \tag{10}$$

From (9)

$$p''[0] + \cdots + p''[na''] = p[n - na''] + p[n - na'' + 1] + \cdots + p[n] = \tfrac{1}{2}. \tag{11}$$

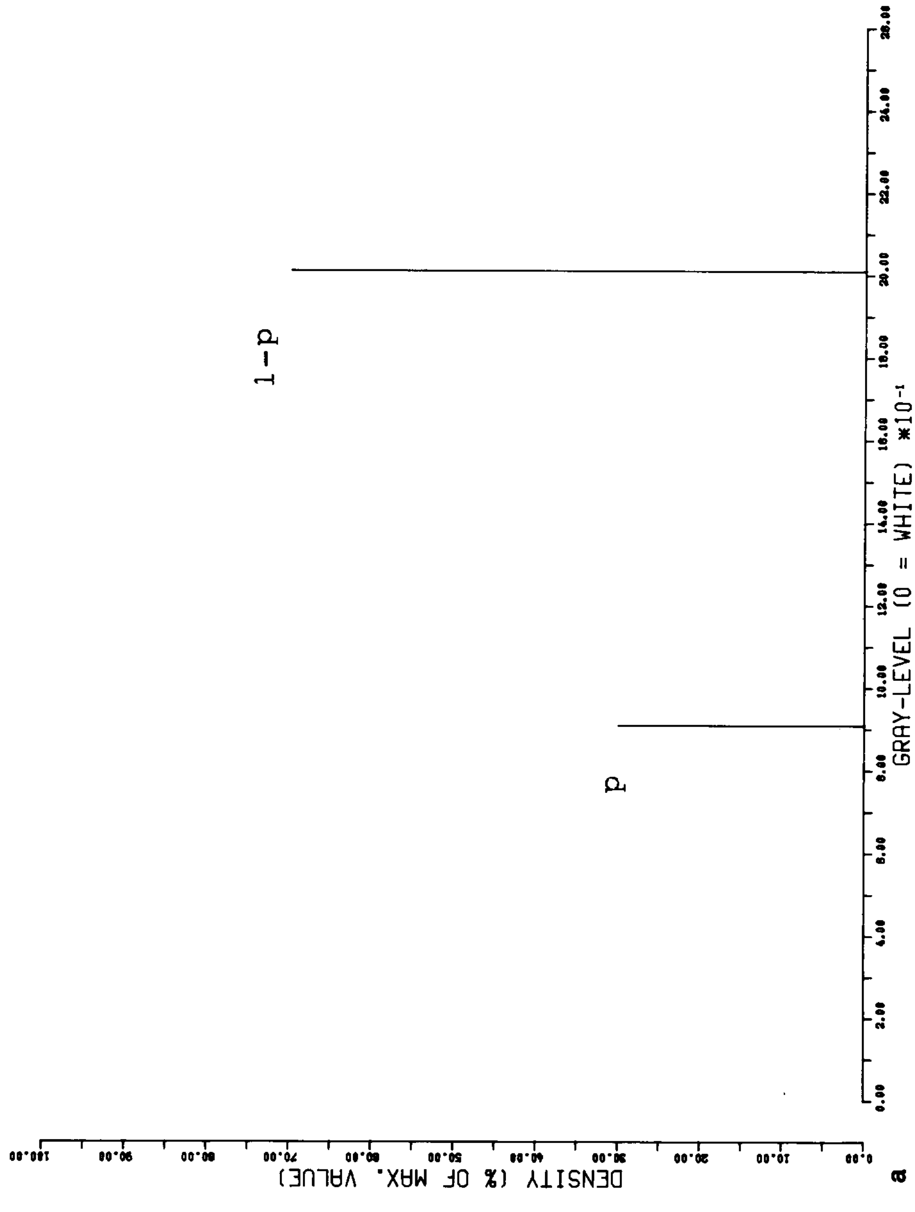


FIG. 2. a. Two-peak histogram (probabilities p and $1 - p$), b. anisotropy coefficient as a function of p.

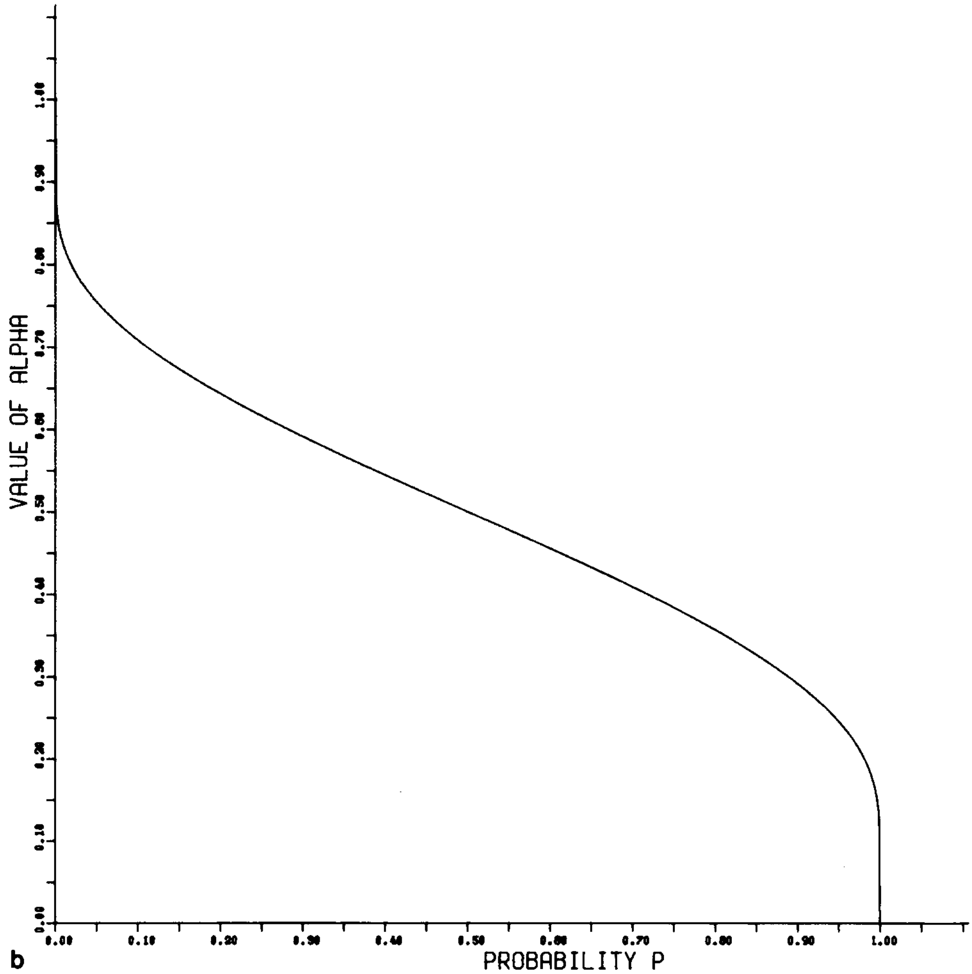


FIG. 2—*Continued.*

With (5), we have

$$p[na+1]+\cdots+p[n]=\tfrac{1}{2}. \tag{12}$$

Identifying (11) and (12), we obtain

$$na+na''=n-1. \tag{13}$$

From (6), and due to the symmetry (9),

$$\alpha''=\frac{\sum_{0}^{na''}p''[i]\cdot \mathrm{lb}(p''[i])}{\sum_{0}^{n}p[i]\cdot \mathrm{lb}(p[i])}. \tag{14}$$

Using (9) and (13), and with $j = n - i$, the numerator of (14) becomes

$$\sum_{0}^{na''} p''[i] \cdot \mathrm{lb}(p''[i]) = \sum_{na+1}^{n} p[i] \cdot \mathrm{lb}(p[i]). \tag{15}$$

Thus

$$\alpha + \alpha'' = 1. \tag{16}$$

This shows that a symmetry in the shapes leads to a symmetry in the anisotropy coefficients. An extreme case is given by two equalized histograms: $\alpha = \alpha'' = \frac{1}{2}$. At the other extreme, a very quick decaying histogram, with α close to 0, has a symmetric $p''[i]$ characterized by α close to 1. This will be used in the derivation of the method for selecting the threshold.

3. THE NEW ENTROPIC THRESHOLDING

3.1. Selection of the Threshold

We have seen in Section 1.1 that it is difficult to say whether a threshold is good or not. The only guidelines we have are intuitive. In the case of an equalized histogram, or for one with central symmetry, a "good" choice seems to put the threshold in the middle of the dynamic range. This is what we would do for a bimodal histogram composed of two equal peaks. We would then obtain an equal number of white and black pels.

For an histogram containing only one nonzero value, intuition would tell us to accept for black points either all pels or no pel at all.

Between those two extreme cases, what would be the choice for a quick decaying histogram? These kind of histograms arise from high-pass filtered images, where the relevant information is contained in the contours. Thus, while there are fewer contours than other points in a real world scene, we would like to have a number of black points smaller than half, and related to the slope of the histogram.

All these considerations lead to the following definition of the *entropic threshold*; it will be selected at the level s such that

$$\sum_{0}^{s} p[i] = \left\{\tfrac{1}{2} + \mathrm{abs}\left(\tfrac{1}{2} - \alpha\right)\right\} = \begin{cases} 1 - \alpha & \text{if} \quad \alpha \leq \frac{1}{2}, \\ \alpha & \text{if} \quad \alpha \geq \frac{1}{2} \end{cases}. \tag{17}$$

It is easy to demonstrate from (17) that the percentage of black points after the thresholding operation is

$$100 \cdot \alpha \text{ if } \alpha \leq \tfrac{1}{2} \qquad \text{and} \qquad 100 \cdot (1 - \alpha) \text{ if } \alpha \geq \tfrac{1}{2}. \tag{18}$$

Then, the thresholding operation will be performed in the following way:

$$\text{if } i \geq s, i \leftarrow 1 \qquad \text{and} \qquad \text{if } i < s, i \leftarrow 0. \tag{19}$$

The right term of (17) can be viewed as a kind of "*distance to the equalized histogram*." If α is $\frac{1}{2}$, the threshold s will be equal to na (middle of the dynamic

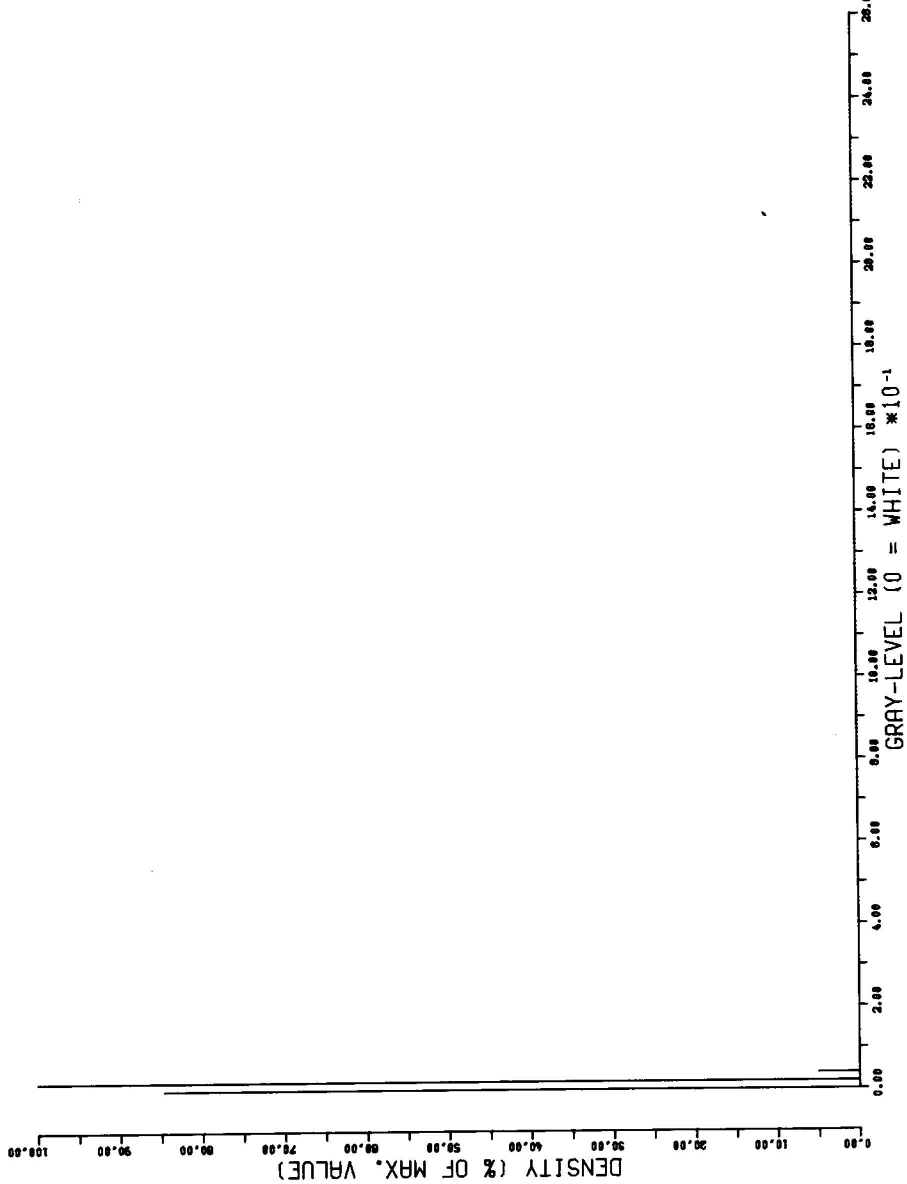
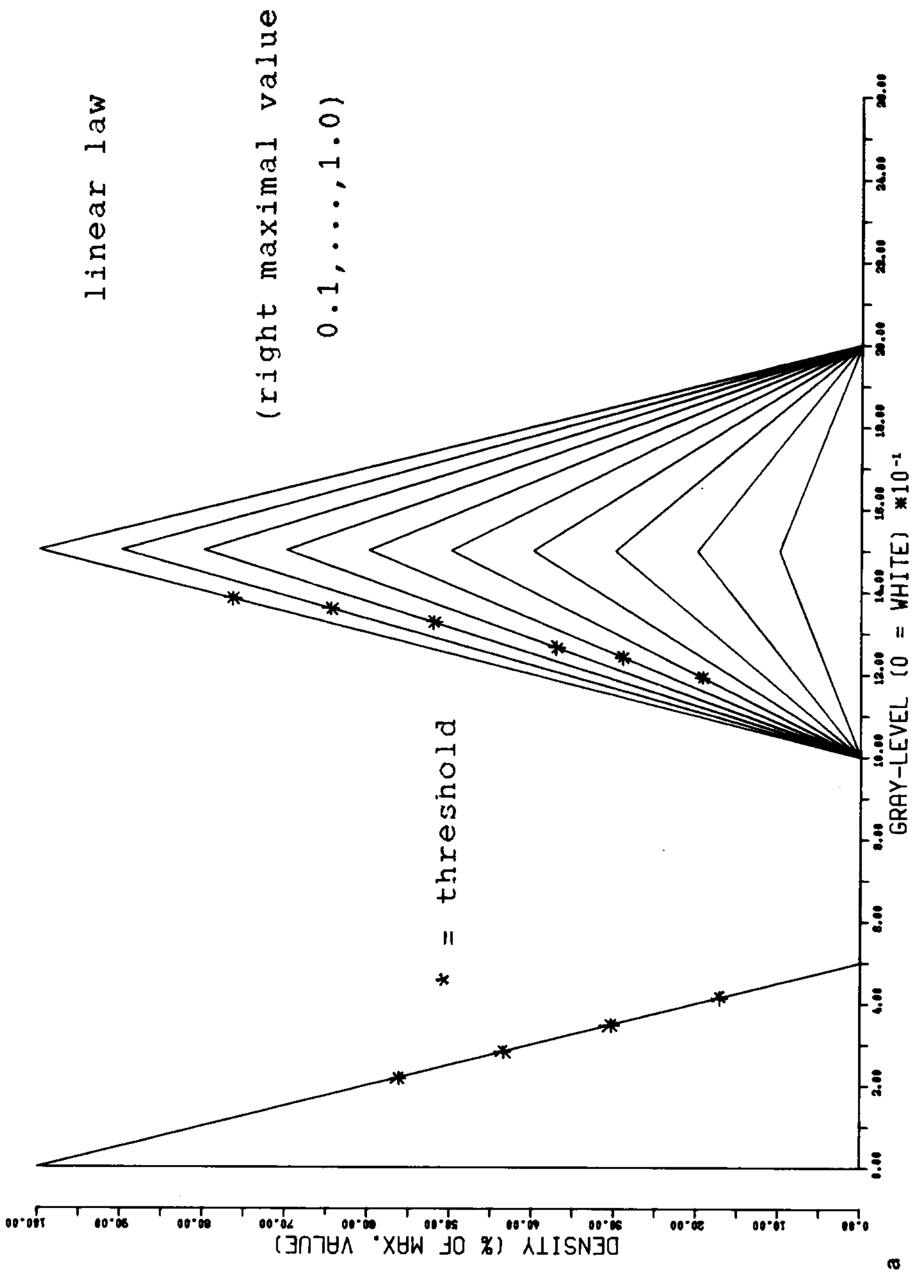


FIG. 3. Histogram of a small (32 × 32 pels) noisy subimage, extracted from Fig. 7a.



a

FIG. 4. Two kinds of histograms: a. bimodal linear law, b. quick decaying law.

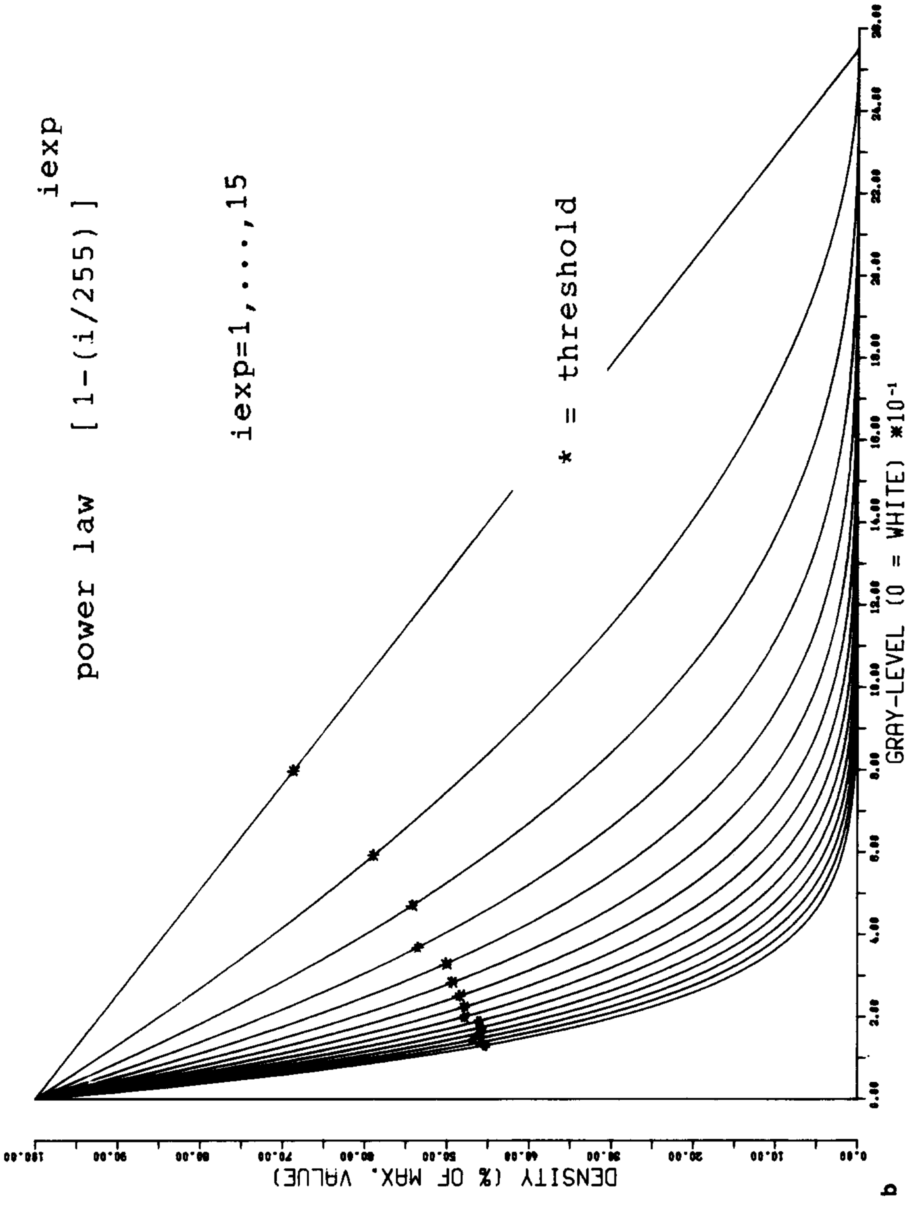


FIG. 4—*Continued.*

range). For a quick decaying histogram, with α close to zero, we will have a small number of black points.

The *absolute value* used in (17) arises for two reasons. First, suppose we have two symmetric quick decaying histograms, whose anisotropy coefficients are related by (16). Since in both cases we would like to obtain the same number of black points, the absolute value is needed. Second, let us consider the extreme case of a very small image (or subimage) containing only low-level background noise. Due to the very low number of levels, it can occur that α becomes greater than 0.5 although the histogram is a "left one" (see Fig. 3). To avoid noise, thus to have a high threshold, the absolute value will also be necessary.

3.2. Examples

The method has been tested experimentally with two kinds of artificial histograms. First, *linear bimodal* ones, like those of certain real images. Second, *quick decaying* ones, like those of high-pass filtered images.

If we denote by $f(i)$ the laws presented in Fig. 4, the probability law of the histogram will be

$$p[i] = a \cdot f[i], \tag{20}$$

where a is a normalization factor,

$$a = 1/\{f[0] + \cdots + f[n]\}. \tag{21}$$

In order to compare equivalent laws, we have to compare the values of α/a instead of the values of α.

In the linear bimodal case (Fig. 4a), α/a increases linearly with the maximal value of the right mode, thus linearly with its area. This is due to the fact that the histogram becomes a "*right*" one. The chosen thresholds are indicated by stars (Fig. 4). It follows from (17) that they cannot be in the valley, since there are no nonzero levels in it.

In the quick decaying case (Fig. 4b), α/a decreases when the histogram becomes "*left*." The thresholds seem to be on a quite regular curve, situated between the zero level and the maximal curvature level.

4. EXPERIMENTAL RESULTS

4.1. The Four Pictures Used

Four pictures were used for the experiments, having various kinds of structure. They are shown in Fig. 5; Fig. 5a is the "building," Fig. 5b the "cameraman," Fig. 5c the "humans" and Fig. 5d the "crowd." They are all digitized with a raster of 256×256 points and quantized to 256 levels.

The entropic threshold was applied both to the original pictures (Section 4.2) and to preprocessed pictures (Section 4.3).

The *preprocessing* consists of two stages. First, a 4×4 *median filter* [9, 10] is applied; here, the median of the gray levels of the window will be the 9th value. This size of 4×4 was determined experimentally, in order to decrease noise while keeping a rather good definition. Second, a *high-pass filter* (2×2 Mero–Vassy

FIG. 5. The four original images used in this study: a. building, b. cameraman, c. humans, d. crowd.

FIG. 5—*Continued.*

operator) has been used for contour extraction. Figure 6 shows the results of the median filter, and Fig. 7 the results after all the preprocessing.

4.2. Thresholding the Original Pictures

For reasons of space, results for only two of the four pictures (building and crowd) are presented here. Figure 8 shows their gray-level histograms, with indication of the values of the anisotropy coefficient, the selected threshold and the percentage of black points. Figure 9 presents the thresholded images.

We observe that α is quite close to 0.5; this is an illustration of Section 2.3. We also verify that the percentage of black points is equal to $100 \cdot \alpha$ (18).

The evaluation of the interpretability of these pictures is not evident. This question can also be formulated as follows: when a gray-level picture is thresholded, which are the interesting points that have to become black? However, the pictures shown in Fig. 9 are quite satisfying.

4.3. Thresholding Preprocessed Pictures

We are now trying to extract the relevant *contour information* from the original pictures. Figure 10 presents the gray-level histograms of the four images shown in Fig. 7, with values of α, threshold and percentage of black points. With this kind of histogram, it is evident that no geometrical criterion (such as the "valley" one) could be used.

We verify that α is greater for Fig. 10d than for Fig. 10b. This is due to the "left–right" effect: the histogram shown in Fig. 10d decays slower than the other one (see also Fig. 4b).

Instead of thresholding the full picture with the same value, we can subdivide the image into blocks [8]. Then, more details will be extracted, but also more noise will appear. Figure 12 gives an example of this effect. Nevertheless, it is less strong than with the use of the old "entropic thresholding" method.

5. CONCLUDING REMARKS

5.1. Extension to Multithresholding

The proposed method can be extended to *multithresholding*. For example, if k *thresholds* have to be selected, we have to find k *levels* $na[1], na[2], \ldots, na[k]$ dividing the histogram into $k+1$ equal parts. We can then define k *anisotropy coefficients* $\alpha 1, \ldots, \alpha k$ as follows:

$$\alpha 1 = \frac{\sum_{0}^{na[1]} p[i] \cdot \mathrm{lb}(p[i])}{\sum_{0}^{n} p[i] \cdot \mathrm{lb}(p[i])} \tag{22}$$

...

$$\alpha k = \frac{\sum_{na[k-1]+1}^{na[k]} p(i) \cdot \mathrm{lb}(p[i])}{\sum_{0}^{n} p[i] \cdot \mathrm{lb}(p[i])}. \tag{23}$$

FIG. 6. Original images after 4 × 4 median filtering. a, b, c, d: Same as Fig. 5.

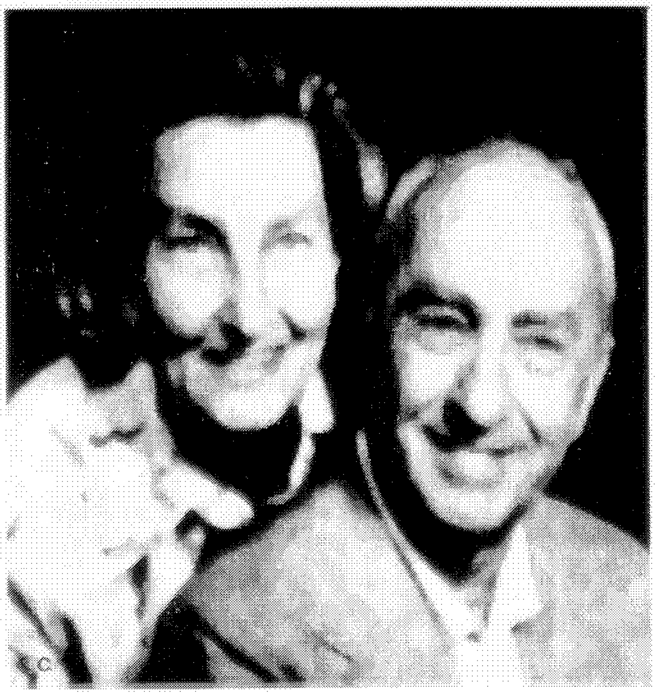
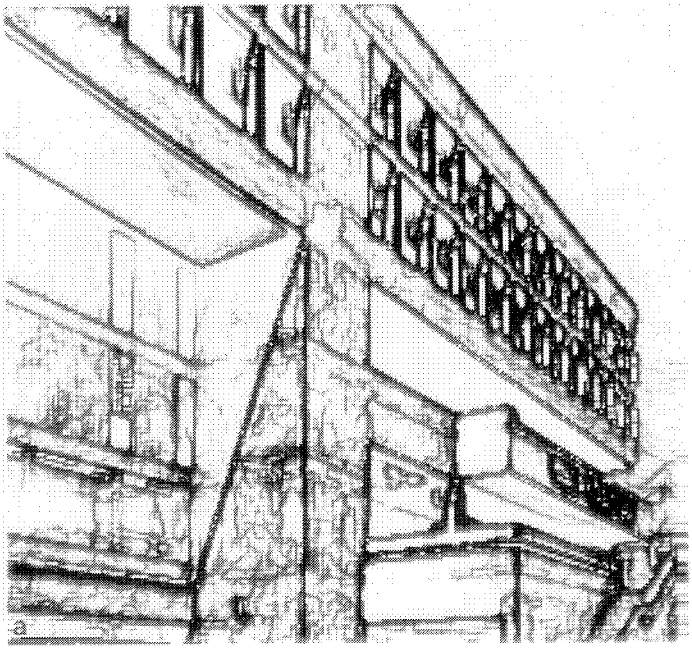


FIG. 6—*Continued.*



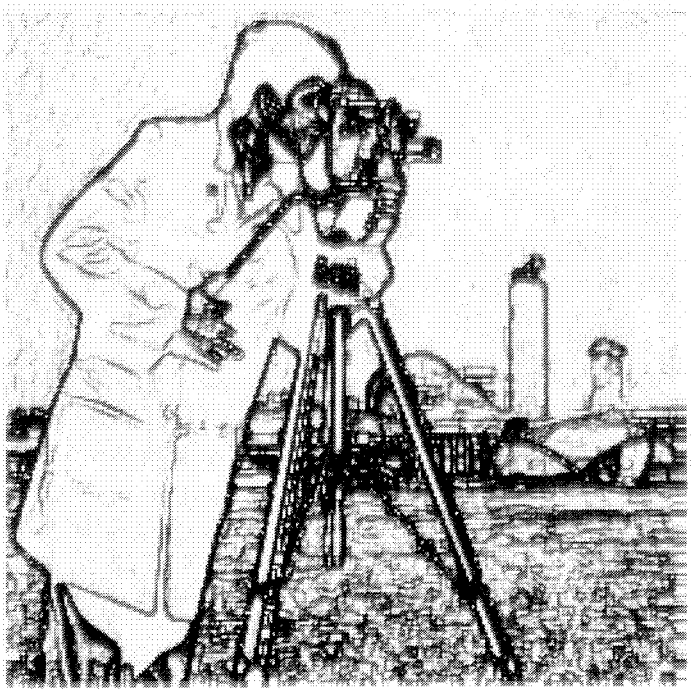


FIG. 7. Original images after all the preprocessing. a, b, c, d: Same as Fig. 5.

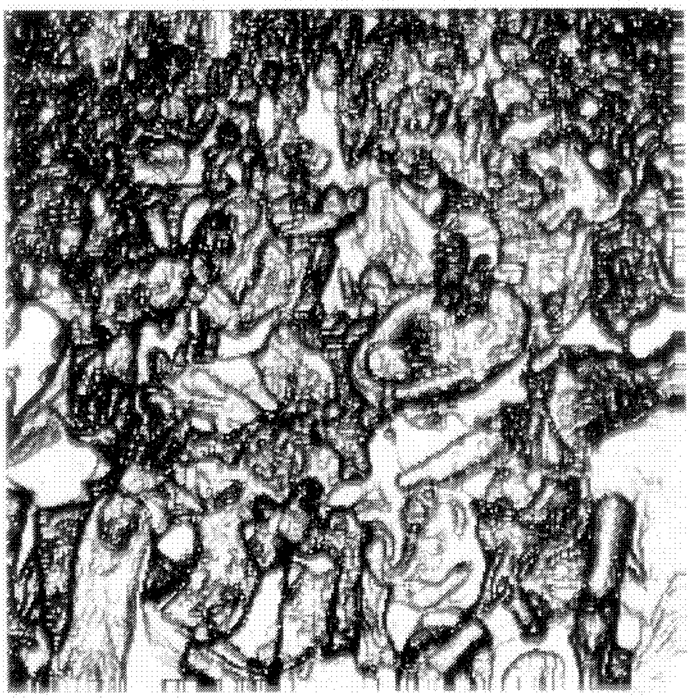
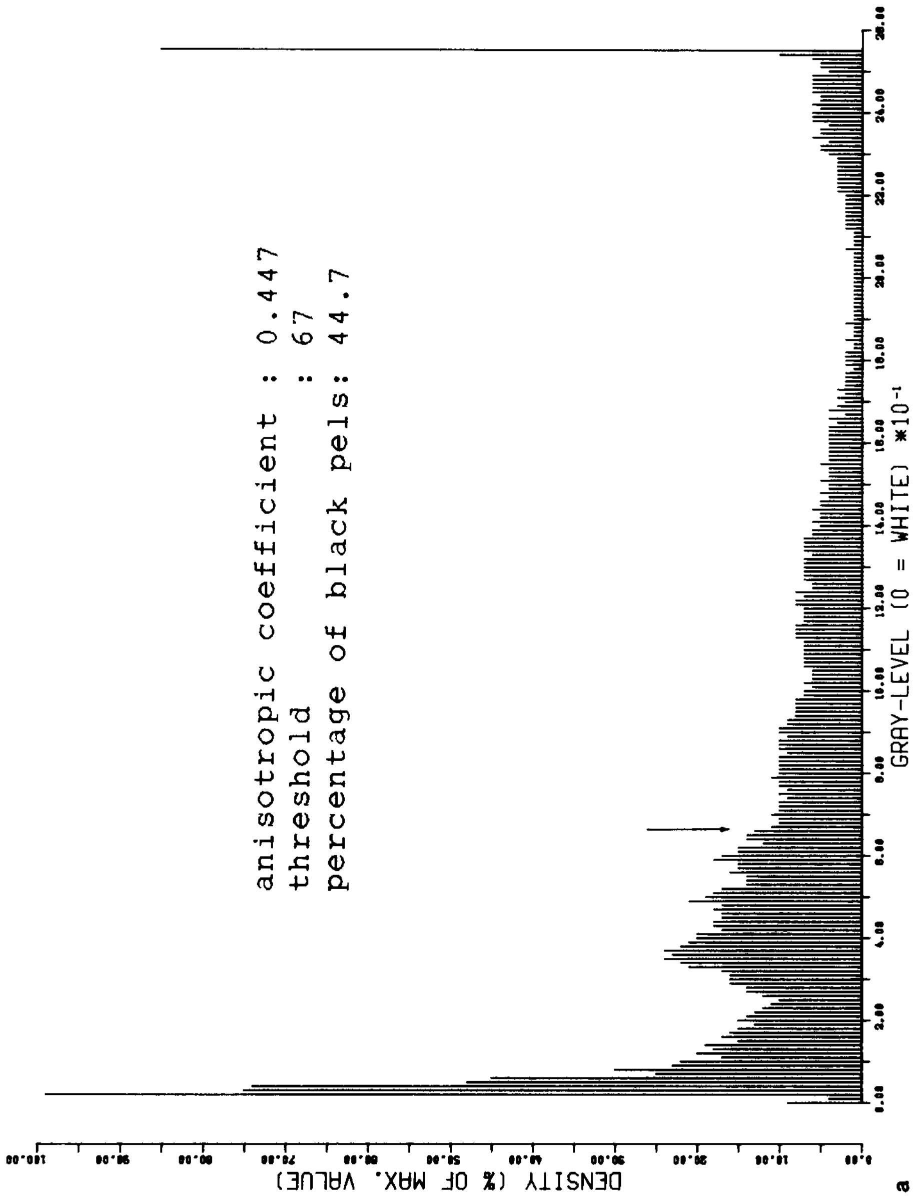
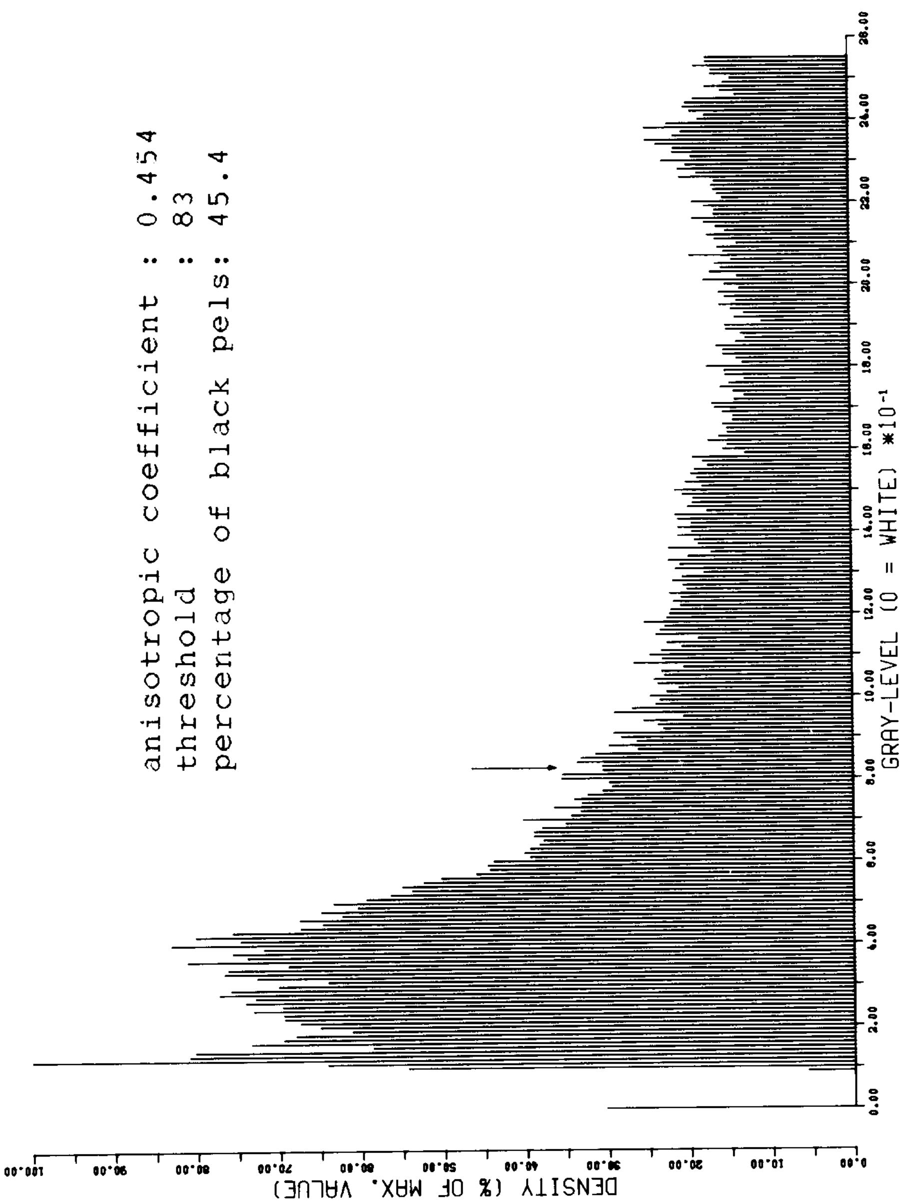


FIG. 7—*Continued.*



a

FIG. 8. Gray-level histograms: a. for Fig. 5a, b. for Fig. 5d.



b

FIG. 8—*Continued.*

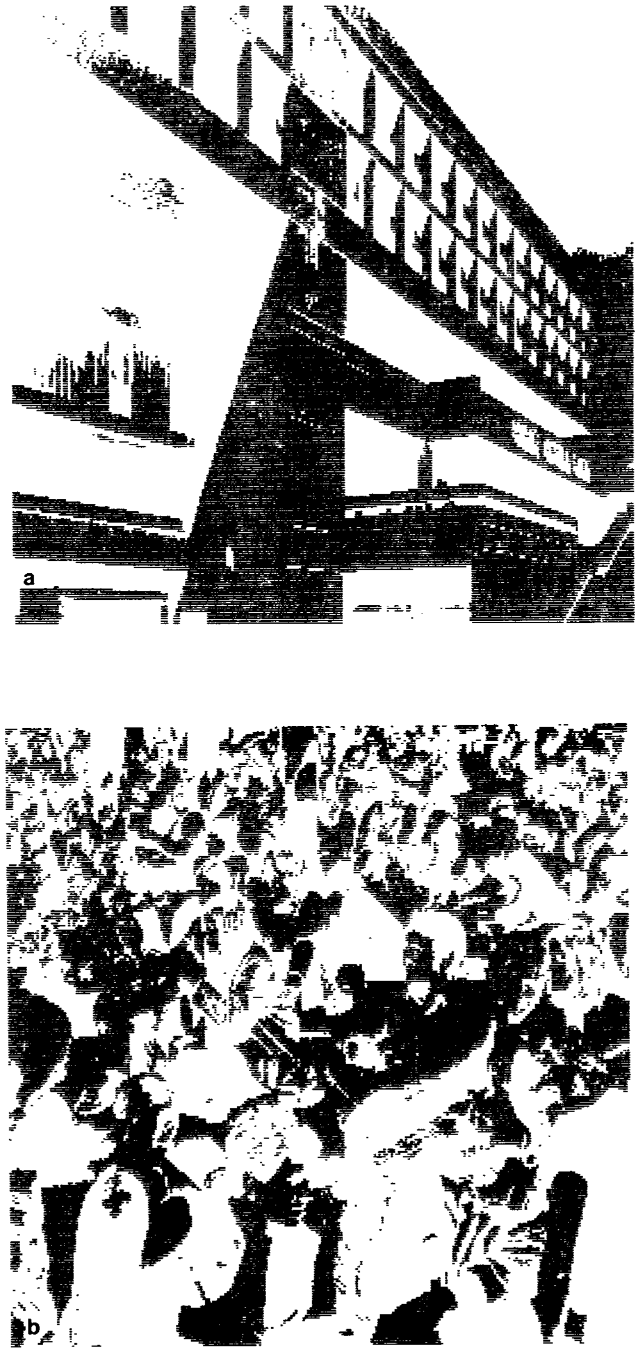
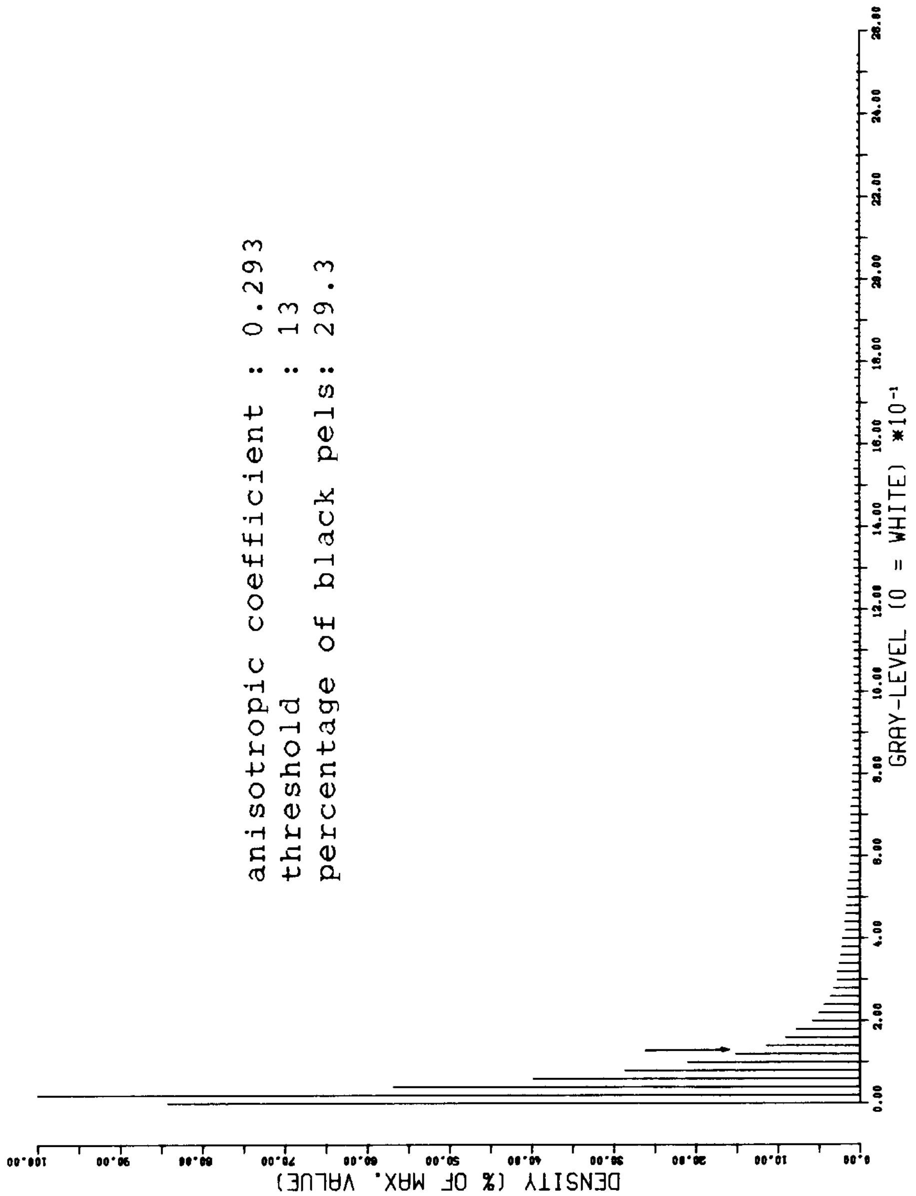


FIG. 9. Thresholded pictures (values of threshold in Fig. 8): a. for Fig. 5a, b. for Fig. 5d.


anisotropic coefficient : 0.293
threshold : 13
percentage of black pels: 29.3
GRAY-LEVEL (0 = WHITE) *10⁻¹
DENSITY (% OF MAX. VALUE)

a

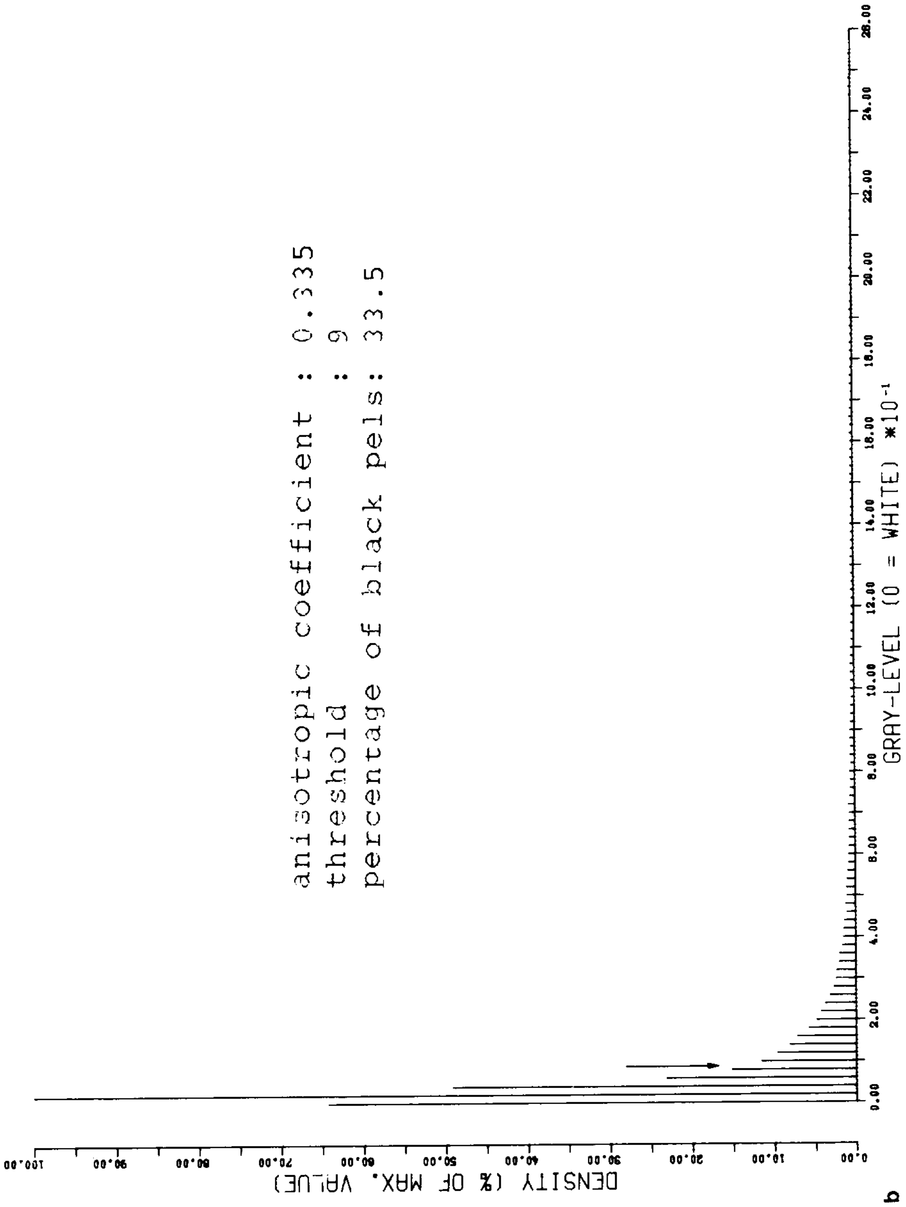
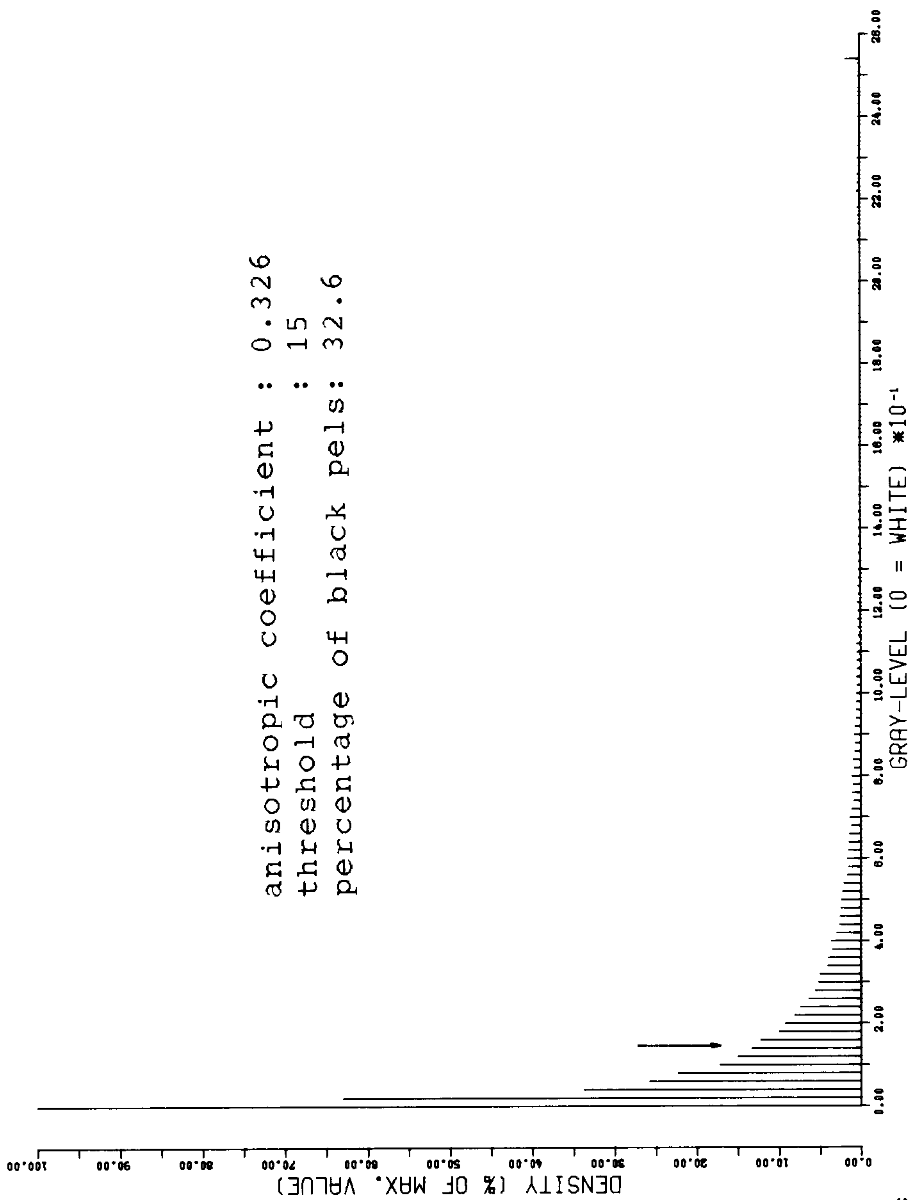
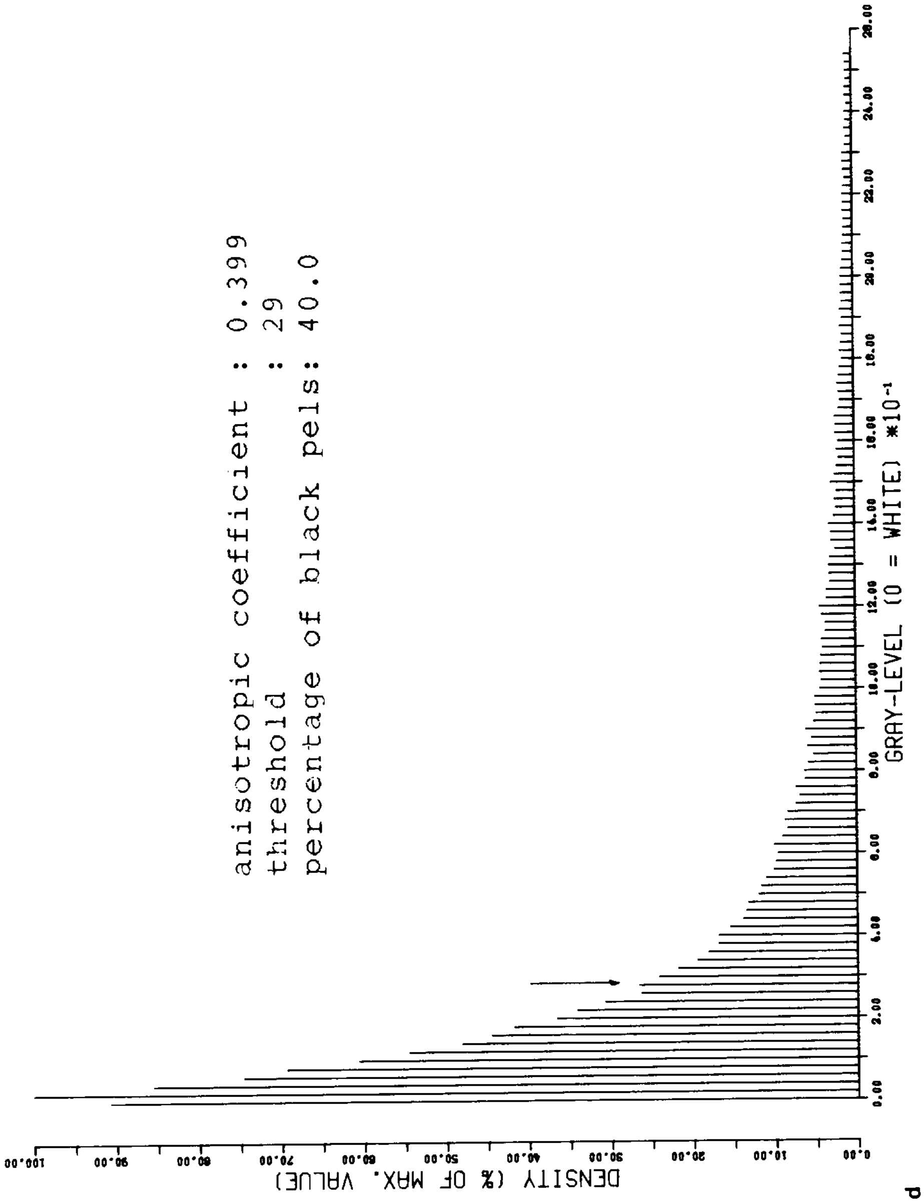


FIG. 10. Gray-level histograms of Figs. 7a, 7b, 7c, 7d. a, b, c, d: Same as Fig. 5.



c

FIG. 10—*Continued.*



d

FIG. 10—*Continued.*

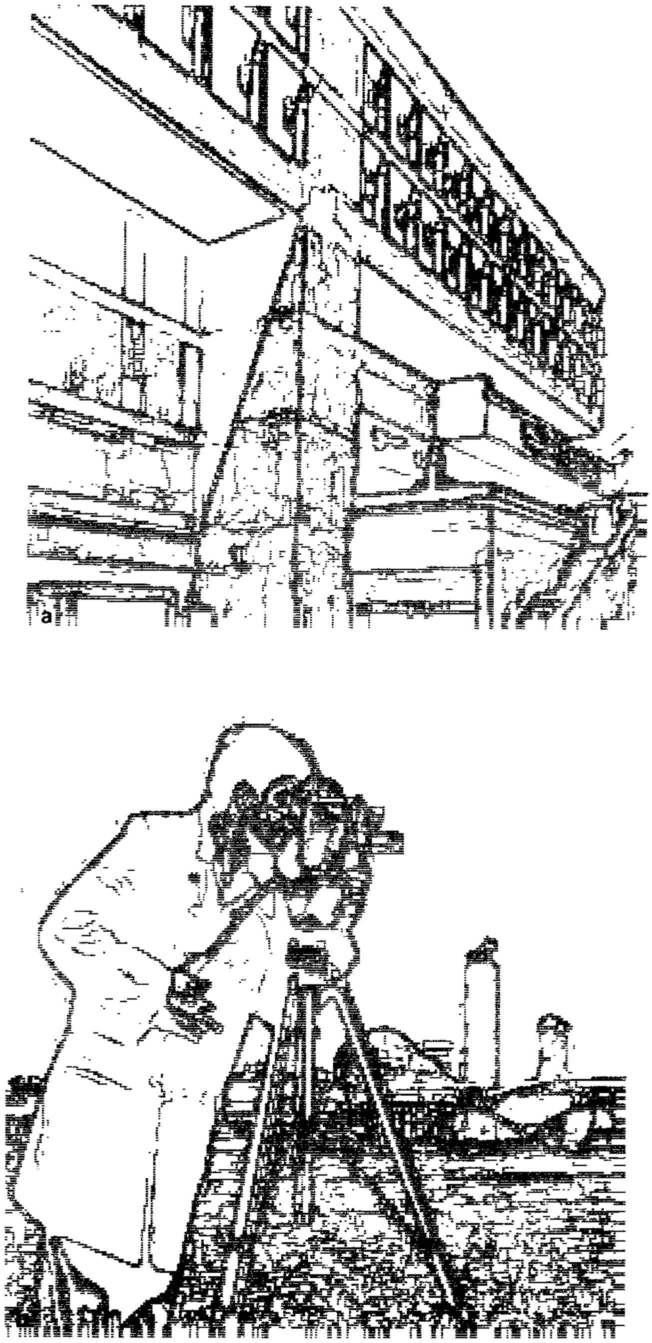


FIG. 11. Thresholded pictures obtained from Figs. 7a, 7b, 7c, 7d. a, b, c, d: Same as Fig. 5.

FIG. 11—*Continued.*

FIG. 12. Picture obtained by thresholding blocks of 64 × 64 pels (from Fig. 7a).

The problem, which is under investigation, is how to extend formula (17). A possible way of achieving that is to have k *thresholds* $s[1],\ldots,s[k]$ satisfying

$$\sum_{0}^{s[1]} p[i] = \left\{ \sum_{0}^{na[1]} p[i] + \mathrm{abs}\left(\sum_{0}^{na[1]} p[i] - \alpha 1 \right) \right\} \tag{24}$$

...

$$\sum_{0}^{s[k]} p[i] = \left\{ \sum_{0}^{na[k]} p[i] + \mathrm{abs}\left(\sum_{na[k-1]+1}^{na[k]} p[i] - \alpha k \right) \right\}. \tag{25}$$

This extension has been successfully tested experimentally for $k = 2$.

5.2. Extension to Multidimensional Thresholding

When a *multidimensional histogram* has to be thresholded, the problem becomes more difficult. It depends on whether one or more thresholds are to be selected, for example, one for each 1D histogram. It is also not obvious how to define a suitable *entropy measure*. This problem remains open.

5.3. *Open Questions*

Several questions are unsolved. First, what happens if two different pictures do have the same histogram, thus the same threshold? Will it be suitable for both? The answer is perhaps that comparable kinds of black and white image must have about the same number of black pels to be fairly well interpretable, disregarding their semantic content.

This consideration leads to the second question: has this anisotropy coefficient α any meaning other than a geometrical interpretation? How should this concept of first-order entropy be interpreted in the case of images?

Third, would it be useful to have higher-order entropies?

5.4. *Conclusion*

A method for automatic selection of a threshold using the gray-level histogram of a picture has been presented. An anisotropy coefficient is extracted from this histogram, which is closely related to its geometrical shape. It permits the unsupervised selection of one or more threshold values.

This method is numerically very easy to implement. The experimental results are fairly good, either on the original pictures or on high-pass filtered images.

ACKNOWLEDGMENTS

The author wishes to thank Professor T. S. Huang of Purdue University for the very deep and helpful discussions he had with him.

This work was supported by the FNS (Swiss National Science Foundation) under Grant 3.982-0.78 to the Laboratoire de Traitement de Signaux of the Swiss Federal Institute of Technology.

REFERENCES

1. D. P. Panda and A. Rosenfeld, Image segmentation by pixel classification in {gray level, edge value} space, *IEEE Trans. Computers* **C-27**, No. 9, 1978.
2. J. S. Weszka and A. Rosenfeld, Histogram modification for threshold selection, *IEEE Trans. Syst. Man Cybernet.* **SMC-9**, No. 1, 1979.
3. A. Wegmann, "Extraction de contours par histogramme bidimensionnel," Rapport interne LTS, Laboratoire de Traitement des Signaux, Swiss Federal Institute of Technology, June 1980.
4. D. L. Milgram and M. Herman, Clustering edge values for threshold selection, *Computer Graphics and Image Processing* **10**, 1979, 272–280.
5. Y. Nakagawa and A. Rosenfeld, Edge/border coincidence as an aid in edge extraction, *IEEE Trans. Syst Man Cybernet.* **SMC-8**, No. 12, 1978.
6. N. Ahuja and A. Rosenfeld, A note on the use of second order gray-level statistics for threshold selection, *IEEE Trans. Syst. Man Cybernet.* **SMC-8**, No. 12, 1978.
7. R. L. Kirby and A. Rosenfeld, A note on the use of {gray level, local average gray-level} space as an aid in threshold selection, *IEEE Trans. Syst. Man Cybernet.* **SMC-9**, No. 12, 1979.
8. T. Pun, A new method for gray-level picture thresholding using the entropy of the histogram, *Signal Process.* **2**, No. 3, 1980.
9. B. Justusson, "Noise Reduction by Median Filtering", Report TRITA-MAT-1978-7, Math. Inst., Royal Inst. of Technology, Stockholm, Sweden, Feb. 1978.
10. T. S. Huang, G. J. Yang, and G. Y. Tang, A fast two dimensional median filtering algorithm, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-27**, No. 1, 1979.