RZ 1210 (#43647) 2/24/83
Computer Science   24 pages

Research Report

SOME EXPERIMENTS WITH STOCHASTIC EDGE DETECTION

Rainer F. Hauser

IBM Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Typed by Denise Müller using the IBM MC 82

IBM Research Division
San Jose · Yorktown · Zurich

Copies may be requested from:

IBM Thomas J. Watson Research Center
Distribution Services 38-066
Post Office Box 218
Yorktown Heights, New York 10598

RZ 1210 (#43647) 2/24/83
Computer Science 24 pages

SOME EXPERIMENTS WITH STOCHASTIC EDGE DETECTION

Rainer F. Hauser

IBM Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Typed by Denise Müller using the IBM MC 82

ABSTRACT: The mathematical concept of information density is applied to the problem of edge detection for application in digital image processing. The stochastic properties of an image are modeled as Markov chains in x- and y-directions. Experiments are presented for different probability distributions, and the results obtained are compared with the well-known Sobel algorithm.

February 2, 1983

1. Introduction

At present, many research efforts in digital image processing deal with the general theme of image understanding [1], the objective of which is to clarify issues related to human and machine vision. In this context, two aspects are of primary importance: The representation and the interpretation of visual data.

Representation of visual data

In any technical system, where the primary purpose of the representation of visual data is intended for the human observer, the data is represented as an array of spatially and amplitude-quantized grey-levels (pixels). This representation is denominated as the deterministic model of an image. However, visual data can also be represented in stochastic terms. An image is then modeled as a random field.

Interpretation of visual data

Psychophysiological experiments show that the content inherent in any pictorial information presented to a human observer can be subdivided into contours and texture. It is also a proven fact that contours play a key role whenever cognition tasks are performed.

In problems related to image understanding, edge-detection and contour-finding algorithms are therefore of primary importance, and the purpose of this paper is to describe a new approach based on stochastic modeling and representation of image data.

To obtain a stochastic representation of an image, a random field has to be constructed from the deterministic image. Ideally, the random field should reflect all the 'knowledge' available in the visual data given (a priori or extracted by other processes). A possible approach is to start the procedure with a primitive random field and subsequently refine it over different processing steps like determining statistical data, contours and other stochastic parameters. The iterative process of the refinement of the random field is illustrated in Fig. 1. The 'knowledge' modeled in the random field is accumulated in different steps, while the 'knowledge' in the deterministic model is constant. If the resulting random field contains just the configurations equivalent with respect to the image-processing application intended, the deterministic model is no longer needed. But the current state of the theory of stochastic image modeling does not allow exclusive use of the random field. Consequently, the random field combined with the deterministic model is used as a representation of the pictorial information given. As practical examples of random fields, Markov fields (Markov chains extended to two dimensions) have been used [2].

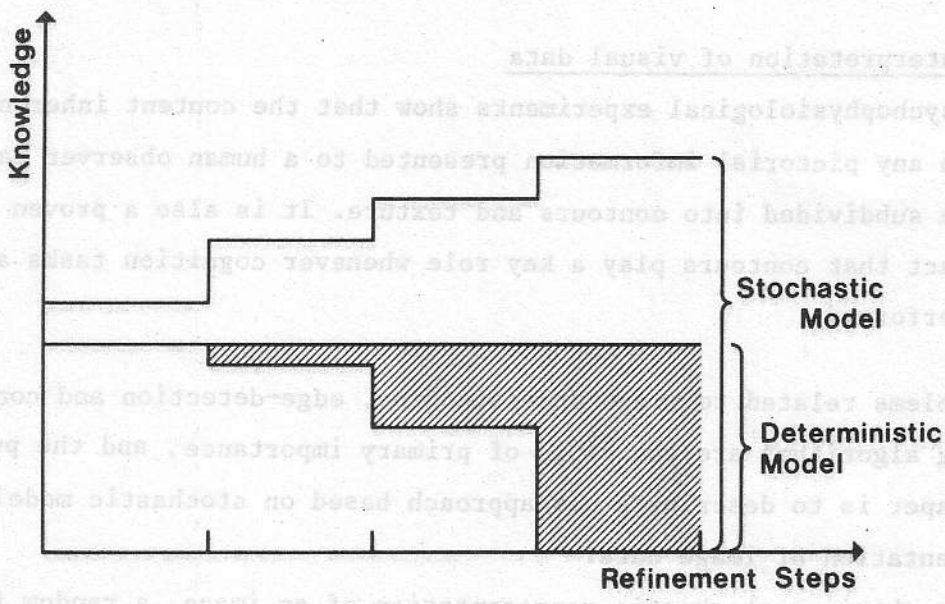


Figure 1. Illustration of the process of refinement for the stochastic model (over-lapping knowledge is hatched).

The purpose of this report is to demonstrate some of the difficulties in determining a good stochastic model for a given image and to show the feasibility of stochastic edge-detection algorithms. We present a version of a stochastic edge-detection algorithm with the following basic idea. By scanning unconsciously through an image, our attention is stimulated by sensitive changes of brightness, color, or texture. Therefore, a possible interpretation of the mechanism of human visual perception of color, shape etc., is to model this scanning process as a prediction from the known part of the image to the part next to be seen. A contour point is assumed when our prediction fails. In images, prediction can be stochastically modeled as random functions but we need a measure for the failure of our prediction in an image point. For this purpose, the concept of information density has been defined [3]. It is a value attached to each image point expressing the amount of information the point adds to the total information content of the image. When this value is higher than the average amount of information for one point, an edge point is assumed. In our case, the stochastic models used are very primitive, and an edge point can be assumed when the information-density value is higher than in the independent case.

In our experiments, we approximated the dependences with Markov chains in the x- and y-directions of the image. The Markov chains are truncated after two neighbors on both sides of the current point. Two different probability distributions on the image and some variants of them are studied and the results compared with the Sobel algorithm. These results are encouraging even for the rough approximations used in the stochastic models.

2. Information Density

In this section, the necessary mathematical definitions and results are presented, and a formula for stochastic edge detection is determined.

2.1. Random Functions

Stochastically modeled images are arrays of random variables

$$\{\xi_{ij} \mid (i,j) \in R_x \times R_y\}$$

together with a given sample

$$\{s_{ij} \mid (i,j) \in R_x \times R_y\}.$$

In other words, images in this section are finite random functions

$$\{\xi_t \mid t \in T\},$$

where T is a finite set of indices, and ξ_t is a random variable with values in a finite set G for each t in T. The given sample $\{s_t \mid t \in T\}$ restricts the random functions to cases with $P(\xi_t = s_t) \neq 0$ for all $t \in T$. This restriction is necessary because the direction goes here from sample to random function and not vice versa. The marginal probability of a subset S of T is given as

$$P(\bigwedge_{t \in S}(\xi_t = s_t))$$

for which we use the abbreviation $P(S)$. [In the same way, we write $P(t)$ for $P(\xi_t = s_t)$, and so on.] The conditional probability for $S_1, S_2 \subset T$ is defined by

$$P(S_1 | S_2) = \frac{P(S_1 \cup S_2)}{P(S_2)}.$$

2.2. Information Theory of Finite Random Functions

The information content of the event $\{\xi_t \mid t \in S\}$ with $S \subset T$ is $I(S) = -\log P(S)$, where the logarithm to the base 2 is usually taken.

With this definition, information terms are attached to the subsets of image points. In image processing, one is not interested in the marginal information content of a subset but in a measure for the information content of one image point t. This localized information content should reflect the dependences between different points. For this purpose, the concept of information density was defined [3] as

$$J_T(t) = \frac{1}{|T|} \sum_{S \subset T \setminus \{t\}} \frac{1}{\binom{|T| - 1}{|S|}} I(t \mid S), \tag{1}$$

and has the following three properties:

1) For all $t \in T$, the information density is not negative: $J_T(t) \geq 0$.

2) Summation over all t's results in the global information content: $\sum_{t \in T} J_T(t) = I(T)$.

3) If all points are stochastically independent [i.e., $I(S) = \sum_{t \in S} I(t)$ for all $S \subset T$], the information density $J_T(t)$ will be equal to the information content $I(t)$ of the point t.

For more details, see [3].

2.3. Markov Chains

Finite Markov chains are finite random functions $\{\xi_t \mid t \in T\}$ with a total ordering relation defined on the index set T. We set $|T| = N$ and $T = \{1, \ldots, N\}$ without loss of generality. The Markovian property is usually presented in the form [4]

$$P(\xi_n = g_n \mid \xi_{n-1} = g_{n-1}, \ldots, \xi_{n-i} = g_{n-i}) = P(\xi_n = g_n \mid \xi_{n-1} = g_{n-1})$$

which leads to the concept of transition matrices. When the transition matrix from one point to the next is constant, the Markov chain is called homogeneous. Two transition matrices can be multiplied to determine the transition probabilities from one point to the point after the next point, and it is clear that each subset S of T defines itself a Markov chain $\{\xi_t \mid t \in S\}$ with the induced total ordering relation.

Let us take one subset $S \quad T$ with $S = \{i_1, \ldots, i_n\}$ and $1 \leq i_1 < i_2 < \ldots < i_n \leq N$. With the Markovian property, we get

$$P\left(\bigwedge_{j=1}^{n} (\xi_{i_j} = g_{i_j})\right)$$

$$= P(\xi_{i_1} = g_{i_1}) \cdot P(\xi_{i_2} = g_{i_2} \mid \xi_{i_1} = g_{i_1}) \cdot \ldots \cdot P(\xi_{i_n} = g_{i_n} \mid \bigwedge_{j=1}^{n-1} \xi_{i_j} = g_{i_j})$$

$$= \frac{\prod_{j=1}^{n-1} P(\xi_{i_j} = g_{i_j} \wedge \xi_{i_{j+1}} = g_{i_{j+1}})}{\prod_{j=2}^{n-1} P(\xi_{i_j} = g_{i_j})}$$

which trivially results in

$$P\left(\xi_{i_m} = g_{i_m} \;\middle|\; \bigwedge_{\substack{i=1 \\ j \neq m}}^{n} (\xi_{i_j} = g_{i_j})\right) = P(\xi_{i_m} = g_{i_m} \mid \xi_{i_{m-1}} = g_{i_{m-1}} \wedge \xi_{i_{m+1}} = g_{i_{m+1}}). \tag{2}$$

This formula expressed in words says that each random variable is conditionally independent of the rest of the chain given its nearest neighbors.

2.4. Information Density of Markov Chains

We set $T_t^+ = \{t' \mid t' \in T \wedge t' > t\}$ and $T_t^- = \{t' \mid t' \in T \wedge t' < t\}$ and get with formula (2)

$$I(t \mid S) = \begin{cases} I(t) & \text{if } S \cap T_t^- = \phi,\ S \cap T_t^+ = \phi \\ I(t \mid \min\limits_{j \in S \cap T_t^+} j) & \text{if } S \cap T_t^- = \phi,\ S \cap T_t^+ \neq \phi \\ I(t \mid \max\limits_{j \in S \cap T_t^-} j) & \text{if } S \cap T_t^- \neq \phi,\ S \cap T_t^+ = \phi \\ I(t \mid \min\limits_{j \in S \cap T_t^+} j,\ \max\limits_{j \in S \cap T_t^-} j) & \text{if } S \cap T_t^- \neq \phi,\ S \cap T_t^+ \neq \phi\ . \end{cases}$$

The information density can be written in the form

$$J_T(t) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{\binom{N-1}{k}} E_k^N(t),$$

where $E_0^N(t) = I(t)$, and where for $k > 0$ the following three cases can occur:

1) $S \cap T_t^- \neq \phi, \quad S \cap T_t^+ = \phi$

2) $S \cap T_t^- = \phi, \quad S \cap T_t^+ \neq \phi$

3) $S \cap T_t^- \neq \phi, \quad S \cap T_t^+ \neq \phi.$

The first case gives

$$\sum_{j=1}^{t-1} \binom{j-1}{k-1} I(t \mid j),$$

the second

$$\sum_{i=t+1}^{N} \binom{N-i}{k-1} I(t \mid i),$$

and the third

$$\sum_{\ell=0}^{k-2} \sum_{j=1}^{t-1} \binom{j-1}{\ell} \sum_{i=t+1}^{N} \binom{N-i}{k-2-\ell} I(t \mid i,j),$$

where $\binom{n}{m} = 0$ for $m < 0$ or $n < m$. For $k > 0$, this results in

$$E_k^N(t) = \sum_{j=1}^{t-1}\binom{j-1}{k-1}(I(t,j) - I(j))$$

$$+ \sum_{i=t+1}^{N}\binom{N-i}{k-1}(I(t,i) - I(i))$$

$$+ \sum_{\ell=0}^{k-2}\sum_{j=1}^{t-1}\sum_{i=t+1}^{N}\binom{j-1}{\ell}\binom{N-i}{k-2-\ell}(I(t,i) + I(t,j) - I(t) - I(i,j)).$$

The information density is

$$J_T(t) = \frac{1}{N}\left[\sum_{m=1}^{N} a(N,t,m)\, I(m) + \sum_{n=2}^{N}\sum_{m=1}^{n-1} b(N,t,m,n)\, I(m,n)\right],$$

where

$$a(N,t,m) = \begin{cases} -\displaystyle\sum_{k=1}^{N-1}\frac{\binom{m-1}{k-1}}{\binom{N-1}{k}} & \text{if } m < t \\ -\displaystyle\sum_{k=1}^{N-1}\frac{\binom{N-m}{k-1}}{\binom{N-1}{k}} & \text{if } m > t \\ 1 - \displaystyle\sum_{k=1}^{N-1}\sum_{\ell=0}^{k-2}\sum_{j=1}^{t-1}\sum_{i=t+1}^{N}\frac{\binom{j-1}{\ell}\binom{N-i}{k-2-\ell}}{\binom{N-1}{k}} & \text{if } m = t, \end{cases} \qquad (3a)$$

and

$$
b(N,t,m,n) = \begin{cases} -\displaystyle\sum_{k=1}^{N-1}\sum_{\ell=0}^{k-2} \frac{\binom{m-1}{\ell}\binom{N-n}{k-2-\ell}}{\binom{N-1}{k}} & \text{if } m < t < n \\[2ex] \displaystyle\sum_{k=1}^{N-1}\left[\frac{\binom{m-1}{k-1}}{\binom{N-1}{k}} + \sum_{\ell=0}^{k-2}\sum_{i=t+1}^{N}\frac{\binom{m-1}{\ell}\binom{N-i}{k-2-\ell}}{\binom{N-1}{k}}\right] & \text{if } m < t = n \\[2ex] \displaystyle\sum_{k=1}^{N-1}\left[\frac{\binom{N-n}{k-1}}{\binom{N-1}{k}} + \sum_{\ell=0}^{k-2}\sum_{j=1}^{t-1}\frac{\binom{j-1}{\ell}\binom{N-n}{k-2-\ell}}{\binom{N-1}{k}}\right] & \text{if } m = t < n \\[2ex] 0 & \text{otherwise.} \end{cases} \tag{3b}
$$

Instead of 2^N information terms $I(S)$ with $S \subset T$, we only have N terms $I(i)$ and $N \cdot (N-1)/2$ terms $I(i,j)$ for Markov chains.

2.5 Approximations for Edge Detection

Markov chains are the easiest example of random functions. We shall now apply them to edge detection. We assume that

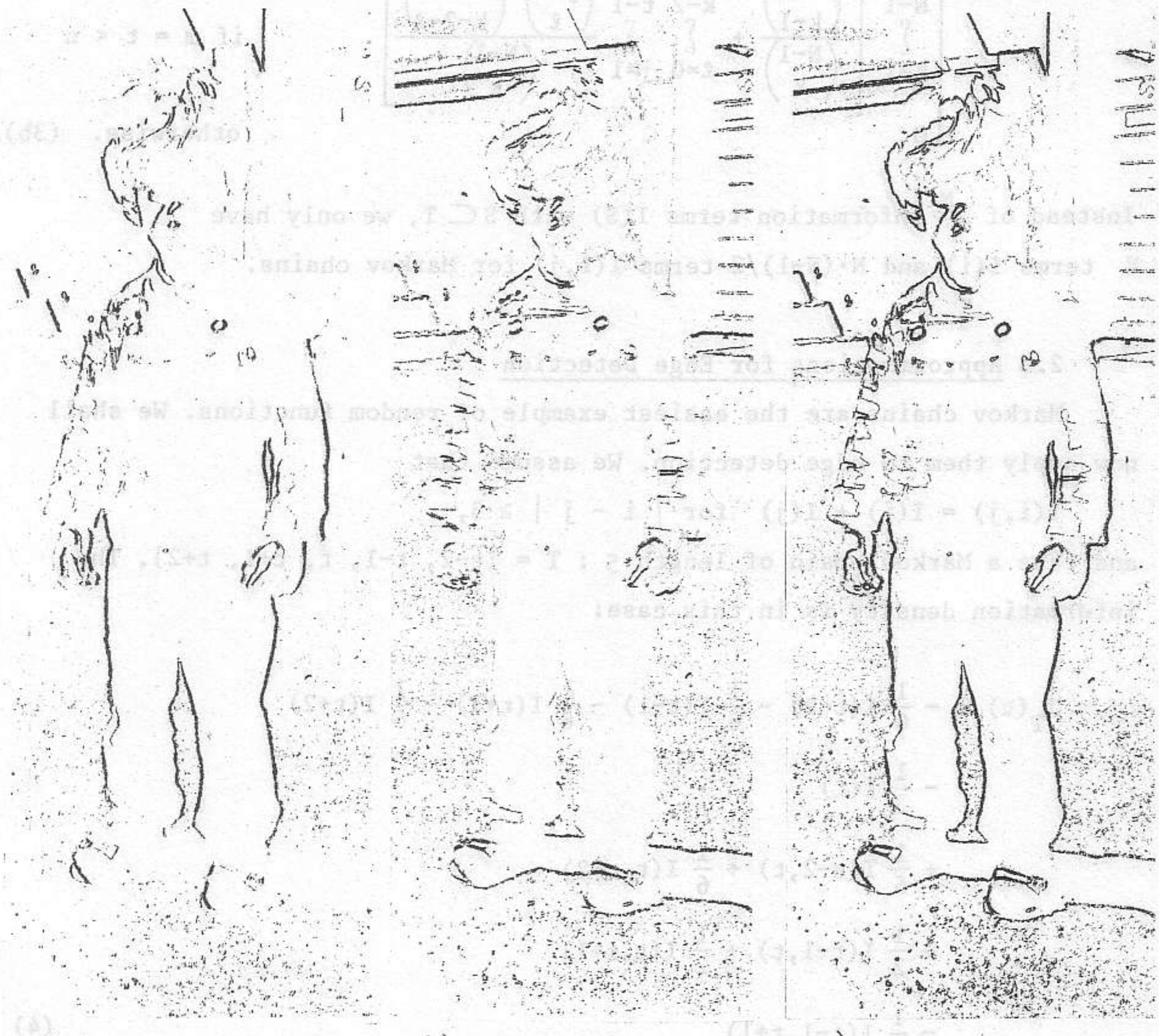
$$I(i,j) = I(i) + I(j) \quad \text{for } |\, i - j \,| \geq 3,$$

and take a Markov chain of length 5 : $T = \{t-2,\ t-1,\ t,\ t+1,\ t+2\}$. The information density is in this case:

$$
\begin{aligned}
J_T(t) = &- \frac{1}{6} I(t-2) - \frac{1}{6} I(t-1) - \frac{1}{6} I(t+1) - \frac{1}{6} I(t+2) \\
&- \frac{1}{3} I(t) \\
&+ \frac{1}{6} I(t-2,t) + \frac{1}{6} I(t,t+2) \\
&+ \frac{1}{2} I(t-1,t) + \frac{1}{2} I(t,t+1) \\
&- \frac{1}{3} I(t-1,t+1)
\end{aligned} \tag{4}
$$

as can be seen from formula (3).

With formula (4), we can develop an edge-detection algorithm. We assume that the current image point together with its two nearest neighbors on each side in the same column form a Markov chain of length 5. No dependences between the columns are modeled. This algorithm can only find contours orthogonal to the columns (see Fig. 2a). The same algorithm with the rows modeled as Markov chains detects edges only orthogonal to the rows (see Fig. 2b). Combined, the process running in both directions results in a bitmap with edges detected in all directions (see Fig. 2c).



a) b) c)

Figure 2. Markov-chain-based stochastic edge detection. a) x-direction only, b) y-direction only, c) both directions combined.

The image for which Fig. 2 presents the detected contours is shown as Fig. 3.

Figure 3. Original image.

To obtain the bitmap as shown in Fig. 2c, the following threshold value is used:

$$B(t) = \begin{cases} 0 & \text{if } J_T(t) \leq I(t) \\ 1 & \text{if } J_T(t) > I(t). \end{cases}$$

An edge point is assumed when the information density is higher than in the independent case.

We used here a rough approximation for the information density of a Markov chain because points with a distance of more than two points are approximated as independent. Also, the use of one-dimensional random functions on images is a poor simplification, since one would like to model dependences in the x- and y-directions simultaneously.

3. Stochastic Models and Edge Detection

We have deduced the formula for the information density of Markov chains which is used in this section. In this formula, the information terms I(i), I(i,i+1), and I(i,i+2) occur. We now discuss solutions to determine these values.

3.1. First and Second-Order Statistics

In an image of sufficient size, we can determine the probabilities $P(\xi_i = g)$, $P(\xi_i = g_1, \xi_{i+1} = g_2)$, and $P(\xi_i = g_1, \xi_{i+2} = g_2)$ from the first and second-order statistics. The general idea behind this stochastic model is the assumption that the probability $P(\xi_i = g)$ only depends on the value g and not on the value i, and that the probability $P(\xi_i = g_1, \xi_j = g_2)$ only depends on g_1, g_2 and $|i - j|$. In this sense, the Markov chain is assumed to be stationary. It must be remarked that this approach does not guarantee the Markovness of this probability distribution. The Markovness can be forced by determining $P(\xi_i = g_1, \xi_{i+2} = g_2)$ with matrix multiplication from the transition probabilities computable with $P(\xi_i = g)$ and $P(\xi_i = g_1, g_{i+1} = g_2)$. The algorithm with the stochastic model based on first and second-order statistics can detect contours (see Fig. 4).

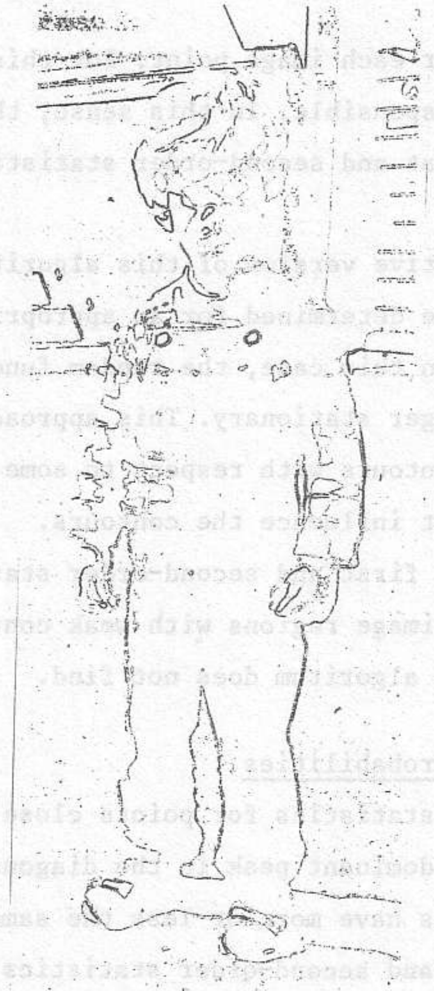


Figure 4. Result of the stochastic edge-detection algorithm with first and second-order statistics only.

We should mention an interesting gedanken experiment. If we take an arbitrary permutation of the finite set G of grey levels, in most cases the resulting image will be a more-or-less homogeneously grey image, where an observer cannot see what it represents. But the algorithm described here detects the same contours as in the original image. The second-order statistics have no peak in the diagonal, but the information density is the same for each image point. For this surprising effect, the quantization is responsible. In this sense, the stochastic edge-detection algorithm based on first and second-order statistics is an analysis tool for images.

To obtain an adaptive version of this algorithm, the first and second-order statistics can be determined for an appropriate neighborhood of the current image point. In this case, the random function approximated as a Markov chain is no longer stationary. This approach reflects the fact that human observers see contours with respect to some neighborhood, but image regions far away do not influence the contours.

The approach with first and second-order statistics has the advantage that also contours in image regions with weak contrast are found. Contours are detected the Sobel algorithm does not find.

3.2. Artificial Probabilities

The second-order statistics for points close together in natural images usually show a dominant peak in the diagonal, because the probability is high that the points have more or less the same grey level. Instead of determining the first and second-order statistics, we can construct artificially uniform first-order statistics and second-order statistics with the dominant peak in the diagonal (see Fig. 5). This approach gives results similar to those produced by the Sobel algorithm.

A variant of this approach is to determine the correct first-order statistics and to construct the second-order statistics only.

This approach has the disadvantage that the probability distributions do not reflect any characteristics of the image given. It is not in the sense of stochastic image processing to take one stochastic model for all images possible, where the main idea is to translate the knowledge about the image into a stochastic model.

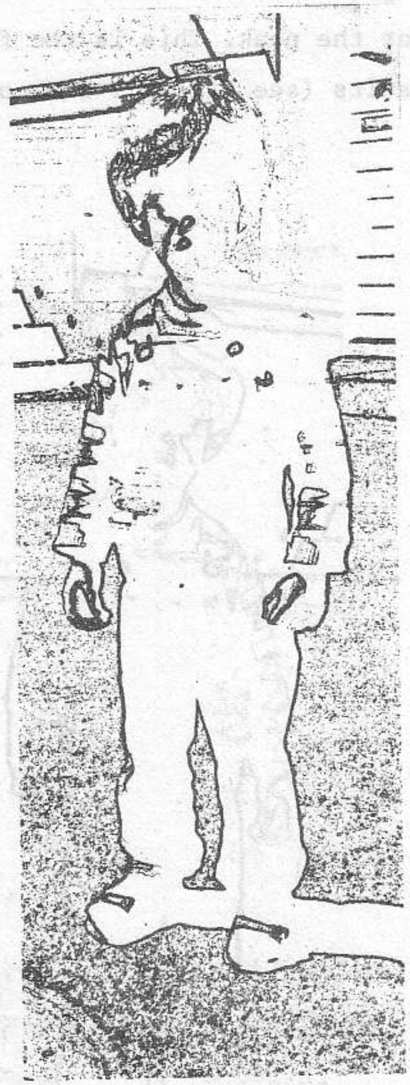


Figure 5. Result of the stochastic edge-detection algorithm with artificial statistics only.

3.3. Mixed Approach

We can improve the approach of the first and second-order statistics by the approach of the artificial probabilities without adding its disadvantage. We retouch the original second-order statistics by adding some points to the diagonal to accent the peak. This is the form of the algorithm in which we discuss its results (see Fig. 6). The contours should be connected and thin.

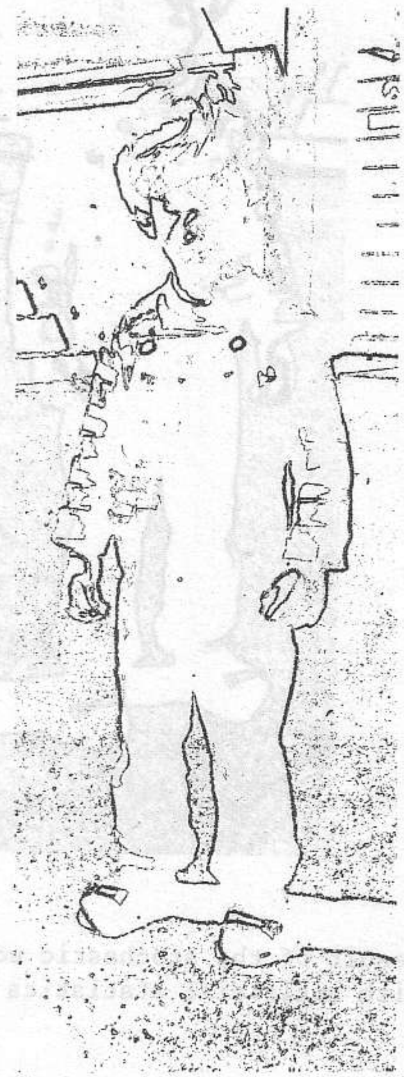


Figure 6. Result of the stochastic edge-detection algorithm with artificially enhanced second-order statistics.

3.4. The Sobel Algorithm

To compare the algorithm described, we use the Sobel algorithm which is easy to implement and produces excellent results [5]. The Sobel algorithm calculates the convolution of the image with the two matrices

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} , \quad \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

and combines the two results with a nonlinear point operation. We implemented the rms point nonlinearity. This algorithm has the disadvantage that the threshold must be optimized by human intervention. In Fig. 7, we tried to suppress all insignificant points and to show all the contours detected.

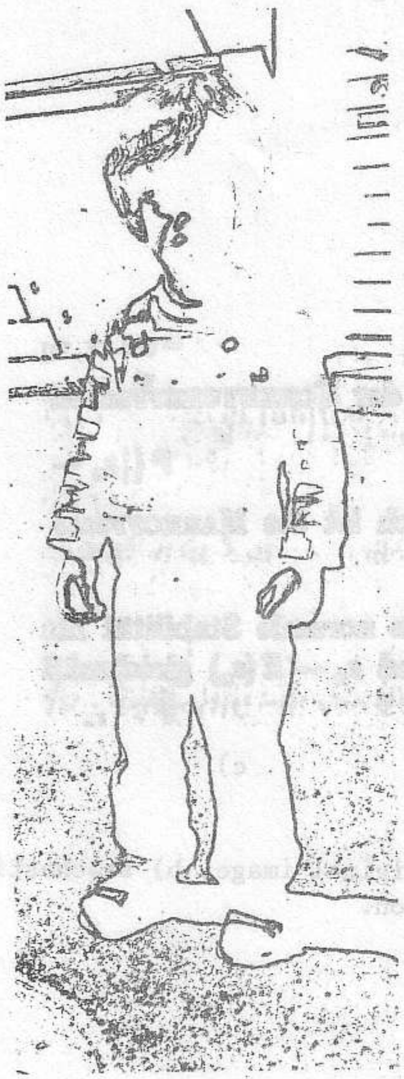


Figure 7. Result of the Sobel edge-detection algorithm.

3.5. Examples

We now present a series of images with the results produced by the stochastic algorithm described and by the Sobel algorithm. The examples are selected to demonstrate special problems of edge-detection in general. The scanned text image we present first shows noise due to the reverse shining through the paper (see Fig. 8). With requantization, the noise can be

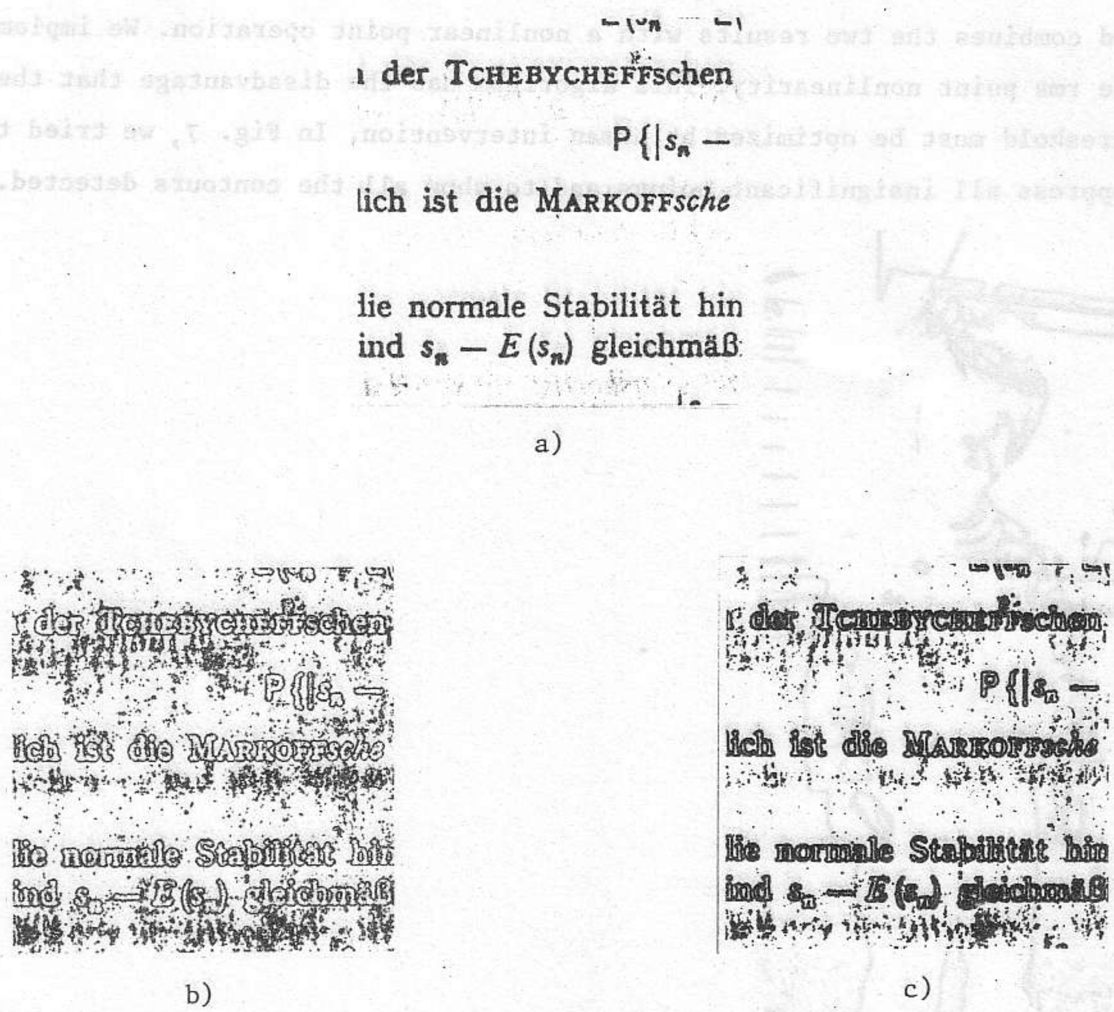
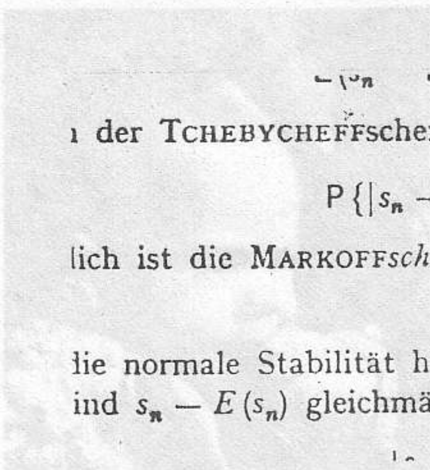
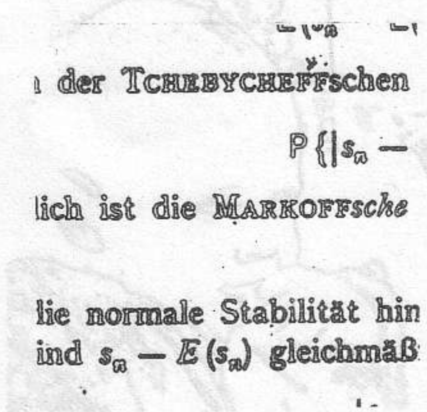


Figure 8. Edge detection on scanned text. a) original image, b) stochastic edge detection, c) Sobel edge detection.

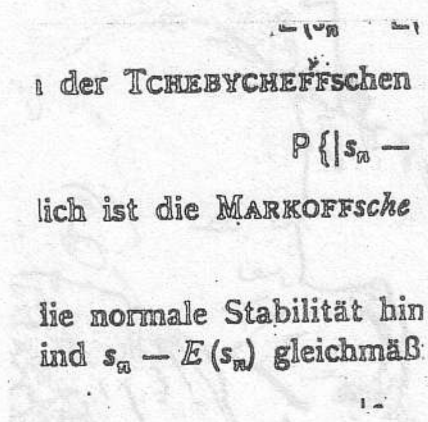
significantly reduced (see Fig. 9). But for this case, the contrast is so good that all edge-detection algorithms will detect more or less the same contours.



a)



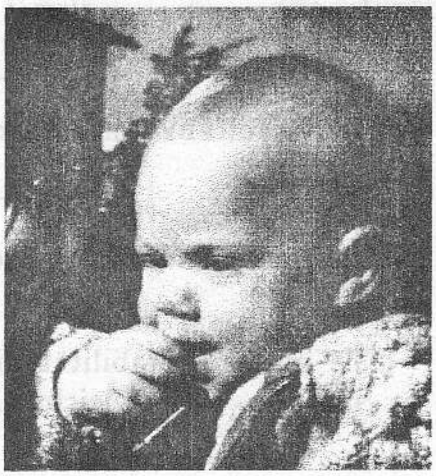
b)



c)

Figure 9. Edge detection on requantized scanned text. a) Original image, b) stochastic edge detection, c) Sobel edge detection.

In Fig. 10, the two algorithms detect different contours. The stochastic algorithm finds contours around the ear, while the Sobel algorithm gives a good contour of the tree on the right side of the house.



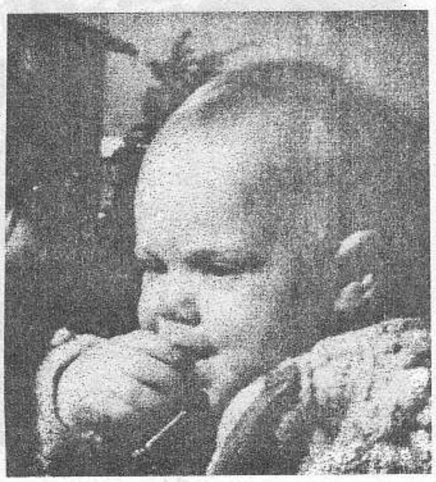
(a)

b)

c)

Figure 10. Edge-detection example. a) Original image, b) stochastic edge detection, c) Sobel edge detection.

Edge detection in noisy images is a problem. We changed 10% of the image points of the image in Fig. 10 randomly into random-generated grey levels. The Sobel algorithm fails, and it should be noted that both algorithms use the same number of image points to find the contours, but arranged in different forms (see Fig. 11).



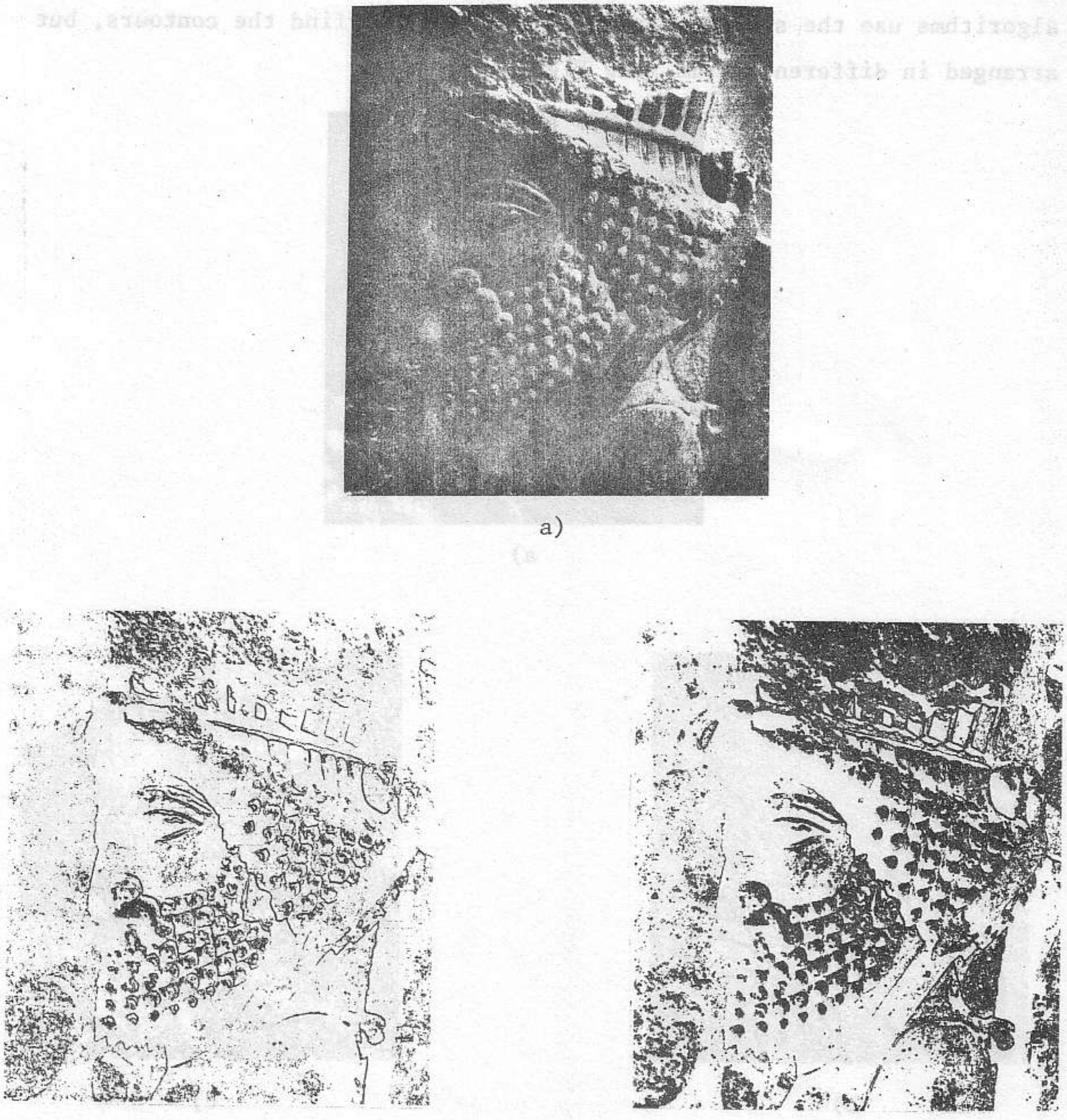
a)

b)

c)

Figure 11. Edge detection on the image of Fig. 10 with noise added. a) Original image, b) stochastic edge detection, c) Sobel edge detection.

We present two other examples in Figs. 12 and 13. The first image contains many changes in contrast which give the impression of noise. The second image has very unsharp contours which are hard to detect.



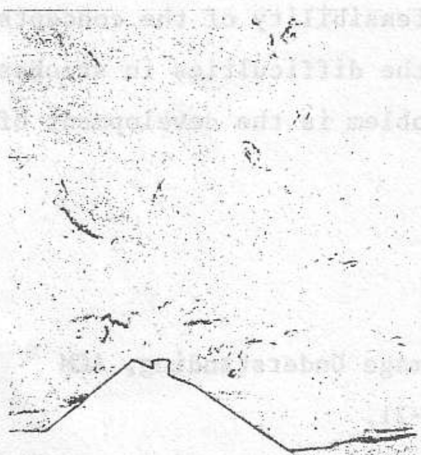
a)

b)

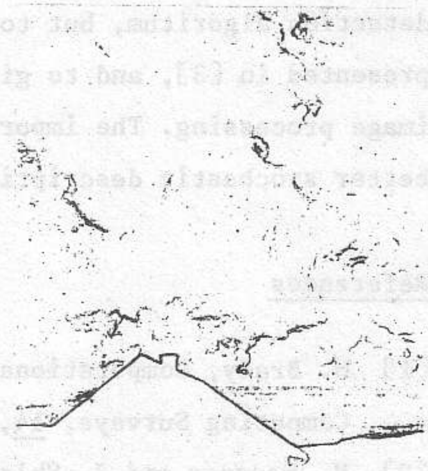
c)

Figure 12. Edge-detection example. a) Original image, b) stochastic edge detection, c) Sobel edge detection.

a)



b)



c)

Figure 13. Edge-detection example. a) Original image, b) stochastic edge detection, c) Sobel edge detection.

The last example should demonstrate that, for stochastic edge detection, larger regions must be taken instead of a cross with nine image points.

4. Conclusion

In two different processes, we modeled the rows and columns of an image as Markov chains. For these, we calculated a very rough approximation of the information density $J_T(t)$. An edge point is assumed when the information density $J_T(t)$ at point t is greater than the value for the independent case $I(t)$ for at least one of the two processes. To calculate the information density effectively, the probability distribution of the image must be given. It can be determined either via first and second-order statistics or via an artificial construction.

It is not satisfactory to divide the process into two processes, one in the x- and the other in the y-direction. In image processing, random functions with two-dimensional index sets should generally be used. The intent of this preliminary solution was not to present a final edge-detection algorithm, but to demonstrate the feasibility of the concepts presented in [3], and to give some ideas of the difficulties in stochastic image processing. The important remaining problem is the development of better stochastic descriptions for images.

References

[1] M. Brady, Computational Approaches to Image Understanding, ACM Computing Surveys, 14, 1, March 1982, 3-71.

[2] M. Hassner and J. Sklansky, The Use of Markov Random Fields as Models of Texture, Computer Graphics and Image Processing, 12, 4, April 1980, 357-370.

[3] R.F. Hauser, Localization of Information on Finite Random Functions, IBM Research Report RZ 1170, 1982.

[4] J.G. Kemeny and J.L. Snell, "Finite Markov Chains", D. van Nostrand Company Inc., Princeton, 1960.

[5] L. Kitchen and A. Rosenfeld, Edge Evaluation Using Local Edge Coherence, IEEE Trans. Systems, Man, and Cybernetics, SMC-11, 9, September 1981, 597-605.