

A VARIATIONAL APPROACH TO EDGE DETECTION

John Canny

338, MIT Artificial Intelligence Laboratory,
545 Technology Square,
Cambridge, MA 02139

Abstract

The problem of detecting intensity changes in images is canonical in vision. Edge detection operators are typically designed to optimally estimate first or second derivative over some (usually small) support. Other criteria such as output signal to noise ratio or bandwidth have also been argued for. This paper describes an attempt to formulate a set of edge detection criteria that capture as directly as possible the desirable properties of the detector. Variational techniques are used to find a solution over the space of all possible functions. The first criterion is that the detector have low probability of error i.e. failing to mark edges or falsely marking non-edges. The second is that the marked points should be as close as possible to the centre of the true edge. The third criterion is that there should be low probability of more than one response to a single edge. The third criterion is claimed to be new, and it became necessary when an operator designed using the first two criteria was found to have excessive multiple responses. The edge model that will be considered here is a one-dimensional step edge in white Gaussian noise although the same technique has been applied to an extended impulse or ridge profile. The result is a one dimensional operator that approximates the first derivative of a Gaussian. Its extension to two dimensions is also discussed.

1. Introduction

Edge detection forms the first stage in a very large number of vision modules, and any edge detector should be formulated in the appropriate context. However, the requirements of many modules are similar and it seems as though it should be possible to design one edge detector that performs well in several contexts. The crucial first step in the design of such a detector should be the specification of a set of performance criteria that capture these requirements. Some previous formulations have chosen the first or second derivative as the appropriate quantity to characterize step edges, and have formed optimal estimates of this derivative over some support. Examples of first derivative operators are the operators of Roberts (1965) and Prewitt (1970), while Modestino and Fries (1977) formed an optimal estimate of the two-dimensional Laplacian over a large support. Marr and Hildreth (1980) suggested the Laplacian of a broad Gaussian since it optimizes the trade-off in localization and bandwidth. There is a second major class of formulations in which the image surface is approximated by a set of basis functions and the edge parameters are estimated from the modelled image surface. Examples of this technique include the work of Hueckel (1971) and Haralick (1982). These methods allow more direct estimates of edge properties such as position and orientation, but since the basis functions are not complete, the properties apply only to a projection of the actual image surface on to the subspace spanned by the basis functions. However, the basis functions are a major factor in operator performance, especially its ability to localize edges. Finally, none of the above methods considers the problem of nearby operators responding to the same edge.

In this paper we begin with a traditional model of a step edge in white Gaussian noise and try to formulate precisely the criteria for effective edge detection. We assume that detection is performed by convolving the noisy edge with a spatial function $f(x)$ (which we are trying to find) and by marking edges at the maxima in the output of this convolution. We then specify three performance criteria on the output of this detector.

- (i) Low probability of error at each point. There should be a low probability of failing to mark real edge points, and low probability of falsely marking an edge point. Since both these probabilities are monotonically decreasing functions of the output signal to noise ratio, this criterion corresponds to maximizing signal to noise ratio.
- (ii) Good localization. The points marked as edges by the operator should be as close as possible to the centre of the true edge.
- (iii) Only one response to a single edge. This is implicitly captured in (i) since when two nearby operators respond to the same edge, one of them must be considered a false edge. However (i) considers a single output point and does not deal with the interaction between operators.

The first result of this analysis for step edges is that (i) and (ii) are conflicting and that there is a trade-off or uncertainty principle between them. Broad operators have good signal to noise ratio but poor localization and vice-versa. A simple choice of mathematical form for the localization criterion gives a product of a localization term and signal to noise ratio that is constant. Spatial scaling of the function $f(x)$ will change the individual values of signal to noise ratio and localization but not their product. Given the analytic form of a detection function, we can theoretically obtain arbitrarily good signal to noise ratio or localization from it by scaling, but not simultaneously. It can be shown that there is a single best shape for the function f which maximizes the product and that if we scale it to achieve some value of one of the criteria, it will simultaneously provide the maximum value for the other. To handle a wide variety of images, an edge detector needs to use several different widths of operator, and to combine them in a coherent way. By forming the criteria for edge detection as a set of functionals of the unknown operator f , we can use variational techniques to find the function that maximizes the criteria.

The second result is that the criteria (i) and (ii) by themselves are inadequate to produce a useful edge detector. It seems that we can obtain maximal signal to noise ratio and *arbitrarily* good localization by using a difference of boxes operator. The difference of boxes was suggested by Rosenfeld and Thurston (1971) and was used by Herskovitz and Binford (1970), who first suggested a criterion similar to (i). If we look closely at the response of such an operator to a step edge we find that there is an output maximum close to the centre of the edge, but that there may be many others nearby. We have not achieved good localization because there is no way of telling which of the maxima is closest to the true edge. The addition of criterion (iii) gives an operator that has very low probability of giving more than one maximum in response to a single edge, and it also leads to a finite limit for the product of localization and signal to noise ratio.

The third result is an analytic form for the operator. It is the sum of four complex exponentials and can be approximated by the first derivative of a Gaussian. A numerical finite dimensional approximation to this function was first found using a stochastic hill-climbing technique. This was done because it was much easier to write the multiple response criterion in deterministic form for a numerical optimization than as a functional of f . Specifically, the numerical optimizer provides candidate outputs for evaluation, and it is a simple matter to count the number of maxima in one of the outputs. To express this constraint analytically we need to find the expectation value of the number of maxima in the response to an edge, and to express this as a functional on f , which is much more difficult. The first derivative of a Gaussian has been suggested before (Macleod 1970). It is also worth noting that in one dimension the maxima in the output of this first derivative operator correspond to zero-crossings in the output of a second derivative operator.

All further results are related to the extension of the operator to two (or more) dimensions. They can be summarized roughly by saying that the detector should be directional, and if the image permits, the more directional the better. The issue of non-directional (Laplacian) versus directional edge operators has been the topic of debate for some time, compare for example Marr (1976) with Marr and Hildreth (1980). To summarize the argument presented here, a directional operator can be shown to have better localization than the Laplacian, signal to noise ratio is better, the computational effort required to compute the directional components is slight if sensible algorithms are used, and finally the problem of combining operators of several orientations is difficult but not intractable. It is for example much more difficult to combine the outputs of operators of different sizes, since their supports will differ markedly. For a given operator width, both signal to noise ratio and localization improve as the length of the operator (parallel to the edge) increases, provided of course that the edge does not deviate from a straight line. When the image does contain long approximately straight contours, highly directional operators are the best choice. This means several operators will be necessary to cover all possible edge orientations, and also that less directional operators will also be needed to deal with edges that are locally not straight.

The problem of combining the different operator widths and orientations is approached in an analogous manner to the operator derivation. We begin with the same set of criteria and try to choose the operator that gives good signal to noise ratio and best localization. We set a minimum acceptable error rate and then choose the smallest operator with greater signal to noise than the threshold determined by the error rate. In this way the global error rate is fixed while the localization of a particular edge will depend on the local image signal to noise ratio. The problem of choosing the best operator from a set of directional operators is simpler, since only one or two will respond to an edge of a particular orientation. The problem of choosing between a long directional operator and a less directional one is theoretically simple but difficult in practice. Highly directional operators are clearly preferable, but they cannot be used for locally curved edges. It is necessary to associate a goodness of fit measure with each operator that indicates how well the image fits the model of a linearly extended step. When the edge is good enough the directional operator output is used and the output of less directional neighbours is suppressed.

Following the analysis the author will outline some simple experiments which seem to indicate that the human visual system is performing similar selections (at some computational level), or at least that the computation that it does perform has a similar set of goals. We find that adding noise to an image has the effect of producing a blurring of the image detail, which is consistent with there being several operator sizes. More interestingly, the addition of noise may enable perception of changes at a large scale which, even though they were present in the original image, were difficult to perceive because of the presence of sharp edges. Our ability to perceive small fluctuations in edges that are approximately straight is also reduced by the addition of noise, but the impression of a straight edge is not.

2. In One Dimension

We consider first the one dimensional edge detection problem. The goal is to detect and mark step changes in a signal that contains additive white Gaussian noise. We assume that the signal is flat on both sides of the discontinuity, and that there are no other edges close enough to affect the output of the operator. The detection criterion is simple to express in terms of the signal to noise ratio in the operator output, i.e. the ratio of the output in response to the step input to the output in response to the noise only. The localization criterion is more difficult, but a reasonable choice is the inverse of the distance between the true edge and the edge marked by the detector. For the distance measure we will use the standard deviation in the position of the maximum of the operator output. By using local maxima we are making what seems to be an arbitrary choice in the mapping from linear operator output to detector output. The choice is motivated by the need for a local predicate which allows the third criterion to be met. Under these assumptions, the signal to noise ratio Σ can be shown to be (Canny 1983)

$$\Sigma = \frac{A \int_{-\infty}^0 f(x) dx}{n_0 \sqrt{\int_{-\infty}^{+\infty} f^2(x) dx}}$$

Where A is the amplitude of the input step, and n_0 is the root mean squared noise amplitude.

The localization Λ , defined as the reciprocal of the standard deviation of the position of the true edge, is given by

$$\Lambda = \frac{A}{n_0} \frac{f'(0)}{\sqrt{\int_{-\infty}^{+\infty} f^2(x) dx}}$$

The derivation of the localization criterion assumes that the function f is antisymmetric. This ensures that the expectation value for the output maximum is at the centre of the input edge.

Analysis of these equations shows that their product is independent of the amplitude of f and that the product is also independent of spatial scaling of f . However the individual terms do change, and if the width of f is increased by w , signal to noise ratio Σ increases by \sqrt{w} while localization Λ is reduced by the factor \sqrt{w} . This is the uncertainty principle relating the two criteria. Another way to view this relationship is that we are simultaneously trying to estimate the amplitude and position of the input step, and that because of image noise we can accurately estimate one only at the expense of the other. Note also that both criteria improve as the signal to noise ratio of the image $\frac{\Lambda}{n_0}$ improves.

We then employ the Calculus of Variations to find the function f which maximizes the product of these criteria. For simplicity, we consider a fixed width function, and assume that it is non-zero only in the range $[-2,2]$. Also since the function is known to be anti-symmetric we consider only the range $[0,2]$. The function f over this range can be shown to be

$$f(x) = a \left(1 - \frac{\cosh(\alpha(x-1))}{\cosh(\alpha)} \right)$$

Where a and α are undetermined constants from the optimization. It turns out that both Σ and Λ increase with α , and that the localization improves without bound. The function f tends to a difference of boxes as α tends to infinity. The third frame of Figure 1 shows this operator and the fourth frame is its response to the noisy edge. The analysis of signal to noise ratio so far has been concerned only with the probability of marking edges when near the centre of an input edge, and has not attempted to reduce the probability of marking edges given that a neighbouring operator has marked an edge. We can reduce the probability of marking multiple edges by constraining the distance between adjacent maxima in the response of the operator to noise. This distance is given by

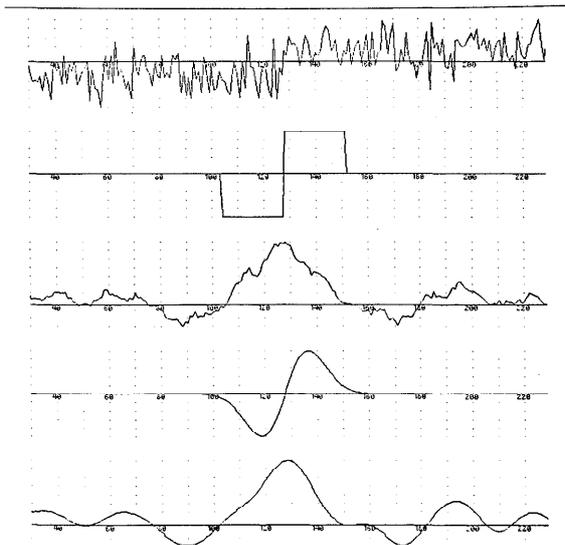
$$x_{maxima} = 2\pi \left(\frac{\int_{-\infty}^{+\infty} f'^2(x) dx}{\int_{-\infty}^{+\infty} f''^2(x) dx} \right)^{\frac{1}{2}}$$

When this constraint is added to the original two criteria, the solution of the variational problem becomes (in the range $[0,1]$)

$$f(x) = a_1 \exp(\alpha x) \cos(\omega x + \theta_1) + a_2 \exp(-\alpha x) \cos(\omega x + \theta_2) - \frac{\lambda_3}{2}$$

When the values of the constants a_1 , a_2 , θ_1 and θ_2 are solved for, we find that the above function can be approximated by the first derivative of a Gaussian, shown in the fifth frame of figure 1. It closely resembles Macleod's (1970) operator, and at least in one dimension it is equivalent to the second derivative zero-crossing operator of Marr and Hildreth (1980). However in two dimensions, the similarity to the zero-crossing operator ends.

Figure 1. Responses of difference-of-boxes and first derivative of Gaussian operators to a noisy step



3. Two or More Dimensions

The step edge is the one dimensional projection of the boundary between two image regions of differing intensity. In two dimensions it is a directional entity, and the direction of the edge contour will be normal to the direction of principal slope at the edge. The direction of principal slope gives us a coordinate to project the two dimensional edge in to a one dimensional step. In principal the projection can be done from any number of dimensions. The projection should use a smooth projection function, since this minimizes the rate at which the detected edge "zigzags" about its mean value. The simplest implementation of the detector uses a projection function that is a Gaussian of the same width as the detection function. This greatly simplifies the computation for the operator, and enables it to be expressed as the composition of convolution with a symmetric two-dimensional Gaussian followed by the application of the following non-linear differential predicate

$$\frac{\partial^2}{\partial n^2} G * I = 0$$

where $n = \frac{\nabla G * I}{|\nabla G * I|}$, G is a symmetric two-dimensional Gaussian, and I is the image

This operator actually locates either maxima or minima, by locating the zero-crossings in the second derivative in the edge direction. This is the derivative of the output of the operator in the projected coordinate. In practice only maxima in magnitude are used, since only these correspond to step edges. This form is readily extensible to higher dimensions, and the convolution with an n-dimensional Gaussian is highly efficient because the Gaussian is decomposable into n linear filters. Second directional derivative zero-crossings have been proposed by Haralick (1982).

It is worthwhile here to compare the performance of this kind of directional second derivative operator with the Laplacian. First we note that the two-dimensional Laplacian can be decomposed into components of second derivative in two arbitrary orthogonal directions. If we choose to take one of the derivatives in the direction of principal gradient, we find that the operator output will contain one contribution that is essentially the same as the operator described above, and also one from a detector that is aligned along the direction of the edge contour. This second component contributes nothing to localization or detection, but increases the output noise by 60%. At best we can expect the Laplacian of Gaussian to be worse by 60% in localization and detection than the simplest form of the detector. This has been verified in experiments.

We can gain considerable improvements in Σ and Λ by extending the projection function. However, a simple decomposition into a convolution and a differential operator is no longer possible. Instead several operators of fixed orientations are needed. The more the operator is extended along the contour direction, the more directional it becomes and the more operators are needed at each point to cover all possible edge orientations. The detector has been implemented using masks with 6 and 8 discrete orientations. The results are very promising. The implementation still uses a decomposition strategy to obtain computational efficiency. The image is first convolved with a symmetric Gaussian, and x and y components of gradient are found. Then the various oriented masks are obtained by convolving the slope components with sparse linear masks in the mask direction. Typically there are only 7 multiplications per mask per point.

A complete implementation of the detector using masks of differing widths and lengths exists, and the heuristics for combining operator outputs are being refined. One of the most interesting

questions involves the combination of the outputs of operators of different widths, since these will respond to intensity changes at different scales. While in most cases it is desirable to choose one operator to represent the intensity change at a particular point in an image, there are cases where changes are superimposed, and two or more edges need to be marked at the point. In particular, this is necessary whenever shading edges occur over surface detail.

4. A Simple Demonstration

The similarity of the detection scheme described above with the human visual system was never a motivating factor in the design of the algorithm. But once the algorithm has been formulated, it is interesting to inquire whether there are similarities. One of the results of the analysis is that small operators are always to be preferred when their signal to noise ratio is high enough, and that larger operators are progressively chosen as the signal to noise ratio in the image decreases. We might ask if a similar switching occurs in the human visual system. A well-known perceptual anomaly involves a coarsely quantized picture of a human face. Since the image is sampled very coarsely, the relevant perceptual information is at the low spatial frequencies. However the quantization adds sharp edges at high spatial frequencies which tend to override the coarse information. The high frequency detail can be reduced by low-pass filtering, i.e. blurring of the image. In the present scheme the transition from narrow to broad filters can also be accomplished by reducing the image signal to noise ratio.

Figure 2 shows a series of coarsely quantized images of a human face. The only difference between the images is that they contain increasing amounts of Gaussian noise. Remarkably, the addition of noise makes the later images easier to identify as a human face. Fig 3 shows a series of lines with sinusoidally varying direction. The latter image (which is the same set of lines with noise added) is perceived by many as containing straight lines.

The second effect is a less convincing test of the hypothesis that the human visual system gives preference to highly directional operators. The lines are closely spaced so that there will be improvement in signal to noise ratio with operator width only while the operator width is smaller than the projected distance between the lines on the retina. So in this context, improvements in signal to noise ratio can only be had by using more directional operators. The lines are approximately straight so that the directional operators will have sufficient "quality". Thus we would expect the highly directional masks (which are less sensitive to rapid changes in edge orientation) to be preferred, giving the impression of straight lines.

An example of the output from two operators of different size on a textured image is shown in figure 4. The image contains both sharp intensity changes due to material boundaries and slow changes due to reflectance variation. The two operators differ by a factor of four in width, and each one has equal width and length.

5. Conclusions

This paper has outlined the design of an edge detector from an initial set of goals related to its performance. The goals were carefully chosen with minimal assumptions about the form of an optimal edge operator. The constraints imposed were that we would mark edges at the maxima in the output of a linear shift-invariant operator. By expressing the criteria as functionals on the impulse response of the edge detection operator, we were able to optimize over a large solution space, without imposing constraint on the form of the solution.

Using this technique with an initial model of a step edge in white Gaussian noise, we found that there was a fundamental limit to the simultaneous detection and localization of step edges. This led to a natural uncertainty relationship between the localizing and detecting abilities of the edge detector. This relationship in turn led to a powerful constraint on the solution, i.e. that there is a class of optimal operators all of which can be obtained from a single operator by spatial scaling. By varying the width of this operator it is possible to vary the trade-off in signal to noise ratio versus localization, at the same time ensuring that for any value of one of the quantities, the other will be maximized.

The technique has also been applied to the derivation of optimal detectors for other types of feature, specifically for "roof" and "ridge" edges (Herskovits and Binford 1970). A ridge detector has been implemented is being tested on images of printed text. It is also possible to apply the technique to non-white Gaussian noise models (Canny 1983). This can be done by deriving a "whitening" filter for the noise and designing an optimal detector for the feature which results from the application of the whitening filter to the actual feature.

6. References

- Canny J. F. "Finding Edges and Lines in Images", S.M. Thesis, Department of Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge Mass., 1983.
- Haralick R. M. "Zero-crossing of Second Directional Derivative Edge Operator," S.P.I.E. Proceedings on Robot Vision, Arlington Virginia, 1982.
- Herskovits A. and Binford T. "On Boundary Detection," M.I.T. Artificial Intelligence Laboratory, Cambridge Mass., AI Memo 183, 1970.
- Hueckel M. H. "An Operator Which Locates Edges in Digitized Pictures," *JACM* 18, No. 1 (1971), 113-125.
- Macleod I. D. G. "On Finding Structure in Pictures," *Picture Language Machines* S. Kanefl ed. Academic Press, New York, pp 231, 1970.
- Marr D. C. "Early Processing of Visual Information," *Phil. Trans. R. Soc. Lond. B* 275 (1976), 483-524.
- Marr D. C. and Hildreth E. "Theory of Edge Detection," *Proc. R. Soc. Lond. B* 207 (1980), 187-217.
- Modestino J. W. and Fries R. W. "Edge Detection in Noisy Images Using Recursive Digital Filtering," *Computer Graphics and Image Processing* 6 (1977), 409-433.
- Prewitt J. M. S. "Object Enhancement and Extraction," *Picture Processing and Psychopictorics* B. Lipkin & A. Rosenfeld Eds, Academic Press, New York, pp 75-149, 1970.
- Roberts L. G. "Machine Perception of 3-Dimensional Solids," *Optical and Electro-Optical Information Processing* J. Tippett, D. Berkowitz, L. Clapp, C. Koester, A. Vanderbergh Eds, M.I.T. Press, Cambridge, pp 159-197, 1965.
- Rosenfeld A. and Thurston M. "Edge and Curve Detection for Visual Scene Analysis," *IEEE Trans. Computers* C-20, No. 5 (1971), 562-569.

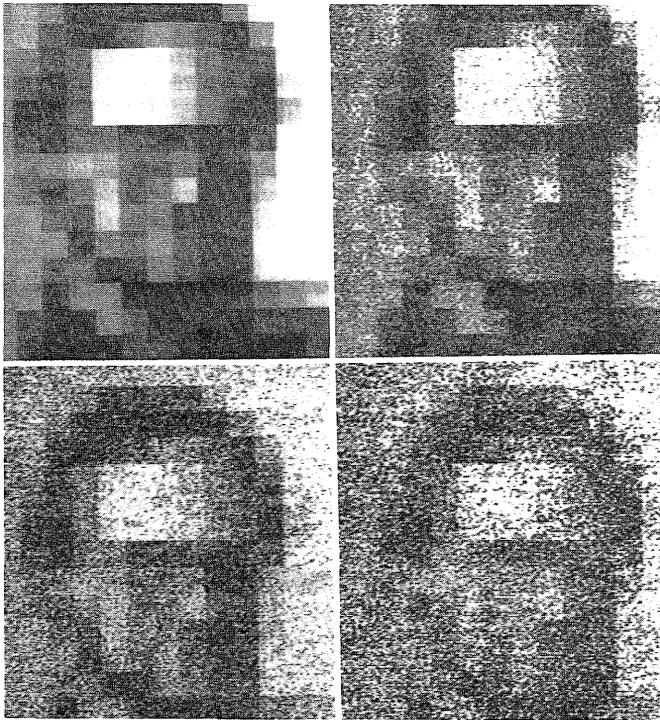


Figure 2. Coarsely sampled image of a human face with varying amounts of additive Gaussian noise

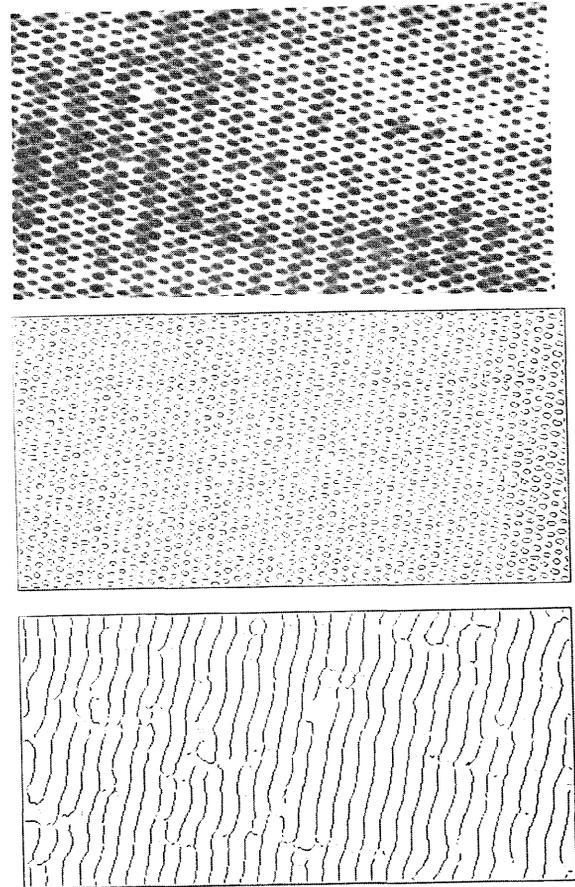


Figure 4. Outputs from two operators on a textured image

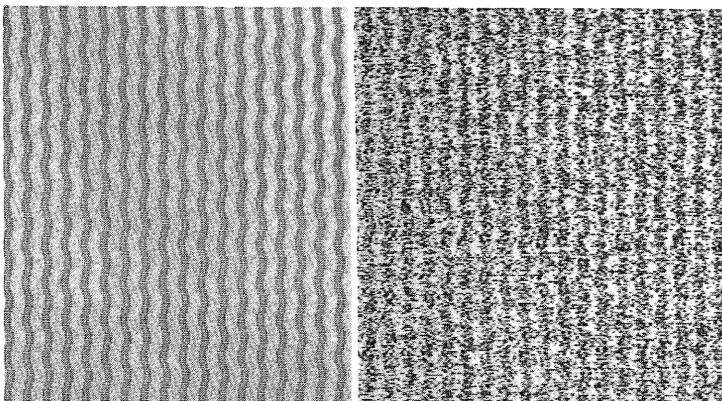


Figure 3. Parallel lines with sinusoidal variation in direction and additive Gaussian noise
