

What Can We Learn from Discrete Images about the Continuous World?*

Ullrich Köthe

Multi-Dimensional Image Processing Group, University of Heidelberg, Germany
Email: ullrich.koethe@iwr.uni-heidelberg.de

Abstract. Image analysis attempts to perceive properties of the continuous real world by means of digital algorithms. Since discretization discards an infinite amount of information, it is difficult to predict if and when digital methods will produce reliable results. This paper reviews theories which establish explicit connections between the continuous and digital domains (such as Shannon's sampling theorem and a recent geometric sampling theorem) and describes some of their consequences for image analysis. Although many problems are still open, we can already conclude that adherence to these theories leads to significantly more stable and accurate algorithms.

1 Introduction

As far as computer vision is concerned, the real world can be considered as a continuous domain.¹ With few exceptions, discrete atoms and molecules are below the relevant scales of image analysis. In contrast, computer-based reasoning is always discrete – an infinite amount of information is inevitably lost before algorithms produce answers. Experience tells us that these answers are nevertheless useful, so enough information is apparently preserved in spite of the loss. But failures in automatic image analysis are not infrequent, and the fundamental question under which conditions discrete methods provide valid conclusions about the analog real world is still largely unsolved.

This problem can be approached in different ways. In one approach (which is, for example, popular in physics), theories and models are formulated in the analog domain, e.g. by means of differential equations. Discretisation is then considered as an implementation issue that doesn't affect the theory itself. The correctness of discrete solutions is ensured by asymptotic convergence theorems: Discretization errors can be made as small as desired by choosing sufficiently fine

* The author gratefully acknowledges essential contributions by Hans Meine and Peer Steldinger and many fruitful discussions with Bernd Neumann, Hans-Siegfried Stiehl, and Fred Hamprecht.

¹ To avoid ambiguity, we shall use the term *analog* domain in the sequel. Otherwise, it might remain unclear whether a “continuous image” is an image function without discontinuities, or an image defined on a non-discrete domain. We will call the latter “analog image” instead.

discretizations, and automatic or interactive refinement mechanisms adaptively adjust the resolution as required.

Unfortunately, this approach doesn't work well in image analysis: Usually, we don't have the opportunity to take refined images if the original resolution turns out to be insufficient. Thus, we need *absolute* accuracy bounds (rather than asymptotic ones) to determine the correct resolution beforehand, or to decide what kind of additional information can be utilized if the desired resolution cannot be reached. Moreover, new discretisation schemes may be required because conventional ones are often unable to arrive at optimal results when the resolution is fixed. For example, [14] show that orientation artifacts in non-linear diffusion on regular grids can be significantly reduced by means of a better discretization of the diffusion operator.

A different approach has been taken by purely discrete methods such as digital geometry. Here, theories are formulated directly in the discrete domain, and implementations impose no further information loss. Algorithms of this type are automatically reliable within the scope of the underlying theories and models. However, the fundamental question of discretization remains unanswered: Digital theories take their (digital) input for granted and don't consider where it comes from and how it relates to the real world. This is no problem as long as one refers to naturally discrete entities of the real world (e.g. well distinguished objects). But pixels in digital images are not natural: They only arise from a particular image acquisition method, and their relation to the real world is largely coincidental. When the theoretical treatment only starts after discretization, one can never be sure that the desired information was present in the original pixel values to begin with.

A satisfactory discretization theory should therefore refer to the analog and digital domains simultaneously. It should establish connections between analog models and their digital representations which lead to realistic, yet tractable problem formulations and provable solutions. Well-defined transformations between the two domains should be possible, so that one can perform each task in the most suitable domain, without losing correctness guarantees. This paper attempts to give an overview over important results in this area.

2 Consequences of Shannon's Sampling Theorem

The information loss from the analog to the digital domain occurs in three stages (cf. figure 1): First, the 3-dimensional world is projected onto a 2-dimensional surface. Regardless of the particular projection method (perspective, fish-eye, catadioptric etc.), the depth dimension is lost in this transformation, but the resulting representation is still an *analog image* with infinite resolution. We refer to the mathematical projection of a real-world scene as the *ideal geometric image*. However, this geometric image is an idealization – any real optical device (e.g. a camera lens) imposes certain imprecisions such as geometric distortions and blurring. The second stage models the effect of these errors and results in the

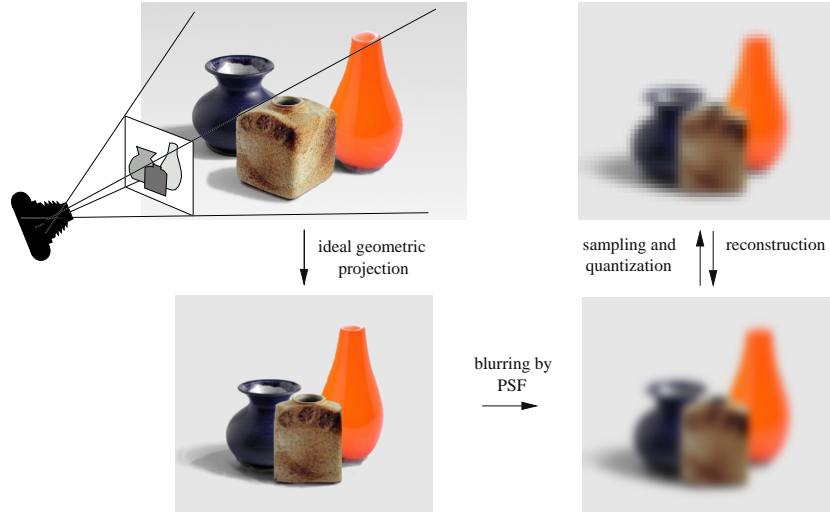


Fig. 1. The digitization model employed in this paper: the transformation from the real world to the digital image (top right) involves several stages of information loss.

real analog image. It still resides in the analog domain (i.e. contains an infinite amount of information), and is not observable.

The third stage, discretization plus addition of noise (including photon counting noise, electronic noise, and quantization round-off noise), finally produces an observable *digital image* that only contains a finite amount of data. In the sense of set cardinality, the information loss during the last stage (from uncountable to finite) is very dramatic. Yet, much less effort has been devoted to understanding the consequences of discretization than to those of projection.

In this paper, we will not deal with projection but assume that the ideal geometric image is given and geometric distortions can be neglected or have been accounted for by camera calibration. The ideal image will be defined in terms of partitions of the plane that represent a given set of ground-truth boundaries:

Definition 1. A partition of the plane is defined by a tuple $[V, A, R]$, where $V = \{v_i \in \mathbb{R}^2\}$ is a finite set of points and $A = \{a_i \subset \mathbb{R}^2\}$ a finite set of pairwise disjoint arcs such that each a_i is a bijective mapping of the open interval $(0, 1)$ into the plane, the end points $\lim_{t \rightarrow \{0,1\}} a_i(t)$ of each arc are in V , but no point of V belongs to an arc. R is a set of regions r_i which are defined as maximally connected components of the points not in V or any of the a_i . The union $B = \{\mathbf{x} : \mathbf{x} \in V \vee \mathbf{x} \in a_i\}$ of points and arcs determines the boundaries of the partition.

Regions can be described by indicator functions χ_{r_k} taking the value 1 in the interior of region r_k and 0 outside. An ideal geometric image is now the superposition of a set of continuous functions f_{r_k} describing the intensities in every

region

$$f_{\text{ideal}}(\mathbf{x}) = \sum_{r_k} \chi_{r_k}(\mathbf{x}) f_{r_k}(\mathbf{x})$$

such that f_{ideal} is discontinuous across region boundaries, i.e. at almost every boundary point $\mathbf{x} \in B$.² An important special case is the well-known *step-edge model*, where all f_{r_k} are constant functions. When ideal geometric images were observable, image segmentation would be trivial: one could simply detect image discontinuities. Unfortunately, these discontinuities are lost in the transformation to the observed digital image, so that segmentation is considerably more complicated. Therefore, the quality of a low-level segmentation method can be judged by *how well it recovers the boundaries of the ideal image from the actual digital image*.

To keep the transition from the ideal to the real analog image tractable, we assume that the amount of blurring does not vary within the field of view, so that it can be modeled by a convolution of the ideal image. Additional blurring is introduced by the sensitivity profile of the CCD sensor elements (which necessarily have non-zero area). All sources of blurring are conveniently combined into a convolution with a single *point spread function* (PSF) of the image acquisition device:

$$f_{\text{real}}(\mathbf{x}) = f_{\text{ideal}}(\mathbf{x}) \star \text{psf}(\mathbf{x})$$

The Fourier transform of the PSF is called the *optical transfer function* (OTF), and its magnitude is the *magnitude transfer function* (MTF). An upper bound for all possible OTFs is defined by the OTF of the ideal *diffraction-limited system*: Since any real lens has finite extent, some blurring is inevitably caused by diffraction at the lens' exit pupil. If there exist no other aberrations, the OTF is

$$\text{OTF}_{\text{ideal}}(\nu, \nu_0) = \begin{cases} \frac{2}{\pi} \left(\cos^{-1} \left(\frac{\nu}{\nu_0} \right) - \frac{\nu}{\nu_0} \sqrt{1 - \left(\frac{\nu}{\nu_0} \right)^2} \right) & \text{if } \nu < \nu_0 \\ 0 & \text{otherwise} \end{cases}$$

where ν is the radial coordinate in the Fourier domain, and $\nu_0 = \frac{d}{z\lambda}$ is the *band-limit* of the OTF (with d - diameter of the exit pupil, z - distance between exit pupil and image, λ - wave length of the light). No real lens can perform better than the diffraction limited lens with the same exit pupil and magnification: $|\text{OTF}_{\text{real}}(\nu)| \leq |\text{OTF}_{\text{ideal}}(\nu, \nu_0)|$. Experimentally one finds that many real PSFs are well approximated by Gaussian kernels

$$\text{psf}_{\text{real}}(\mathbf{x}) \approx \frac{1}{2\pi\sigma_{\text{PSF}}^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma_{\text{PSF}}^2}\right)$$

with $0.45 < \sigma_{\text{PSF}} < 0.9$ (measured in pixel distances). The practical band-limit of this PSF is even lower than that of the diffraction-limited system because the corresponding Gaussian OTF rapidly converges to zero, see figure 2.

² Since B is set of measure zero, the exact values of f_{ideal} on B can be arbitrary as long as they remain finite.

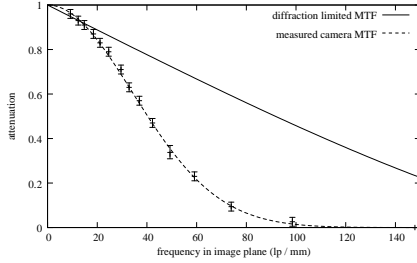


Fig. 2. The magnitude of the OTF of the ideal diffraction limited system and a real digital SLR camera. The sampling frequency of the latter’s sensor is about 120 pixels per mm.

The fact that real OTFs are band-limited has an important consequence: Since Shannon’s sampling theorem states that band-limited functions can be discretized and reconstructed without information loss (see e.g. [2]), this applies to real images as well, provided the sampling frequency ν_S is high enough: $\nu_S \geq 2\nu_0$ (Nyquist limit). In practice, the reconstruction will never be perfect (mainly due to noise, but also because camera designers increase subjective image sharpness – at the price of some aliasing artifacts – by slight under-sampling), but it is good enough to consider the reconstructed analog image a close approximation of the real analog image

$$f_{\text{reconstructed}}(\mathbf{x}) = f_{\text{real}}(\mathbf{x}) + n(f_{\text{real}}(\mathbf{x}), \mathbf{x}) \quad (1)$$

where $n(f_{\text{real}}(\mathbf{x}), \mathbf{x})$ is a noise process with band-limit $\nu_S/2$ that may depend on image position and intensity and subsumes all sources of statistical reconstruction error. In many cases, the noise is or can be transformed into homogeneous, intensity-independent Gaussian noise, which is the easiest to handle. Performing this reconstruction is not always necessary in practice, although it can be done efficiently and accurately by means of cubic or quintic splines

$$f_{\text{reconstructed}}(\mathbf{x}) \approx \sum_{k,l} f_{kl} s_m(x_1 - k) s_m(x_2 - l) \quad (2)$$

where $f_{kl} = f_{\text{real}}(k, l) + n_{kl}$ is the digital image, and $s_m(\cdot)$ is an m^{th} -order cardinal spline reconstruction kernel [13]. Formulas (1) and (2) are key ingredients for the desired correspondence between analog and discrete domains: they ensure that an *analog theory* derived for f_{real} can actually be implemented with high accuracy by *discrete algorithms*, provided the noise is handled in a statistically sound way.

Due to convolution with the PSF, the real analog image does no longer exhibit any discontinuities, so that edge detection must be based on an alternative boundary definition. Many popular definitions involve the computation of some non-linear filter, for example the squared magnitude of the Gaussian gradient. Here we find another important consequence of Shannon’s sampling theorem: When the image analysis algorithm includes *non-linear steps*, it is usually necessary to perform the analysis at a higher resolution than that of the original image in order to avoid further information loss. In other words, the image must be interpolated to a higher resolution *before* image analysis starts [7].

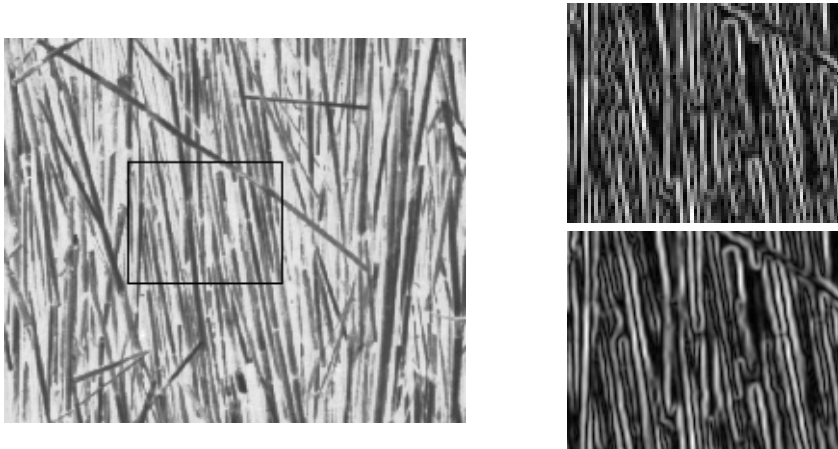


Fig. 3. Example image (Brodatz texture D15, left) where aliasing of the gradient magnitude at the original resolution is clearly visible (top right), whereas no aliasing occurs when the image is interpolated two-fold before gradient computation (bottom right).

We use the gradient to explain this seemingly counter-intuitive fact. The local maxima of the gradient squared magnitude along the gradient direction can be used as a boundary definition that replaces the (unobservable) discontinuities of the ideal image. Now suppose that the real analog image is band-limited according to the sampling rate of the sensor. When we apply any linear filter (e.g. a smoothing filter in order to reduce noise, or a derivative filter), the band-limit does not change, and the original resolution is still sufficient. However, the computation of the gradient squared magnitude involves the *square* of two linearly filtered images:

$$|\nabla f|^2 = f_x^2 + f_y^2$$

where f_x and f_y are derivatives along the horizontal and vertical directions respectively. As is generally known, a multiplication in the spatial domain corresponds to a convolution in the Fourier domain. If a band-limited spectrum is convolved with itself, the *band-limit is doubled*. Consequently, the gradient squared magnitude must be represented with doubled resolution in every direction (i.e. four times as many pixels). Figure 3 shows an example where severe aliasing artifacts occur when this is not done. Similar conclusions can be drawn for other non-linear methods, and it can generally be observed that image analysis results improve when the image is interpolated at the beginning.

3 Geometric Sampling Theorems

Shannon's sampling theorem guarantees a close relationship between a band-limited analog image function and its digitization. It is therefore a fundamental

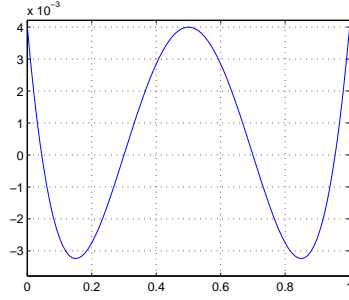


Fig. 4. A band-limited function according to (3) with 4 zero crossings in a single pixel.

justification for digital image analysis in a continuous world. However, it only makes statements about iconic representations of the world, i.e. about the data at the initial stages of image analysis. But image analysis ultimately aims for descriptions at much higher levels of abstraction which are outside the scope of the sampling theorem.

We can demonstrate the limitations of Shannon’s theorem by looking at basic geometric properties of band-limited 1-dimensional functions. Suppose we are interested in object edges defined by a level line of the image function. For simplicity, we choose level zero. When the image function is band-limited according to the grid’s Nyquist limit, there will be at most one zero crossing per pixel *on average* (corresponding to the highest frequency component of the data). But the actual distribution of zero crossings can be rather uneven – there may be large regions without zero crossings, but there can also be many zeros in a very small interval. In fact, we can construct band-limited functions with as many zero-crossings in a given interval as we desire by means of a superposition of integer translates of the function $\text{sinc}(x - k) = \frac{\sin(\pi(x-k))}{\pi(x-k)}$ (I’m indebted to P. Stelldinger for suggesting this construction). Since sinc-functions are band-limited, their superpositions are band-limited as well. Now consider, for example, the interval $[0, 1]$. A pair of two sinc-functions

$$s(x, k) = \text{sinc}(x - k - 1) + \text{sinc}(x + k)$$

has zeros at 0 and 1. Using two such pairs, we arrive at the function

$$f(x) = s(x, 1) + \frac{5}{3}s(x, 2) + 0.004 \quad (3)$$

which has four zeros in the interval $[0, 1]$ (approximately at 0.045, 0.3, 0.7, and 0.955, see figure 4). Obviously, more pairs can be added to this function to achieve arbitrary complicated behavior in the interval $[0, 1]$. In other words, even if the image function is band-limited, regions can become arbitrary narrow. Thus, the property of band-limitation is not sufficient to derive resolution requirements for the representation of *geometric features* such as region boundaries.

To overcome this limitation, it is necessary to complement the signal-theoretic sampling theorem with dedicated *geometric sampling theorems*. Such theorems

must be based on certain assumptions on the geometry of the regions in the ideal geometric image, e.g. sufficient region size and limited boundary curvature. Necessary resolution requirements are then derived from these assumptions in conjunction with an adequate model of the image acquisition process. The challenge in developing these theories is to make them realistic while maintaining mathematical and algorithmic tractability – many existing attempts employ simplifications that exclude most of the practically relevant images.

One typical simplification is the notion of *r-regular shapes* first introduced in [9,11]. Here it is assumed that all objects in the ideal geometric image are invariant under morphologically opening and closing with a disc of radius r . This definition has a very undesirable consequence: Since the boundary curvature of r -regular shapes cannot exceed $1/r$, shapes are not allowed to have corners. This may not look like a very severe restriction at first because one can approximate corners as accurately as desired by choosing a small r . But this refinement strategy does not always work. Most importantly, it fails at occlusions: Occlusions give rise to configurations where the boundaries of three objects (the background, the occluding object, and the occluded object) meet in a single point, and it is impossible to arrange the configuration without a sharp corner in at least one of the three regions. Unfortunately, occlusions abound in many real images due to the very nature of projections from 3D to 2D.

Another unrealistic simplification is the direct use of the ideal geometric image in sampling analysis. For example, digital straight line algorithms are usually motivated by the direct digitization of a Euclidean straight line, for example by means of grid intersection or subset digitization, cf. [6]. That is, pixels (or interpixel cracks) are assigned to a digital line according to rules describing the relationship between the grid and the ideal Euclidean shape. However, this approach is far from what is actually happening in a camera: In reality, the ideal geometric image cannot be observed, and the intrinsic blurring and noise of the image acquisition process must be taken into account. We already mentioned the most important consequence of this: Convenient boundary definitions on the basis of set intersections and discontinuities are not applicable to the real camera image and must be replaced.

A third shortcoming in traditional geometric sampling analysis is the introduction of unnecessarily hard requirements on the quality of the digitization. In particular, it is often required that the digital image be topologically equivalent, i.e. homeomorphic, to the ideal geometric image. This excludes certain digitizations that we would intuitively consider as correct. For example, a junction of degree 4 in the ideal image is often digitized into a pair of junctions of degree 3 which cannot be continuously transformed into a 4-junction as required by topological equivalence (figure 5 left). Similarly, when a closed contour in the ideal image (i.e. a 1-dimensional manifold) is mapped onto a closed chain of 8-connected pixels (i.e. a 2-dimensional union of squares), this mapping cannot be continuous (figure 5 right). On the other hand, purely topological quality criteria are insufficient because they disregard geometric similarity between the ideal and digital shapes – since homeomorphisms can be arbitrary continuous

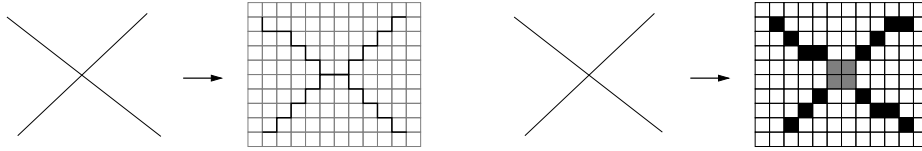


Fig. 5. Two cases where an intuitively correct boundary digitization is not homeomorphic to the ideal geometric image.

distortions, topologically “correct” digitizations may actually fail to be close to the original shape by any intuitively plausible criteria. For example, the letters “L” and “G” are very different, yet topologically equivalent. In other words, a notion of geometric closeness should also be part of digitization analysis.

We addressed these shortcomings in [12]: There, we relaxed the requirement of topological equivalence to homotopy equivalence. In 2D, this is easily captured by the notion of *homotopy trees*:

Definition 2. *The homotopy tree of a partition P of the plane is defined as follows: The infinite region (containing the point at infinity) is the root of the tree. All connected boundary components adjacent to the infinite region are its children. In turn, the regions adjacent to these boundary components (if not already in the tree) become their children. This process is repeated recursively, alternating between boundary and regions levels, until all elements of P have been added to the tree.*

It has been shown by Serra [11] that two partitions of the plane are of the *same homotopy type* if and only if their homotopy trees are isomorphic, which is relatively easy to check. The definition ensures that the regions of a partition are correctly reconstructed, whereas certain errors in the boundary are permitted. For example, the digitizations in figure 5 are correct (have the same homotopy type as the original) under the new requirement. Homotopy allows us to relax the strict requirement of r -regular shapes into a much more realistic shape class [8,12]:

Definition 3. *A partition P of the plane is called r -stable when the boundary B of P can be morphologically dilated with an open disc of radius r without changing the homotopy tree of the partition.*

This definition essentially means two things: First, regions must be large enough for an r -disc to fit into them. Second, regions may not have narrow waists where boundary dilation would lead to splitting of a region into two or more disconnected parts. On the other hand, shapes with corners and junctions are allowed. Dilation extends the true boundary into a band of width $2r$ wherein the detected boundary may be distorted without changing the homotopy type of the partition. The key effect of this possibility is that segmentation needs no longer be error free: As long as errors are confined to the band, correct reconstruction is still possible. In fact, we have been able to prove the following geometric sampling theorem:

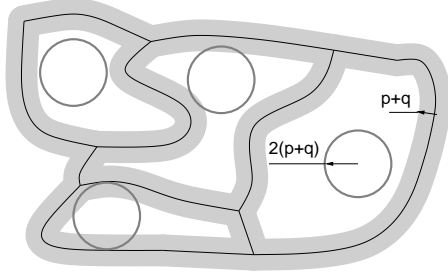


Fig. 6. Any r -regular shape (black lines) can be correctly segmented when boundary detection errors are confined to the gray strip around the true boundary, and every region contains a sufficiently large circle.

Theorem 1. *Let S be a finite set of points approximating the true boundary B of a partition P of the plane such that the distance of any point on the true boundary from S is at most p , and the distance of any point in S from the true boundary is at most q . Then it is possible to link the points of S into the boundary B' of a reconstructed partition P' which has the same homotopy type as P , provided the following conditions are fulfilled: (i) P is r -stable with $r > 2q$ and $r > p+q$, and (ii) every region of P contains a disc with radius $r' > 2(p+q)$.*

The theorem is illustrated in figure 6. Its proof, as well as a detailed description of the reconstruction algorithm (based on Delaunay triangulation and α -shapes) can be found in [8,12]. The theorem does in no way depend on the method applied to compute the points in S , as long as the errors do not exceed the given bounds. Or, putting it the other way, correct segmentation is possible and indeed guaranteed when the regions in the true partition are, in a very precise sense, large enough with respect to the errors of the segmentation algorithm.

4 Error Analysis of Segmentation Methods

The geometric sampling theorem outlined in the last section poses the challenge of estimating the errors of a segmentation method. We must distinguish two kinds of error: systematic and statistical ones. The former arise mainly from the fact that the ideal, discontinuity-based boundary definition has to be replaced with some alternative definition (cf. section 2). Due to the unavoidable information loss incurred by a band-limiting pre-filter (i.e. the blurring effect of the PSF), these alternative definitions cannot in general reproduce the true boundaries exactly. Statistical errors are, on the one hand, a consequence of the inherent noise of any real image acquisition system. On the other hand, aliasing and round-off errors arising due to digitization (spatial sampling and intensity quantization) can also be treated as noise to good approximation.

Let us first look at systematic errors. Obviously, their nature and magnitude depend on the boundary definition applied. One particularly well understood possibility are the zero-crossings of the oriented second Gaussian derivative along the direction of the Gaussian gradient. They are defined as follows

$$f_x^2 f_{xx} + 2f_x f_y f_{xy} + f_y^2 f_{yy} = 0 \quad (4)$$

where f_x , f_y , f_{xx} , f_{yy} , and f_{xy} denote first and second derivatives of the reconstructed analog image $f_{\text{reconstructed}}$ after filtering with a Gaussian at scale σ_{filter} . Recall that $f_{\text{reconstructed}}$ is a very good approximation of f_{real} when the PSF is band-limited and noise is not too strong, cf. equation (1). Also note that the zero-crossings differ from the true edges due to two sources of blurring: first by the PSF at scale σ_{PSF} , and then by the noise reduction filter at scale σ_{filter} . Since both blurring kernels are assumed to be Gaussians, the total blur can be expressed by a single scale parameter $\sigma = \sqrt{\sigma_{\text{PSF}}^2 + \sigma_{\text{filter}}^2}$. It can be shown that this blurring causes the following errors:

1. When the true edge is perfectly straight, and the intensities conform to the step-edge model (i.e. are constant within each region), no systematic errors occur, i.e. the true edge position can be estimated without bias.
2. When the true edge is circular with radius R and the regions still conform to the step edge model, the detected zero-crossing edge is located at the position $r = r_0$ which is the solution to the implicit equation

$$R \left(I_0 \left(\frac{r R}{\sigma^2} \right) + I_2 \left(\frac{r R}{\sigma^2} \right) \right) - 2r I_1 \left(\frac{r R}{\sigma^2} \right) = 0$$

where I_k are modified Bessel functions of order k [1]. This is the essential part of the second derivative of the image along the gradient direction, expressed in polar coordinates with origin at the center of curvature. It turns out that the detected edge is shifted toward the concave side of the true edge, and the bias can be well approximated by

$$\frac{r_0 - R}{\sigma} = 0.52 \sqrt{0.12^2 + \left(\frac{\sigma}{R} - 0.476 \right)^2} - 0.255$$

when $\sigma < R/2$ (which is true in most practical situations). For $\sigma < R/5$, the even simpler approximation $\frac{r_0 - R}{\sigma} \approx -\frac{\sigma}{2R}$ holds. The absolute bias $\Delta x = r_0 - R$ is thus essentially proportional to σ^2 , the *square of the blurring scale*.

3. When the edge is straight, but the adjacent regions are shaded (i.e. do not conform to the step edge model), the bias depends on the orientation and strength of the shading gradient. When shading occurs perpendicular to the edge, and the resulting intensity profile can be described by two linear equations separated by a discontinuity of height b , the bias is

$$\Delta x = \frac{a}{b} \sigma^2$$

where a is the slope difference between the shading in the two regions (measured in the ideal geometric image). Again, the bias is proportional to the square of the scale σ .

4. At a sharp corner with opening angle β , the bias is maximal along the bisector of the corner. [10] derived the formula

$$\Delta x = \sigma x' \sqrt{1 + \left(\tan \frac{\beta}{2} \right)^2}$$

where x' is the solution of the implicit equation

$$\frac{1}{\sqrt{2\pi}} e^{-x'^2/2} - \left(\tan \frac{\beta}{2}\right)^2 \frac{x'}{2} \left(1 + \operatorname{erf}\left(\frac{x'}{\sqrt{2}}\right)\right) = 0 \quad (5)$$

It can be seen that the displacement increases as the corner gets sharper. For $\beta = 90^\circ$, the bias is about 0.7σ , but it reaches 2.2σ for $\beta = 15^\circ$ and increases without bound for $\beta \rightarrow 0$.

5. At junctions of degree 3 and higher, the bias depends on the angles and intensities of the adjacent regions in very complicated ways, see e.g. [4]. Maximum errors can be determined experimentally and may be as high as 4σ . This mainly reflects the fact that equation (4) is only a very coarse approximation of the ideal geometric image near junctions. However, it is not at all easy to come up with an alternative boundary definition that produces consistently superior results, unless one is willing to give up generality and make heavy use of application-specific knowledge. A general definition is an unsolved problem of low-level image analysis.

The above theoretical predictions of the systematic errors of Gaussian zero-crossings are well confirmed by experiment, provided that edge positions are computed with floating-point accuracy and not rounded to integer coordinates. It should be stressed that errors frequently increase linearly or even quadratically with scale. This suggests that noise reduction filters should be chosen as small as possible in order to keep the induced bias to a minimum, unless we are certain that the image does not contain boundaries of the critical types.

Expected statistical errors can be derived from a suitable noise model. Photon counting sensors usually have a Poisson noise characteristic, i.e. the noise variance increases linearly with intensity I : $\sigma_{\text{noise}}^2 = sI + t$, where the parameters s and t are determined by camera calibration. Since it is very inconvenient to handle intensity-dependent noise in algorithms, it is preferable to transform Poisson noise into unit-variance Gaussian noise by means of a noise normalization transformation

$$I' = \frac{2}{s} \sqrt{sI + t} - \frac{2\sqrt{t}}{s} \quad (6)$$

where I' is the intensity after the transformation [5].³ Other noise sources (electronic, quantization) are typically of much smaller magnitude. We can therefore assume that the image is corrupted by additive white Gaussian noise. The noise standard deviation matters only in relation to the step height of the relevant discontinuities. We define the signal-to-noise ratio $SNR = b/\sigma_{\text{noise}}$ as the quotient of the step height and the noise standard deviation.

The effect of intensity noise on the localization accuracy of zero-crossing edges has, for example, been studied by [3]. Without loss of generality, one can

³ Interestingly, most consumer cameras are internally performing a similar transformation, called γ -correction. Its main purpose is the creation of a perceptually more uniform color space (usually sRGB), but approximate noise normalization is achieved as a side effect since γ -correction curves are very similar to (6).

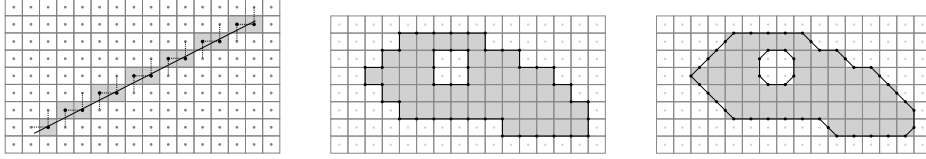


Fig. 7. Grid intersection digitization, subset digitization, mid-crack digitization of the reconstructed boundary.

assume that the true edge runs vertically through the coordinate origin. Then all y -derivatives in (4) vanish, and the estimated zero-crossing will be displaced along the x -axis by Δx . Formula (4) can be expanded into a Taylor series around $x = 0$, so that we get the estimated zero-crossing as the solution of

$$f_{xxx}(x)|_{x=0} \Delta x + n_{xx}(\Delta x) = 0$$

where $f_{xxx}(x)|_{x=0} = -\frac{b}{\sqrt{2\pi}\sigma^3}$ is the third derivative of the noise-free image blurred by both the PSF and noise reduction filters (i.e. with a total scale of $\sigma = \sqrt{\sigma_{\text{PSF}}^2 + \sigma_{\text{filter}}^2}$), and n_{xx} is the second derivative of the filtered noise (with only σ_{filter} , because the PSF acts before noise is added). The expected value of $|\Delta x|^2$ can be obtained as the quotient between the variance $N_{xx}^2 = \frac{3\sigma_{\text{noise}}^2}{16\pi\sigma_{\text{filter}}^2}$ of the second noise derivative, and the square of the third signal derivative. The standard deviation of the statistical edge displacement is therefore

$$\text{StdDev}[\Delta x] = \frac{\sigma_{\text{noise}}}{b} \frac{\sqrt{6}}{4} \left(1 + \frac{\sigma_{\text{PSF}}^2}{\sigma_{\text{filter}}^2}\right)^{3/2} \quad (7)$$

In typical situations, $\sigma_{\text{PSF}} < \sigma_{\text{filter}}$, so that $\text{StdDev}[\Delta x] \approx 1/\text{SNR}$. Typical signal-to-noise ratios in images of standard quality are above 10, so that $\text{StdDev}[\Delta x] < 0.1$ pixels is perfectly possible. The maximum error is below $3 \text{StdDev}[\Delta x]$ in 99.7% of the cases. Therefore, the maximum noise-induced edge displacement is of the order of $3/\text{SNR}$, provided that edges are represented as polygons with floating-point knot coordinates.

When edges are represented on the grid instead, coordinate round-off errors must be accounted for as well. The magnitude of these errors depends on the particular method of rounding. We distinguish three common cases, cf. figure 7:

1. Grid intersection digitization: Whenever the polygonal edge crosses a grid line (i.e. a line connecting two directly adjacent pixel centers), the edge is rounded to the nearest pixel coordinate (both coordinates are integers).
2. Subset digitization: Pixels are marked as being either inside or outside of an object of interest. The resulting boundary is therefore an interpixel boundary, and the edge coordinates are rounded to pixel corners (both coordinates are half-integers).

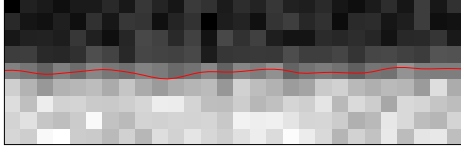


Fig. 8. Error correlation along a noisy straight edge at $SNR = 10$ and $\sigma_{\text{filter}} = 1$. The true edge runs in the image center. On average, the error of the detected edge has the same sign over about 4 pixels.

3. Mid-crack digitization: The boundary is defined as in subset digitization, but the coordinates are rounded to the midpoint between two directly adjacent pixels (one coordinate – either x or y – is integer, the other one half-integer).

In either case, round-off errors are uniformly distributed between the maximum possible displacements to the left and right of the polygonal boundary. It is easily verified that the standard deviation of a uniform distribution is $\text{StdDev}[\Delta x] = l/\sqrt{12}$, where l is the length of the permissible interval, and the maximum error is $l/2$. In cases 1 and 2, this interval is the pixel diagonal, i.e. $l = \sqrt{2}$, whereas in case 3 the interval is the pixel side length, i.e. $l = 1$. Since the round-off error is independent of the noise error, the total maximum error is the sum of the two maximum errors. In other words, the round-off error is significantly larger than the noise error.⁴ We conclude once again that it is a good idea to perform image analysis at a higher resolution than the original image, either by representing edges as polygons with floating-point accuracy, or by interpolation to a denser grid. If the latter approach is chosen, mid-crack edges are to be preferred over other possibilities because they minimize the error at a given resolution.

Another consequence of the above noise analysis is the fact that the displacements of neighboring edge points are highly correlated. In case of noise-induced displacements according to (7), the error auto-correlation along the edge, or equivalently the noise power spectrum can be explicitly computed

$$\|\mathcal{F}[\Delta x](u)\|^2 = 2\sqrt{\pi}\sigma_{\text{filter}} \text{StdDev}[\Delta x]^2 e^{-4\pi^2 u^2 \sigma_{\text{filter}}^2}$$

where $\mathcal{F}[\Delta x]$ denotes the Fourier transform of the error. An immediate conclusion from this formula is that the correlation length along the edge is rather high: On average, the error Δx keeps its sign over an interval of $\sqrt{2}\pi\sigma_{\text{filter}}$ pixels along the boundary. That is, even for small filters with $\sigma_{\text{filter}} = 1$, the edge is displaced to the same side of the true boundary over more than 4 pixels, see figure 8. The situation with round-off error is similar: As is well known, a digital straight line contains only two different directions, and the error sign remains equal within one half of the longest single-direction run. Thus, the correlation is highest for edges running on or close to the principal grid directions. In fact, the error of a perfectly horizontal or vertical digital line will have the same sign throughout.

⁴ This is in contrast to the situation for intensity quantization. Here, the noise standard deviation (usually about 3 gray levels) is much larger than the quantization standard deviation of $1/\sqrt{12}$ gray levels.

This error correlation has important consequences for certain methods that attempt to improve an initial boundary (or properties derived from it) by filtering the edge coordinates *along* the boundary. This includes, in particular, methods that estimate the boundary tangent direction by means of derivative filters or geometric fits applied to the points of the boundary: Due to correlation, small filtering or fitting windows will not lead to significant improvements of the estimates from the raw boundary – correlated errors can only be averaged out when the windows are significantly larger than the correlation length. Sufficiently large windows will only be available for rather large objects. Consequently, it can be shown both theoretically and experimentally that tangent angle estimates derived from the direction $\phi_{\text{tangent}} = \arctan(-f_x/f_y)$ perpendicular to the simple Gaussian gradient of the gray levels are often vastly superior to much more complicated estimates derived from the boundary itself.

5 Conclusions

This contribution demonstrated that image analysis theories can and, in fact, should be constructed so that they apply to both the analog and digital domains. This approach offers the possibility to prove that computed properties actually correspond to properties of the analog real world. In contrast, no such guarantees can be established when the two domains are considered in isolation. We reported a number of useful and encouraging results, but the existing body of knowledge barely scratches the surface of this important area, and much remains to be desired.

First, the realism of the theories has still to be enhanced in order to be of universal practical value. For example, the geometric sampling theorem 1 is based on bounds for maximum edge detection errors. However, these bounds are quite large for certain boundary configurations such as junctions, and may become arbitrarily big if the edge detector produced some artifact edges (outliers). We would like to reduce these errors by using a more accurate and more robust boundary definition for blurred noisy images, but we would also like to generalize the theorem so that it gives probabilistic guarantees when the requirements can only be fulfilled with certain probabilities. Moreover, it would be highly desirable to prove similar theorems for higher dimensions and other image analysis tasks like object recognition – these bounds may differ in interesting ways from the bounds for segmentation.

Another important achievement would be the development of a truly multidimensional signal theory. Up to now, Fourier transforms, filters and related techniques are often generalized to 2- and higher dimensional spaces by simple outer products of 1-dimensional spaces. In practice, this often means that the desirable rotational invariance of the Euclidean space is lost during digitization, so that digital algorithms exhibit significant anisotropy. Solutions to this problem might eventually be built into dedicated image analysis cameras that are optimized for the requirements of digital computers, and not for that of human viewers like most current cameras.

References

1. H. Bouma, A. Vilanova, L.J. van Vliet, F.A. Gerritsen: "Correction for the Dislocation of Curved Surfaces Caused by the PSF in 2D and 3D CT Images", IEEE Trans. Pattern Analysis and Machine Intelligence, 27(9):1501–1507, 2005
2. R.N. Bracewell: "The Fourier Transform and its Applications", New York: McGraw-Hill, 1978
3. J. Canny: "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679-698, 1986
4. R. Deriche, G. Giraudon: "A computational approach for corner and vertex detection", Intl. Journal of Computer Vision, 10(2):101-124, 1993
5. W. Förstner: "Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images", Proc. Summer School on Data Analysis and the Statistical Foundations of Geomatics, Lecture Notes in Earth Science, Berlin: Springer, 1999
6. R. Klette, A. Rosenfeld: "Digital Geometry", Amsterdam: Elsevier, 2004
7. U. Köthe: "Edge and Junction Detection with an Improved Structure Tensor", in: B. Michaelis, G. Krell (Eds.): Pattern Recognition, Proc. of 25th DAGM Symposium, Magdeburg 2003, Lecture Notes in Computer Science 2781, pp. 25-32, Berlin: Springer, 2003
8. U. Köthe, P. Stedding, H. Meine: "Provably Correct Edgel Linking and Subpixel Boundary Reconstruction", in: K. Franke, K.-R. Müller, B. Nikolay, R. Schäfer (Eds.): Pattern Recognition, Proc. DAGM 2006, Lecture Notes in Computer Science 4174, pp. 81-90. Berlin: Springer, 2006
9. T. Pavlidis: "Algorithms for Graphics and Image Processing", Rockville: Computer Science Press, 1982
10. K. Rohr: "Localization Properties of Direct Corner Detectors", Journal of Mathematical Imaging and Vision 4:139-150, 1994
11. J. Serra: "Image Analysis and Mathematical Morphology", New York: Academic Press, 1982
12. P. Stedding, U. Köthe, H. Meine: "Topologically Correct Image Segmentation Using Alpha Shapes", in: DGCI'06, Lecture Notes in Computer Science, Berlin: Springer, 2006
13. M. Unser, A. Aldroubi, M. Eden: "B-Spline Signal Processing", IEEE Trans. Signal Processing, 41(2), pp. 821-833 (part I), 834-848 (part II), 1993
14. J. Weickert, H. Schar: "A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance", J. Visual Comm. Image Repr., 13(1/2):103-118, 2002