Errors in Binary Images and an L^p Version of the Hausdorff Metric

A.J. Baddeley
CWI, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Department of Mathematics and Computer Science
University of Leiden, The Netherlands

This paper introduces a metric on the class of binary images, which can be used to assess the quality of image processing algorithms. Traditionally such assessment is done using the 'statistical' misclassification error rate, or Pratt's figure of merit. However, we show that these measures have practical weaknesses. Further we show that they have undesirable topological properties in the continuous case. The classical Hausdorff metric generates the desired topology (the myopic topology on compact subsets) but is highly sensitive to noise and thus unsuitable for image processing purposes.

We introduce a metric Δ which is an L^p modification of the Hausdorff metric. It generates the myopic topology yet has many features in common with L^p metrics, including better robustness against noise. The various error measures are compared on synthetic image data, including examples of edge detection and the Besag ICM algorithm.

0. Introduction

Most image processing tasks require us to find an algorithm which approximates or estimates a 'true' image as closely as possible. Examples are image reconstruction (de-blurring, tomographic reconstruction, registration), encoding (compression, discretisation, hierarchical representation) and low-level scene analysis (classification, segmentation and edge detection).

To objectively assess the performance of such algorithms, one needs a numerical measure $\Delta(f, g)$ of the discrepancy between two digital images f, g. In theoretical treatments Δ is typically a function space metric on the class of all possible images, and an optimal filter is one which achieves minimum (expected) error. In computer experiments Δ may be a more general "loss function".

Grey-scale images are traditionally compared using the root mean squared difference of corresponding pixel values,

$$||f-g||_2 = \left[\frac{1}{N}\sum_{x\in X}(f(x)-g(x))^2\right]^{1/2} \tag{1}$$

where $f(x)$ denotes the brightness value of image f at pixel x, and X is the raster of N pixels. This has many theoretical and computational advantages and is the basis of the usual optimal linear filtering theory [28, 49, 73]. Alternatives are the L^1 metric [22, 34] and Sobolev norms [38].

Error metrics are particularly important in the discussion of segmentation and classification [2, 8, 9, 32, 37, 48, 59, 60, 61, 65, 67] and edge detection in computer vision [1, 14, 25, 27, 29, 30, 45, 46, 64, 66]. Here the pixel values are binary, or unordered labels, and images are often compared using the *misclassification error rate*

$$\epsilon(f,g) = \frac{1}{N}\text{number}\{x \in X : f(x) \neq g(x)\}. \tag{2}$$

However, it is widely agreed that (1)-(2), and other similar L^p metrics, are inadequate to express the human observer's sense of reconstruction fidelity or quality [13, 26, 44, 48, 61, 58, 59]. They are insensitive to errors which severely affect 'shape' but involve relatively few pixels. See §3.2. In recent work on Bayesian methods of segmentation and classification [8, 9, 23, 24, 32, 37, 48, 47] it has been observed [9, discussion, p. 299], [48, pp. 97,110] that ϵ does not adequately detect degenerative effects such as over-smoothing in the ICM algorithm. See §6.4.

For binary images, a popular error measure is PRATT's [1, 46] *figure of merit*. We discuss this in §3.3 below. It too shows a number of undesirable features in practice.

A theoretically attractive, but practically unusable, alternative is the *Hausdorff metric* (e.g. [35, p. 15], [56, p. 72 ff.]), discussed in §2, §3.4 below. This is fundamentally important in mathematical morphology [20, 35, 56] and random set theory [31, 39, 63, 70]. The Hausdorff metric generates the myopic topology [35, p.12] on the class of compact subsets of X. SERRA [56, p. 72 ff.] argues that it is very desirable that image processing operations be continuous with respect to the myopic topology. However, the Hausdorff metric itself is extremely sensitive to 'noise' and even to changes in a single pixel. It cannot be used in practical experiments and its use as the basis of an optimal filtering theory is questionable.

In this paper we study the failings of the abovementioned metrics from a practical and theoretical viewpoint, and propose a new metric Δ combining their desirable features. It is topologically equivalent to the Hausdorff metric, satisfying the desiderata of [56, p. 72 ff.]; it is defined as an L^p mean of pixel contributions so that it has an 'average risk' interpretation and is reasonably stable to noise. The underlying theory [3] is also applicable to grey-level images, but here we discuss only the binary case; see also [4]. Related work is in [2, 65, 72].

Section 1 lists notation and assumptions. Section 2 introduces the desired topology, the myopic topology, and the associated Hausdorff metric. In section 3 we study error measures that are in current use, giving some examples of undesirable properties in practice, and some topological characterisations. Section 4 puts forward a list of precise conditions for choosing a suitable metric. Section 5 introduces the new metric Δ and examines its topological properties. Finally in section 6 we give some practical examples of application to digital images.

1. Setup

A discrete binary image is a function $f : X \to \{0, 1\}$ where X is the image raster. In applications, X is usually a finite subset of a two- or three-dimensional regular square or hexagonal lattice [46, 56]. The value 1 will be interpreted as logical 'true' and displayed as black. A binary image b can of course be identified with a subset $B \subseteq X$, namely the set of black pixels, $B = \{x \in X : b(x) = 1\}$.

Images defined on more general X, such as arbitrary finite graphs, have recently been studied. We follow Serra [56, 57] in allowing X to be a general topological space, including Euclidean space $\mathbb{R}^d$.

Definition 1. *Let (X, ρ) be a locally compact, second countable metric space. A* **binary image** *is a nonempty compact subset $B \subseteq X$. The class of all binary images is denoted by $\mathcal{K}'$. An* **image metric** *is a metric Δ on the space $\mathcal{K}'$.*

In the continuous case $X = \mathbb{R}^d$ the metric ρ will usually be Euclidean distance $\rho(x, y) = ||x - y||$. In the discrete case where X is a finite subset of a regular square or hexagonal lattice, ρ will be a shortest path length metric [11, 12, 50, 51].

Note that the formulation is not symmetric in 'black' and 'white', i.e. the set complement of an element of $\mathcal{K}'$ is not in $\mathcal{K}'$.

We also need a measure ν on X; in the discrete case this will be $\nu(B) = n(B) =$ number of points in B, while in the continuous case $X = \mathbb{R}^d$ typically ν will be Lebesgue measure.

Definition 2 (Assumptions). Assume that X is equipped with a Radon measure ν which is Borel regular and satisfies

$$\inf_{x \in X} \nu(D(x, r)) > 0 \tag{3}$$

for any fixed $r > 0$, where

$$D(x, r) = \{y \in X : \rho(x, y) \leq r\}$$

is the closed unit ball of radius $r > 0$ and centre $x \in X$.

2. Hausdorff metric and myopic topology

2.1. Hausdorff metric

Definition 3. Let (X, ρ) be a metric space as in Definition 1. The **distance function** of a nonempty subset $A \subseteq X$ is

$$d(x, A) = d_\rho(x, A) = \inf\{\rho(x, a) : a \in A\}, \quad x \in X$$

i.e. $d(x, A)$ is the shortest distance from pixel $x \in X$ to a pixel in A.

The distance function has a Lipschitz property

$$d_\rho(x, A) \leq d_\rho(y, A) + \rho(x, y) \tag{4}$$

for any $x, y \in X$. In particular, $d_\rho(\cdot, A)$ is uniformly continuous. Closed sets are characterised by their distance functions: $d_\rho(\cdot, A) \equiv d_\rho(\cdot, B)$ iff $A = B$.

It is interesting to note that, when X is a two- or three-dimensional rectangular or hexagonal grid and ρ is a shortest path metric, the function $d(\cdot, A)$ can be computed very rapidly by a recursive algorithm [12, 50, 51] based on repeated application of (4).

Definition 4. *The* **Hausdorff distance** *(e.g. [35, p. 15], [56, p. 72 ff.]) between two nonempty subsets $A, B \subseteq X$ is*

$$H_\rho(A, B) = max\left\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\right\} \tag{5}$$

i.e. this is the maximum distance from a point in one set to the nearest point in the other set.

Important representations of H_ρ are the following (see [19, 2.10.21]).

Lemma 1. For any ρ and for nonempty compact $A, B \subseteq X$,

$$H_\rho(A, B) = \sup_{x \in X} |d_\rho(x, A) - d_\rho(x, B)| \tag{6}$$

This is trivial using (4). It is an immediate consequence that H_ρ is a finite-valued metric on $\mathcal{K}'$.

Lemma 2.

$$H_\rho(A, B) = \inf\left\{r \geq 0 : A \subseteq B^{(r)},\ B \subseteq A^{(r)}\right\} \tag{7}$$

where

$$A^{(r)} = \{x \in X : d_\rho(x, A) \leq r\} \tag{8}$$

is called the r-envelope *of* A.

In the case $X = \mathbb{R}^d$ with $\rho(x,y) = ||x-y||$, the envelope is familiar as the Minkowski dilation $A \oplus rD$ of mathematical morphology [56, p. 43] where rD is the ball of radius r centred at the origin in $\mathbb{R}^d$. In case X is a discrete rectangular or hexagonal grid with a translation-invariant metric (cf. [12, 50, 51]) again $A^{(r)} = A \oplus D_r$ where D_r is the ball of radius r in the given metric.

Interpretation (7) connects Hausdorff distance with the partial order of set inclusion.

2.2. Myopic topology

Topologies on spaces of subsets were introduced by Michael [36], Fell [20] and Matheron [35]. See also [6, 7, 15, 21, 33, 71]. Related topologies have been constructed for uppersemicontinuous functions [52, 68], Radon measures [10, 31], and capacities in general [39, 40, 41, 42, 43, 69, 70]. The present section collects definitions and important facts from the above and [3].

DEFINITION 5. The **myopic topology** *on* $\mathcal{K}'$ [36], [35, p. 12] *is the weakest topology on* $\mathcal{K}'$ *generated by all classes*

$$\{K \in \mathcal{K}' : K \cap F = \emptyset\} \quad \textit{for all} \quad F \in \mathcal{F}$$
$$\{K \in \mathcal{K}' : K \cap G \neq \emptyset\} \quad \textit{for all} \quad G \in \mathcal{G}$$

where $\mathcal{F}, \mathcal{G}$ *are the classes of all closed and open subsets of* X*, respectively.*

The myopic topology is locally compact, second countable and Hausdorff [35, p. 13] so that $K_n \to K$ myopically in $\mathcal{K}'$ iff

$$K \cap F = \emptyset \;\Rightarrow\; K_n \cap F = \emptyset \text{ eventually}$$
$$K \cap G \neq \emptyset \;\Rightarrow\; K_n \cap G \neq \emptyset \text{ eventually}$$

for all $F \in \mathcal{F}, G \in \mathcal{G}$, where 'eventually' means 'for all but finitely many n'.

The discussion in [35, pp. ix–x, 12-26] and [56, p. 72 ff.] makes it clear that continuity with respect to the myopic topology is a very desirable property for image processing algorithms. The connection with the Hausdorff metric is the following.

PROPOSITION 1. *Let* (X,ρ) *be as in Definition 1. For* $K, K_n \in \mathcal{K}'$*, the following are equivalent:*

(a) $K_n \to K$ *myopically;*

(b) $d(x,K_n) \to d(x,K)$ *uniformly in* $x \in X$;

(c) $H_\rho(K_n, K) \to 0$.

The equivalence of (b),(c) is clear from (6) equivalence of (a), (c) is proved by MATHERON [35, p. 15]. See also [71, theorem 3.1], [6, 7], [15, p.41], [33, p. 41].

The Hausdorff metric thus provides the foundation for mathematical morphology and random set theory [35, pp 12–26], [56, pp. 63-92].

3. Error measures in current use

In this section we study the error measures for binary images surveyed by Peli & Malah [45] and Van Vliet et al. [66].

3.1. General criteria

Important general principles for error measurement were enunciated by Canny [14] in the context of edge detection. He argued that an edge filter should be considered 'good' if it exhibits

1. good detection: low probability of failing to detect an edge, and low probability of incorrectly labelling a background pixel as an edge;
2. good localization: points identified as edge pixels should be as close as possible to the centre of the true edge;
3. unicity: there should be only one response to a single edge.

Canny showed that (in a suitable real-analytic framework) there is an uncertainty principle balancing good detection against good localization.

3.2. Detection performance ('statistical') measures

Let A be the 'true' binary image and B the putative or 'estimated' image. Pixels that belong to B but not A will be called false positives or Type I errors; pixels that belong to A but not B will be called false negatives or Type II errors. In case X is finite, define the type I error rate [45] by

$$\alpha(A,B) = \frac{n(B \setminus A)}{n(X \setminus A)} \tag{9}$$

the type II error rate

$$\beta(A,B) = \frac{n(A \setminus B)}{n(A)} \tag{10}$$

and the overall misclassification rate

$$\epsilon(A,B) = \frac{n(A \triangle B)}{n(X)} \tag{11}$$

where $n(S)$ = number of pixels in S and $\triangle$ denotes set symmetric difference. Some other quantities derived from α, β are discussed in [17, 45].

Two attractive features of ϵ are that it is a metric on $\mathcal{K}'$, and that it is linear. Under any stochastic model for A and B the mean value of $\epsilon(A,B)$ equals the average over all pixels x of the disagreement probability:

$$\mathbb{E}\,\epsilon(A,B) = \frac{1}{n(X)} \sum_{x \in X} \mathbb{P}\{a(x) \neq b(x)\}. \tag{12}$$

where $a(x) = 1$ if $x \in A$, and 0 otherwise. Thus ϵ-optimality has an 'average risk' interpretation: the estimate $\widehat{a}$ of an image a which minimises expected ϵ-error is that which maximises the pixelwise likelihood $\mathbb{P}\{\widehat{a}(x) = a(x)\}$ for every pixel x. There is a similar Bayesian statement.

It is widely acknowledged [9, p.299], [48, pages 97,110], [58, 59] that pixel misclassification errors are a poor measure of reconstruction fidelity. Discrepancies between A and B are measured by the number of disagreements, regardless of the pattern. Errors such as the displacement of a boundary, that affect a large number of pixels but do not severely affect 'shape', are given high values by ϵ; while errors such as the deletion of a linear feature, filling-in of small holes or deletion of small islands, that involve only a small number of pixels but severely affect 'shape', have low ϵ values. An example of an effect which is not well detected by ϵ, α, β is the over-smoothing of segmented images by iterative algorithms such as ICM and deterministic and stochastic relaxation [9, discussion], [23, 48, 47]. See §6.4 for an example. These comments support Canny's observations.

More rigorous statements of this kind can be obtained by considering topological properties of the continuous counterparts of α, β, ϵ. For arbitrary (X, ρ) with measure ν as in Definition 2 define

$$\begin{aligned} \tilde{\alpha}(A,B) &= \nu(B \setminus A) \\ \tilde{\beta}(A,B) &= \nu(A \setminus B) \\ \tilde{\epsilon}(A,B) &= \nu(A \triangle B). \end{aligned}$$

One cannot normalise α and ϵ as done in (9), (11) unless $\nu(X) < \infty$.

LEMMA 3. *For $A_n, A \in \mathcal{K}'$, if $A_n \to A$ myopically then $\tilde{\alpha}(A, A_n) \to 0$. The converse is generally false, and so is the corresponding statement concerning $\tilde{\alpha}(A_n, A)$. The analogous statements for $\tilde{\beta}$ and $\tilde{\epsilon}$ are also generally false.*

PROOF. If $H(A_n, A) \to 0$ then by (7) for every $r > 0$ we have for all sufficiently large n that $A_n \subseteq A^{(r)}$, so that $\nu(A_n \setminus A) \leq \nu(A^{(r)} \setminus A)$. Now $A^{(r)} \downarrow A$ as $r \downarrow 0$ so $\lim_{r \downarrow 0} \nu(A^{(r)}) = \nu(A)$. Thus $\tilde{\alpha}(A, A_n) \to 0$.

The other statements mentioned are generally false, since e.g. the finite sets are dense in $\mathcal{K}'$ with the myopic topology and have zero measure under any nonatomic ν. □

The following result states that $\tilde{\epsilon}$ convergence and myopic convergence are equivalent for images without 'small' features. Define the **inner envelope** of $A \subset X$ as

$$A^{(-r)} = \{x \in A : d(x, A^c) \geq r\}$$

for $r > 0$. Say that A is **r-closed** if

$$\left(A^{(r)}\right)^{(-r)} = A$$

and **r-open** if

$$\left(A^{(-r)}\right)^{(r)} = A.$$

These are analogues of standard definitions in mathematical morphology [56, 57]. If $X = \mathbb{R}^d$ and ρ is Euclidean distance, r-closure implies s-closure for every $0 < s \leq r$ and similarly for openings. Every compact *convex* set in $\mathbb{R}^d$ is r-closed for every $r > 0$.

LEMMA 4. *Let $r > 0$ be fixed and let $\mathcal{C}_r, \mathcal{O}_r$ be the classes of compact subsets of X that are respectively s-closed and s-open for every $0 < s \leq r$.*

(a) *on $\mathcal{C}_r$, $A_n \to A$ myopically implies $\tilde{\epsilon}(A, A_n) \to 0$.*

(b) *on $\mathcal{O}_r$, $\tilde{\epsilon}(A, A_n) \to 0$ implies $A_n \to A$ myopically.*

Thus $\tilde{\epsilon}$ and H are topologically equivalent on $\mathcal{C}_r \cap \mathcal{O}_r$.

PROOF. (a): Suppose $H(A_n, A) \to 0$ for $A_n, A \in \mathcal{C}_r$. For $s \leq r$ we have for all sufficiently large n by (7) that $A \subseteq (A_n)^{(s)}$ and $A_n \subseteq A^{(s)}$. Since the inner envelope operation is increasing ($A \subseteq B \Rightarrow A^{(-r)} \subseteq B^{(-r)}$) we have

$$\begin{aligned} A^{(-s)} &\subseteq \left[(A_n)^{(s)}\right]^{(-s)} \\ &= A_n \end{aligned}$$

since A_n is s-closed. That is

$$A^{(-s)} \subseteq A_n \subseteq A^{(s)}$$

for all sufficiently large n. Thus $\nu(A_n \triangle A) \leq \nu(A^{(s)} \setminus A^{(-s)})$. Now as $s \downarrow 0$ we have $A_{(s)} \downarrow A$ and $A_{(-s)} \uparrow \mathbf{int}(A)$, the topological interior. Since ν is Borel regular the measures of both these sequences converge to $\nu(A)$. Thus $\tilde{\epsilon}(A, A_n) \to 0$.

(b): Given $0 < s \leq r$ suppose $A, B \in \mathcal{O}_r$ are such that $H(A, B) \geq s$. Then without loss of generality there is a point $x \in A$ with $d(x, B) \geq s$. Set $t = s/3$; then since A is t-open, x belongs to some ball $D(y, t)$ of radius t in the metric ρ contained entirely within A. But since $d(x, B) \geq s$ we have $D(y, t) \cap B = \emptyset$ so that $\nu(A \triangle B) \geq \nu(D(y, t))$. The latter is bounded below in y by assumption (Definition 2). Hence

$$\inf\{\tilde{\epsilon}(A, B) : H(A, B) \geq s, \quad A, B \in \mathcal{O}_r\} > 0.$$

It follows that $\tilde{\epsilon}(A_n, A) \to 0$ implies $H(A_n, A) \to 0$. □

Part (a) is similar to the proof [35, p. 68] that Lebesgue measure is myopically continuous on the compact convex sets of $\mathbb{R}^d$.

3.3. Localization performance ('distance') measures

Measures of localization performance discussed by PELI and MALAH [45] for a discrete raster X were the mean error distance

$$\bar{e}(A,B) = \frac{1}{n(B)} \sum_{x \in B} d(x,A), \tag{13}$$

the mean square error distance

$$\overline{e^2}(A,B) = \frac{1}{n(B)} \sum_{x \in B} d(x,A)^2 \tag{14}$$

and Pratt's [1, 46] "figure of merit"

$$\mathrm{FOM}\,(A,B) = \frac{1}{\max\{n(A),n(B)\}} \sum_{x \in B} \frac{1}{1+\kappa\, d(x,A)^2} \tag{15}$$

where κ is a scaling constant, usually set to $1/9$ when ρ is normalized so that the smallest nonzero distance between pixel neighbours equals 1. Here A is the true image and B the estimated image; note that $\mathrm{FOM}\,(A,B) \neq \mathrm{FOM}\,(B,A)$. One has $0 < \mathrm{FOM}\,(A,B) \leq 1$ and $\mathrm{FOM}\,(A,B) = 1$ iff $A = B$. FOM is the most popular of these and is widely used [1, 5, 25, 45, 46].

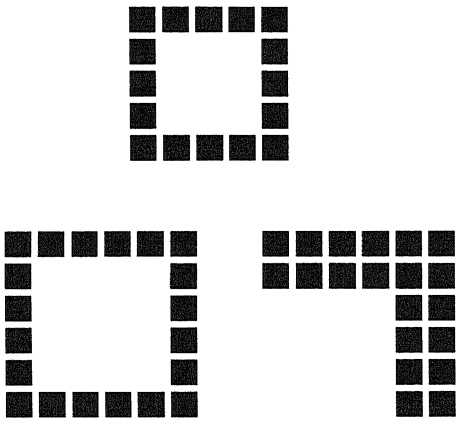

FIGURE 1. Peli-Malah counterexample. True picture A (top) and two error pictures B_1, B_2 (bottom) with the same value of FOM.

The three error measures are insensitive to type II errors. For example, if all errors are of type II, $B \subseteq A$, then $\bar{e} = \overline{e^2} = 0$ while $\mathrm{FOM}\,(A,B) = n(B)/n(A) = 1 - \beta(A,B)$. They are also insensitive to the pattern of type I errors, since they are *averages* of a loss function $f(x,A)$ over all type I error pixels x. A striking example found by PELI and MALAH [45] is shown in Figure 1. If the upper image is taken as the true image A, then the two lower images B_1, B_2 have the same FOM values, $\mathrm{FOM}\,(A,B_1) = \mathrm{FOM}\,(A,B_2)$. Indeed they also have the same values of α, β and ϵ.

PELI and MALAH [45] and VAN VLIET et al. [66] observed cases where

FOM was large, but the visual quality was bad. When FOM was used as a criterion for choosing parameter values in edge detection algorithms [66, p. 186 and section 6] in the case of the classical Laplacian operator the FOM-optimal images often had sections of the true contour missing, or oscillated around the true contour. This behaviour can be explained by observing that for $x \notin A \cup B$ we have $\mathrm{FOM}(A, B \cup \{x\}) < \mathrm{FOM}(A, B)$ if and only if $d(x, A) > \kappa^{-1/2} \left(\mathrm{FOM}(A, B)^{-1} - 1\right)^{1/2}$. If $\mathrm{FOM}(A, B) > 0.9$ and $\kappa = 1/9$ then this translates into

$$\mathrm{FOM}(A, B \cup \{x\}) < \mathrm{FOM}(A, B) \quad \text{iff} \quad d(x, A) \geq 1$$

That is, when FOM is large, preference will always be given to type II errors over type I errors, however innocuous the latter might be.

Similar criticisms apply to $\bar{e}$ and $\overline{e^2}$; these have the additional disadvantage that they are highly sensitive to background noise. If the error image B contains even one single pixel x far distant from A, its distance value will drastically elevate the mean distance. This is connected with the statistical phenomenon of non-robustness of the arithmetic mean.

Note $\bar{e}$ and $\overline{e^2}$ have an average risk interpretation analogous to (12), but FOM does not. Splitting the sum in (15) one has

$$\mathrm{FOM}(A, B) = \frac{1}{\max\{n(A), n(B)\}} \left[n(A \cap B) + \sum_{x \in B \setminus A} \frac{1}{1 + \kappa d(x, A)^2} \right]$$

or

$$1 - \mathrm{FOM}(A, B) = \frac{1}{\max\{n(A), n(B)\}} \sum_{x \in B \setminus A} \frac{d(x,A)^2}{1 + \kappa d(x,A)^2} + \max\{0, 1 - \frac{n(B)}{n(A)}\}.$$

The three error measures seem difficult to interpret because of the normalisation by a variable denominator $n(B)$ or $\max\{n(A), n(B)\}$. For example it is not clear how to compare $\mathrm{FOM}(A, B)$ for fixed A and different B if $n(B) > n(A)$.

Peli and Malah concluded that FOM sometimes gives insufficient information and that a better measure is needed.

The author is not aware of any theoretical justification for $\bar{e}, \overline{e^2}$ or FOM. For general X, define analogues of $\bar{e}$, $\overline{e^2}$ and FOM by replacing $n(\cdot)$ by ν, and sums by integrals with respect to ν, in (13)-(15). These apply only when $\nu(A), \nu(B) > 0$. The (Lebesgue) integrals are always well-defined since the integrand is uniformly continuous by (4) and the domain of integration B is compact.

LEMMA 5. *Suppose $A_n \to A$ myopically where $A_n, A \in \mathcal{K}'$ satisfy $\nu(A) > 0$, $\nu(A_n) > 0$. Then $\bar{e}(A, A_n) \to 0$, and $\overline{e^2}(A, A_n) \to 0$. The converse statements are generally false. The corresponding statements for FOM are also generally false.*

LEMMA 6. *Let C_r be as in Lemma 4. Then on C_r, $A_n \to A$ myopically implies $\mathrm{FOM}(A, A_n) \to 0$, but not conversely.*

The proofs are similar to those for Lemmas 3 and 4a respectively.

3.4. Hausdorff metric

The Hausdorff metric H defined in §2 is never used in practice (to the author's knowledge) as an error measure for digital images. It has been used in some research on numerical analysis and approximation theory [53, 54, 55, 16].

The practical objection to H is its extreme sensitivity to changes in even a small number of pixels, because of the supremum in (4). This is expressed by the following 'minimax property'

LEMMA 7. *For any sets* $A_1, \ldots, A_n$ *and* $B_1, \ldots, B_m$

$$H\left(\bigcup_{i=1}^{n} A_i, \bigcup_{j=1}^{m} B_j\right) = \max\left\{\max_i \min_j H(A_i, B_j), \max_j \min_i H(A_i, B_j)\right\} \quad (16)$$

This is an application of the definition of H to the relation

$$d(t, \bigcup_{i=1}^{n} A_i) = \min_i d(t, A_i).$$

Hence for example

$$H(A, A \cup \{x\}) = d(x, A)$$

is unbounded. Errors such as the addition of a small amount of background noise thus give inappropriately high values of H. The Hausdorff metric is therefore so unstable as to be unusable in this context.

4. DISCUSSION OF PROBLEM

The objections to existing error measures, discussed above, may be summarised as follows. Firstly, in some cases, errors of a particular kind are completely undetected; for example, type II errors are not detected by $\bar{e}, \overline{e^2}$. Such error measures do not have a satisfactory topological interpretation at all. Secondly, in most cases, the relative values given to various possible types of errors are not in the desired relation to each other (cf. discussion in §3.2-3.3). In the cases considered, this imbalance was extreme, in that the continuous-space versions of these error measures are in conflict with the myopic topology. Thirdly, in some cases $(\bar{e}, \overline{e^2}, H)$ the error measure is not 'robust' in that the alteration of a relatively small number of pixel values can have an unbounded effect on the measured error.

We thus arrive at the following list of desiderata for an improved error measure:

1. The error measure should be a metric Δ on $\mathcal{K}'$;
2. Δ should generate the myopic topology;
3. There should be an 'average risk' interpretation of Δ similar to (12);

4. The alteration of a fixed number of pixels should have a bounded effect on Δ, or more generally,

$$\sup\{\Delta(A, A \cup B) \ : \ \nu(B) \leq u\} < \infty$$

for each fixed $A \in \mathcal{K}'$ and $u < \infty$.

See [56, p.72ff] for a discussion of metrics in the context of image processing. While a metric is desirable for theoretical purposes, in particular for developing an optimal filtering theory (e.g. [28, 49, 73]) it is arguable whether the metric property is desirable for practical purposes. The symmetry axiom implies equal treatment of type I and type II errors. The triangle inequality effectively means we cannot normalise the error $\Delta(A, B)$ by some measure of the size of A or B (as done in the construction of FOM, α, β but not ϵ). Yet these objections would be unimportant if one could find a metric that behaved well in practical experiments.

It is also important to distinguish topologies and *uniformities* [18, pp. 200–204] which are not discussed in [56]. Two metrics may generate the same topology, yet not generate the same uniformity. The complaints which we have levelled at existing error metrics can best be understood as differences in the corresponding uniformities in the continuous case. One can change a metric so as to preserve the desired topology but change the undesired uniformity.

A number of authors [62, 65, 72] have proposed modifications of ϵ of the form

$$\zeta(A, B) = \frac{1}{N} \sum_{x \in X} z(x; A, B)$$

where the contribution $z(x; A, B)$ from pixel x depends on pixel values of A, B in a neighbourhood of x. These enjoy an 'average risk' interpretation similar to (12) for ϵ, but are also liable to similar criticisms.

Desideratum 4 above can be achieved by the standard device of concave transformation.

LEMMA 8. *Let w be any continuous function on $[0, \infty]$ that is* **concave**

$$w(s + t) \leq w(s) + w(t)$$

and strictly increasing at 0, $w(t) = 0$ iff $t = 0$. If ρ is a metric on a space S, then $\tau = w \circ \rho$ is also a metric, i.e. $\tau(x, y) = w(\rho(x, y))$. The metrics τ and ρ generate the same topology and the same uniformity.

Examples which transform an unbounded metric to a bounded one include

$$\begin{aligned} w(t) &= \frac{t}{1+t} \\ w(t) &= \tan^{-1}(t) \\ w(t) &= \min\{t, c\} \end{aligned}$$

for a fixed $c > 0$.

The effect of concave transformation on the Hausdorff metric H is particularly simple.

LEMMA 9. *Let w be a concave function as above. If H_ρ denotes the Hausdorff metric defined by a pixel distance metric ρ, then $H_{w\circ\rho} = w \circ H_\rho$, i.e.*

$$H_{w\circ\rho}(A, B) = w(H_\rho(A, B)). \tag{20}$$

Hence $H_{w\circ\rho}$ generates the same topology and uniformity as H_ρ.

5. PROPOSED METRIC

The proposal is simply to replace the supremum in representation (6) of H by an L^p average. Thus in the discrete case

$$\Delta^p(A, B) = \left[\frac{1}{n(X)} \sum_{x\in X} |d(x, A) - d(x, B)|^p\right]^{1/p} \tag{21}$$

for $1 \leq p < \infty$. The result is still a metric, and indeed topologically equivalent to H, since distance functions $d(x, A)$ are equicontinuous Lipschitz functions by (4). More generally we can introduce a concave transformation as follows.

DEFINITION 6 (DELTA METRIC). Let X, ρ, ν be as in Definition 2. Let $w : [0, \infty] \to [0, W]$ be any bounded concave function with $w(0) = 0$. For $1 \leq p < \infty$ define

$$\Delta^p_w(A, B) = \left[\int_X |w(d(x, A)) - w(d(x, B))|^p \, d\nu(x)\right]^{1/p} \tag{22}$$

for $A, B \subseteq \mathcal{K}'$.

PROPOSITION 2. *Assume either*

- X is compact (so that $\nu(X) < \infty$);
- w is eventually constant, $w(t) = W$ for $t \geq t_0$ where $t_0 < \infty$.

Then Δ^p_w is a finite valued metric on $\mathcal{K}'$ generating the myopic topology.

PROOF. For any $K_1, K_2 \in \mathcal{K}'$ the set of points x where $\max d(x, K_i) \leq t_0$ is compact. Thus the integrand of (22) is zero outside a compact region, and since ν is a Radon measure the integral in (22) is finite. It is then clear from standard properties of $L^p(\nu)$ that Δ^p_w is a metric.

Let $K_n, K \in \mathcal{K}'$ and suppose $H_\rho(K_n, K) \to 0$. Then $d_\rho(\cdot, K_n) \to d_\rho(\cdot, K)$ uniformly. Hence $w \circ d_\rho(\cdot, K_n) \to w \circ d_\rho(\cdot, K_n)$ uniformly, which (since the integrals are bounded) implies convergence in L^p, i.e. $\Delta^p_w(K_n, K) \to 0$.

Conversely suppose $K_n \to K$ in Δ^p_w, i.e. $d_\sigma(\cdot, K_n) \to d_\sigma(\cdot, K)$ in L^p, where $\sigma = w \circ \rho$. Then writing

$$R_{n,\epsilon} = \{x \in X : |d_\sigma(x, K_n) - d_\sigma(x, K)| > \epsilon\}$$

for $\epsilon > 0$, we have $\nu(R_{n,\epsilon}) \to 0$ as $n \to \infty$ for each fixed ϵ. Now the Lipschitz property (4) gives, for any $x, y \in X$,

$$|d_\sigma(x, K_n) - d_\sigma(x, K)| \leq |d_\sigma(y, K_n) - d_\sigma(y, K)| + 2\sigma(x, y)$$

so that

$$x \in R_{n,\epsilon} \Rightarrow D_\sigma(x, \epsilon/3) \subseteq R_{n,\epsilon/3}.$$

For fixed $\epsilon > 0$, the ν-measure of the right-hand set converges to zero; yet since closed balls in σ are closed balls in ρ, (3) implies $\nu(D_\sigma(x, \epsilon/3))$ is bounded below for all n. Thus $R_{n,\epsilon}$ must be empty for sufficiently large n, i.e. the σ distance functions converge uniformly on X. Hence the ρ distance functions converge (pointwise, hence uniformly). □

Implementation is straightforward using the distance transform algorithm of ROSENFELD and BORGEFORS [12, 50, 51]. In applications we shall always use the cutoff transformation $w(t) = \min\{t, c\}$ for a fixed $c > 0$. In this case the contributions to the integral in (22) are zero for points x further than c units away from A and B. This has the attractive property that the value of $\Delta^p(A, B)$ does not change if we change the grid size (embed X in a larger space). The possible values of $\Delta^p(A, B)$ then range from 0 to c.

The parameters c and p determine the tradeoff between localization error and misclassification error. The value of c effectively controls scale: roughly speaking, a misclassification error is equivalent to an error in localization by distance c. For small c the effect is similar to misclassification error; as $c \to 0$ on a discrete grid $\frac{1}{c}\Delta^p_w(A, B) \to \epsilon(A, B)^{1/p}$. The value of p determines the relative importance of large localization errors. For large p the effect is similar to the Hausdorff metric; $\Delta^\infty_w(A, B) = H_w(A, B)$.

Since $d(x, A) = 0$ for $x \in A$, the sum in (21) includes contributions $\sum_{x \in A} d(x, B)^p$ and $\sum_{x \in B} d(x, A)^p$ which are analogous to $\bar{e}, \overline{e^2}$ and FOM. However, the sum in (21) also includes other terms for x outside A and B.

Clearly Δ^p_w has an 'expected risk' interpretation analogous to (12): under any stochastic model for A, B

$$\mathbb{E}[(\Delta^p_w(A, B))^p] = \int_X \mathbb{E}|w(d(x, A)) - w(d(x, B))|^p \, d\nu(x). \tag{23}$$

Another version is as follows. Let B be a random compact set [35] and $A \in \mathcal{K}'$ fixed. Then by straightforward calculation

$$\mathbb{E}[(\Delta^p_w(A, B))^p] = \int_X q_B(x, d_\tau(x, A)) d\nu(x) \tag{24}$$

where $\tau = w \circ \rho$ and

$$q_B(x, t) = p \int_t^\infty (s - t)^{p-1} T_B(D_\tau(x, s)) ds + t^p - p \int_0^t (t - s)^{p-1} T_B(D_\tau(x, s)) ds$$

depends only on $d_\sigma(\cdot, A)$ and on the avoidance functional [35] of B,

$$T_B(K) = \mathbb{P}\{B \cap K = \emptyset\}.$$

6. Examples

Throughout this section we have compared the figure of merit FOM for $\kappa = 1/9$ with the Δ metric with $p = 2$ and w being the cutoff transformation (19).

6.1. Peli-Malah example

This example (Figure 1) yields a FOM value of 0.941 for both pictures B_1, B_2. The corresponding Δ^2 values, with cutoff $c = 5$, are 0.323 (left picture) and 0.512 (right picture).

6.2. Artificial data

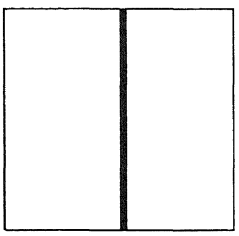

FIGURE 2. Synthetic true image A

Figures 2–4 show synthetic images (32×32 pixels) deviating in various ways from a straight edge. Table 1 reports the computed values of type I error α, type II error β, misclassification error ϵ, Pratt's figure of merit with $\kappa = 1/9$, and the Δ^2 metric with cutoff distance 5.

The most dramatic disagreement between these measures is for the `gaps` image, which scores a very bad grade in FOM, an indifferent grade in β, and scores better than all other images in Δ^2. FOM gives roughly comparable, high scores to `shift`, `bend` and `barbs`, while ϵ and Δ^2 spread them over a wide range.

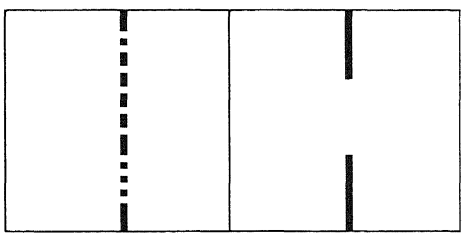

FIGURE 3. Images `gaps` and `lost`

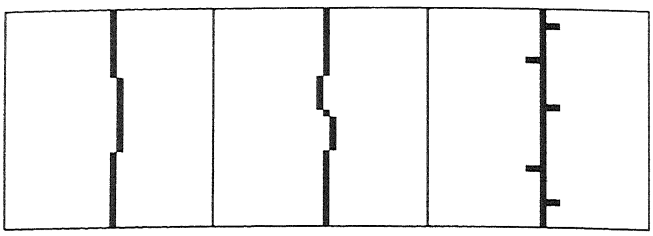

FIGURE 4. Images `shift`, `bend` and `barbs`

Image **B**	α	β	ϵ	FOM	Δ^2
`gaps`	0	0.313	0.0098	0.688	0.149
`lost`	0	0.344	0.0108	0.656	0.682
`shift`	0.011	0.344	0.0214	0.966	0.319
`bend`	0.010	0.313	0.0195	0.969	0.291
`barbs`	0.010	0	0.0097	0.952	0.463

TABLE 1. Error measures for the synthetic images

6.3. Edge detection

The next experiment is modelled on the standard edge detector test of HARALICK [29] (see [27, 30, 66]) and compares optimality under FOM and under Δ^2.

Figure 5 shows the test image, a chessboard pattern with additive Gaussian noise at signal-to-noise ratio 2.0. Figure 6 shows the true edge image, computed before adding the noise, and cropped from 256×256 pixels to 200×200 to standardise the image size for comparisons with filtered images.

The edge detector consisted of Gaussian smoothing with standard deviation σ, followed by the classical 4-connected Laplacian, zero-crossing by thresholding and distance transform [66], then the Lee-Haralick morphological edge strength detector with a pseudocircular mask [66] of size 3 was applied to the smoothed data and the result multiplied by the zero-crossing image. The resulting edge strength image was thresholded at a fixed value to obtain a binary edge image B. For various values of the smoothing parameter σ, we then compared FOM (A, B) with $\Delta^2(A, B)$. The FOM parameter κ was set to the usual $1/9$ and the cutoff parameter of Δ^2 was $c = 5$.

The plot of FOM and Δ^2 values (Figure 7) shows that FOM is almost indifferent to a wide range of σ values near its optimum. This is consistent with our theoretical comments about FOM.

The optimal values of σ under the two measures were quite different: FOM chooses $\sigma = 2.32$ and Δ^2 chooses $\sigma = 2.08$. The corresponding images are shown in Figure 8. The FOM values were 0.941 and 0.939 respectively; the corresponding Δ^2 values were 0.663 and 0.617.

6.4. ICM algorithm

Our final experiment studies the behaviour of Besag's ICM (Iterated Conditional

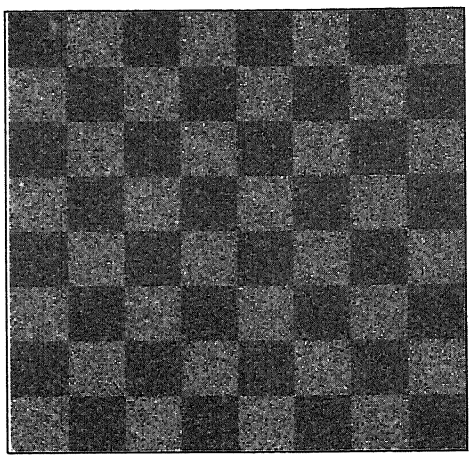

FIGURE 5. Chessboard image with additive Gaussian noise (SNR = 2)

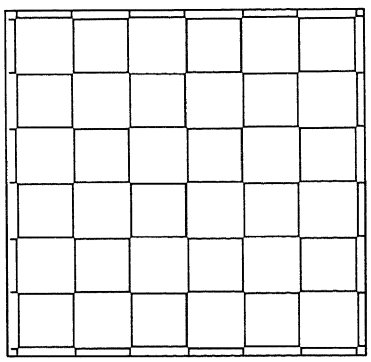

FIGURE 6. True edges of chessboard

Modes) algorithm for classification [9]. The 'true' image is the 200×125 binary image in Figure 9 (originally obtained by thresholding a camera image of a newspaper).

The data image in Figure 10 was generated using independent Gaussian pixel values, with mean 100, standard deviation 40 for foreground (text) pixels and mean 150, standard deviation 20 for background pixels. Heteroscedastic data (i.e. with unequal variances) were used to encourage an imbalance of type II over type I errors, the most challenging situation for FOM.

Figure 11 shows the result of maximum likelihood classification from the data of Figure 10 (i.e. applied independently to each pixel).

We then applied Besag's ICM algorithm [9] (eight-connected neighbourhood, 2×2 coding, constant interaction β for all neighbours) for 11 iterations, with interaction parameter β beginning at 0.4 and increasing in steps of 0.2 until the

value 2.0 was reached.

Figure 12 shows a selection of the ICM results. Features to note are the dramatic improvement over maximum likelihood, after one iteration, and the slow degeneration of the text shapes beyond iteration 4.

Figure 13 shows the error measures Δ^2, $1 - \text{FOM}$ and ϵ for successive ICM iterations. The cutoff for Δ^2 was $c = 4$. To remove irrelevant scale effects Figure 14 shows the same data normalised to have value 1 at iteration 11.

The initial improvement from MLE to iteration 1 is registered clearly by all three error measures. The degeneration of shape from about iteration 4 onwards is almost undetected by FOM and the misclassification error, but quite clearly detected by Δ^2.

ACKNOWLEDGEMENTS

For helpful comments the author is grateful to G. Beer, M. Berman, J.E. Besag, L. Bischof, T.C. Brown, A.J. Cabo, N.I. Fisher, J. Freedman, R.D. Gill, H.J.A.M. Heijmans, M.N.M. van Lieshout, P. Nacken, D.B. Pollard, B.D. Ripley, D.E. Robinson, A.N. Shiryayev, A.G. Steenbeek, C.C. Taylor, T.R. Turner and W. Vervaat, and participants in the Workshop on Statistics and Pattern Recognition, Edinburgh, 1985, 'StatComp', Melbourne 1987, and ANZAAS Congress, Townsville 1987.

REFERENCES

1. I.E. ABDOU and W.K. PRATT, 1979, Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67: pp. 753-763.
2. A.J. BADDELEY, 1987, A class of image metrics. In *Australian and New Zealand Association for the Advancement of Science, Proceedings 11th Congress, Townsville.*
3. A.J. BADDELEY, 1991, Hausdorff metric for capacities, *Research report BS-R9127*, Centrum voor Wiskunde en Informatica, Amsterdam.
4. A.J. BADDELEY, 1992, An error metric for binary images. In W. FÖRSTNER and S. RUWIEDEL, editors, *Robust Computer Vision*, pp. 59-78, Karlsruhe, Wichmann.
5. D.G. BAILEY and R.M. HODGSON 1985, Range filters: intensity subrange filters and their properties. *Image Vision Computing*, 3: pp. 99-110.
6. G. BEER, 1985, On convergence of closed sets in a metric space and distance functions. *Bulletin of the Australian Mathematical Society*, 31: pp. 421-432.
7. G. BEER, 1985, Metric spaces with nice closed balls and distance functions for closed sets. *Bulletin of the Australian Mathematical Society*, 35:81-96.
8. J. BESAG, 1983, Discussion of paper by P. Switzer. *Bulletin of the International Statistical Institute*, 50: pp. 422-425.
9. J. BESAG, 1986, On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, series B*, 48: pp. 259-302.
10. P. BILLINGSLEY, 1968, *Convergence of probability measures.* John Wiley and Sons, New York.

11. G.BORGEFORS, 1984, Distance transformations in arbitrary dimensions. *Computer Vision, Graphics and Image Processing*, 27: pp. 321-345.
12. G. BORGEFORS, 1986, Distance transformations in digital images. *Computer Vision, Graphics and Image Processing*, 34: pp. 344-371.
13. T.M. CANNON, H.J. TRUSSELL, and B.R. HUNT, 1978, Comparison of image restoration methods. *Applied Optics (Opt. Soc. Amer.)*, 17: pp. 3384-3390.
14. J. CANNY, 1986, A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8: pp. 679-698.
15. C.CASTAING and M. VALADIER, 1977, *Convex analysis and measurable multifunctions.* Lecture Notes in Mathematics 580. Springer.
16. P.J. DAVIS, R.A. VITALE, and E.BEN-SABAR, 1977, On the deterministic and stochastic approximation of regions. *Journal of Approximation Theory*, 21: pp. 60-88.
17. E.S. DEUTSCH and J.R. FRAM, 1978, A quantitative study of the orientation bias of some edge detector schemes. *IEEE Transactions on Computing*, 27: pp. 205-213.
18. J.DUGUNDJI, 1966, *Topology.* Allyn and Bacon, Boston, Mass..
19. H. FEDERER, 1969, *Geometric Measure Theory.* Springer Verlag, Heidelberg.
20. J.M.G. FELL, 1962, A Hausdorff topology for the closed subsets of a locally compact non-Hausdorff space. *Proc. Amer. Math. Soc.*, 13: pp. 472-476.
21. Z. FROLÍK, 1960, Concerning topological convergence of sets. *Czechoslovak Mathematical Journal*, 10: pp. 168-180.
22. M. GABBOUJ and E.J. COYLE, 1990, Minimum mean absolute error stack filtering with structural constraints and goals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38: pp. 955-968.
23. S. GEMAN and D. GEMAN, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: pp. 721-741.
24. S. GEMAN and D.E. MCCLURE, *Bayesian image analysis: application to single photon emission tomography.* Preprint, Applied Mathematics Dept., Brown University.
25. J.J. GERBRANDS, E. BACKER and W.A.G. VAN DER HOEVEN, 1986, Quantitative evaluation of edge detection by dynamic programming. In GELSEMA and KANAL, editors, *Pattern Recognition in Practice II*, pp. 91-99. Elsevier, North-Holland, New York.
26. F. GODTLIEBSEN, 1991, Noise reduction using Markov random fields. *J. Magnetic Resonance*, 92: pp. 102-114.
27. W.E.L. GRIMSON and E.C. HILDRETH, 1985, Comments on "digital step edges from zero crossings of second directional derivatives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7: pp. 121-127.
28. R.W. HAMMING, 1983, *Digital filters.* Signal Processing Series. Prentice-Hall, Englewood Cliffs, N.J., second edition.
29. R.M. HARALICK, 1984, Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence. 6: pp. 58-68.
30. R.M. HARALICK, 1985, Author's reply. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7: pp. 127-129.
31. O. KALLENBERG, 1983, *Random measures*. Akademie Verlag/Academic Press, Berlin/New York, third edition.
32. H. KIIVERI and N. CAMPBELL, Allocation of remote sensing data using Markov models for spectral variables and pixel labels. Preprint.
33. E. KLEIN and A.C. THOMPSON, 1984, *Theory of Correspondences*. Canadian Mathematical Society Monographs. Wiley.
34. J.-H. LIN, T. M. SELLKE and E.J. COYLE, 1990, Adaptive stack filtering under the mean absolute error criterion. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38: pp. 938-954.
35. G. MATHERON, 1975, *Random sets and integral geometry*. John Wiley and Sons, New York.
36. E. MICHAEL, 1951, Topologies on spaces of subsets. *Trans. Amer. Math. Soc.*, 71: pp. 152-182.
37. E. MOHN, N. HJORT and G. STORVIK, 1986, A comparison of some classification methods in remote sensing by a Monte Carlo study. Norwegian Computing Centre Report Kart/03/86, Norsk Regnesentral, Oslo.
38. F. NATTERER, 1986, *The mathematics of computerized tomography*. Teubner/Wiley, Stuttgart/Chichester.
39. T. NORBERG, 1986, Random capacities and their distributions. *Probability Theory and Related Fields*, 73: pp. 281-297.
40. T. NORBERG, 1989, Existence theorems for measures on continuous posets, with applications to random set theory. *Mathematica Scandinavica*, 64: pp. 15-51.
41. T.NORBERG, 1990, On the convergence of probability measures on continuous posets. Preprint 1990-08 ISSN 0347-2809, Department of Mathematics, Chalmers University of Technology and the University of Göteborg, Göteborg, Sweden.
42. T. NORBERG and W. VERVAAT, 1989, Capacities on non-Hausdorff spaces. Preprint 1989-11 ISSN 0347-2809, Department of Mathematics, Chalmers University of Technology and the University of Göteborg, Göteborg, Sweden.
43. G.L. O'BRIEN and W. VERVAAT, 1991, Capacities, large deviations and loglog laws. In G. SAMORODNITSKY, S. CAMBANIS and M.S. TAQQU, editors, *Stable Processes and related topics*, pp. 43-83. Birkhäuser.
44. W.A. PEARLMAN, 1978, A visual system model and a new distortion measure in the context of image processing. *J. Opt. Soc. Amer.*, 68: pp. 374-386.
45. T. PELI and D. MALAH, 1982, A study on edge detection algorithms. *Computer Graphics and Image Processing*. 20: pp. 1-21.
46. W.K. PRATT, 1977, *Digital image processing*. John Wiley and Sons, New York.
47. B.D. RIPLEY, 1988, *Statistical inference for spatial processes*. Cambridge University Press.
48. B.D. RIPLEY, 1986, Statistics, images and pattern recognition. *Canad. J.*

Statist, 14: pp. 83-111.

49. A. ROSENFELD and A.C. KAK, 1982, *Digital picture processing.* Academic Press, Orlando, 2nd edition.
50. A. ROSENFELD and J.L. PFALZ, 1966, Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, 13: p. 471,
51. A. ROSENFELD and J.L. PFALZ, 1968, Distance functions on digital pictures. *Pattern Recognition*, 1: pp. 33-61.
52. G. SALINETTI and R.J.B. WETS, 1986, On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Operations Research*, 11: pp. 385-419.
53. B. SENDOV, 1966, Hausdorffsche metrik und Approximation. *Numerische Mathematik*, 9: pp. 214-266.
54. B.SENDOV, 1969, Some questions on the theory of approximation of functions and sets in the Hausdorff metric. *Russian Mathematical Surveys*, 24: pp. 143-183.
55. B. SENDOV, 1990, *Hausdorff approximations*, volume 50 of *Mathematics and its Applications.* Kluwer, Dordrecht, Boston, London.
56. J. SERRA, 1982, *Image analysis and mathematical morphology.* Academic Press, London.
57. J. SERRA, editor, 1988, *Image analysis and mathematical morphology, volume 2: Theoretical advances.* Academic Press, London.
58. L.A. SHEPP, S.K. HILAL and R.A. SCHULZ, 1979, The tuning fork artifact in computerized tomography. *Computer Graphics and Image Processing*, 10: pp. 246-255.
59. L.A. SHEPP and J.A. STEIN, 1976, Simulated reconstruction artifacts in computerized X-ray tomography. In M. M. Ter-Pogossian, editor, *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine.*
60. L.A. SHEPP and Y. VARDI, 1982, Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1: pp. 113-122.
61. B.W. SILVERMAN, M.C. JONES, J.D. WILSON and D.W. NYCHKA, 1990, A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion). *Journal of the Royal Statistical Society, series B*, 52: pp. 271-324.
62. L.J. SPREEUWERS and F. VAN DER HEIJDEN, 1992, An edge detector evaluation method based on average risk. In W.FÖRSTNER and S. RUWIEDEL, editors, *Robust Computer Vision*, pp. 79-89, Karlsruhe, Wichmann.
63. D. STOYAN, W.S. KENDALL, and J. MECKE, 1987, *Stochastic Geometry and its Applications.* John Wiley and Sons, Chichester.
64. H.D. TAGARE and R.J.P. DEFIGUEIREDO, 1990, On the localization performance measure and optimal edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12: pp. 1186-1190.
65. C.C. TAYLOR, 1988, Measure of similarity between two images. Technical

report, University of Strathclyde.

66. L.J. VAN VLIET, I.T. YOUNG and A.L.D. BECKERS, 1989, A nonlinear Laplace operator as edge detector in noisy images. *Computer vision, graphics and image processing*, 45: pp. 167-195.
67. Y. VARDI, L.A. SHEPP and L. KAUFMAN, 1985, A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.*, 80.
68. W. VERVAAT, 1986, Stationary self-similar processes and random semicontinuous functions. In E. EBERLEIN and M.S. TAQQU, editors, *Dependence in probability and statistics*. Birkhäuser, Boston.
69. W. VERVAAT, 1987, Random semicontinuous functions and extremal processes. Research Report MS-R8801, Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands.
70. W. VERVAAT, 1988, Narrow and vague convergence of set functions. *Statistics and Probability Letters*, 6: pp. 295-298.
71. R. WIJSMAN, 1966, Convergence of sequences of convex sets, cones and functions II. *Transactions of the American Mathematical Society*, 123: pp. 32-45
72. B.S. YANDELL, C.C. TAYLOR and B.D. RIPLEY, 1990, Inference for image reconstructions. Draft manuscript.
73. C.K. YUEN and D. FRASER, 1979, *Digital spectral analysis*. CSIRO/Pitman, Melbourne.

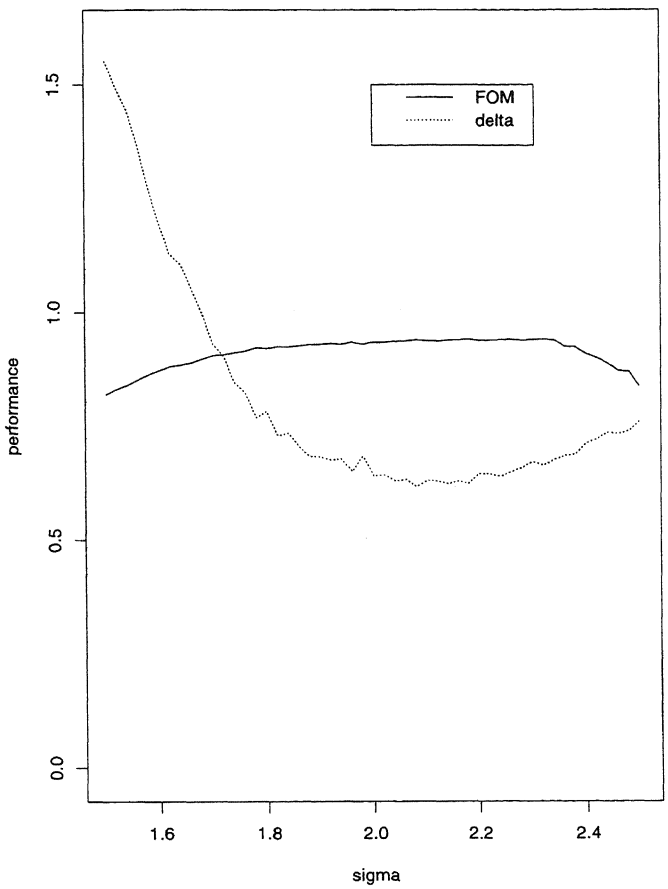

FIGURE 7. FOM and Δ^2 errors against smoothing parameter

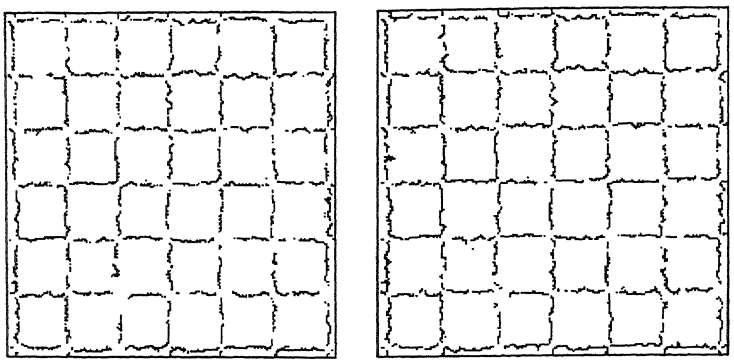

FIGURE 8. Laplace edge detector with FOM-optimal (left) and Δ-optimal smoothing

FIGURE 9. True image of text

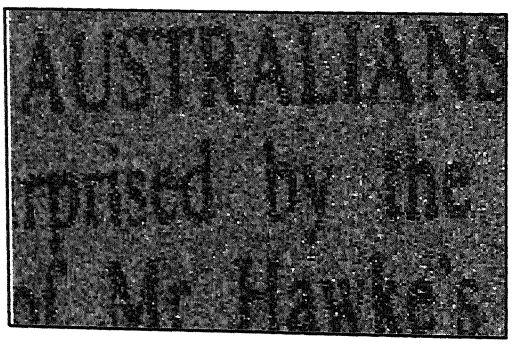

FIGURE 10. Greyscale data image for ICM algorithm

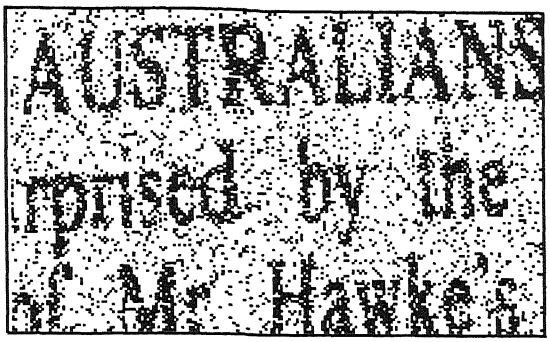

FIGURE 11. Pixelwise maximum likelihood classification

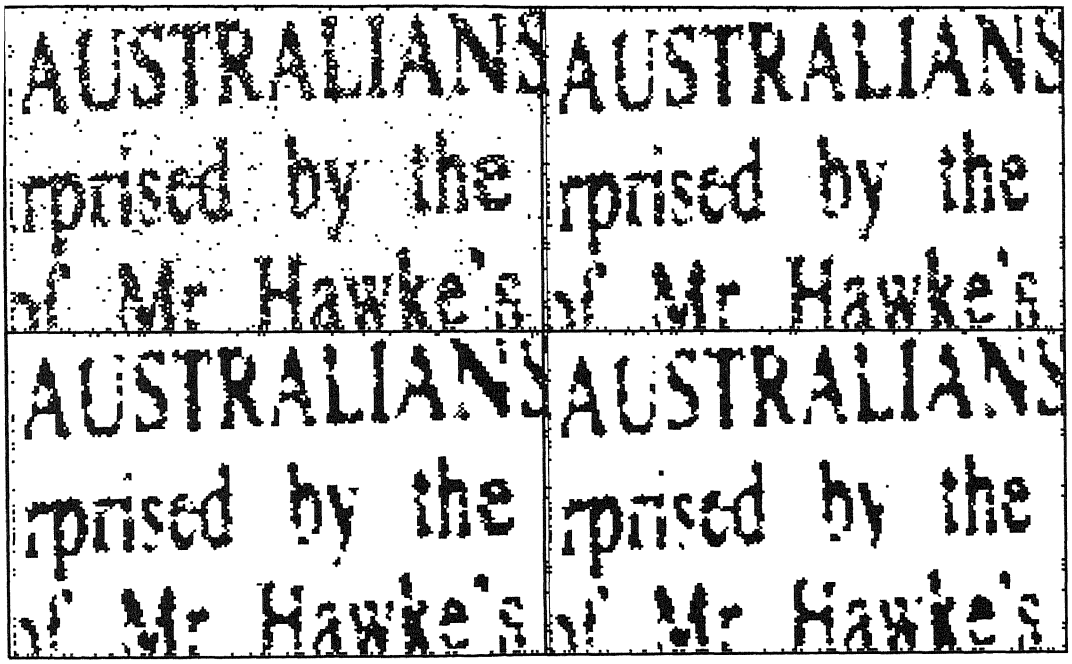

FIGURE 12. ICM iterations 1, 4, 7 and 10

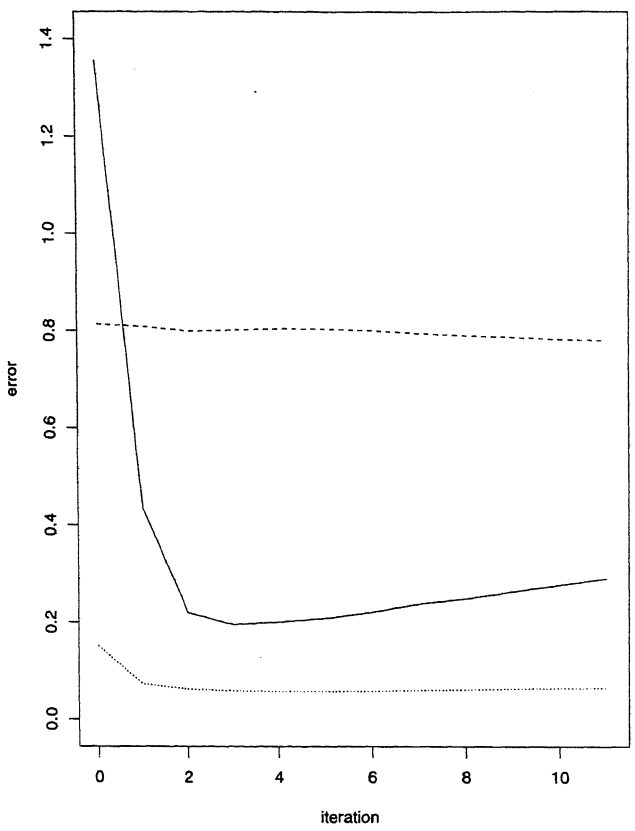

FIGURE 13. Error measured by Δ metric (solid lines), 1 – FOM (dashed lines) and misclassification error ϵ (dotted lines) for successive ICM iterations (iteration 0 is the MLE)

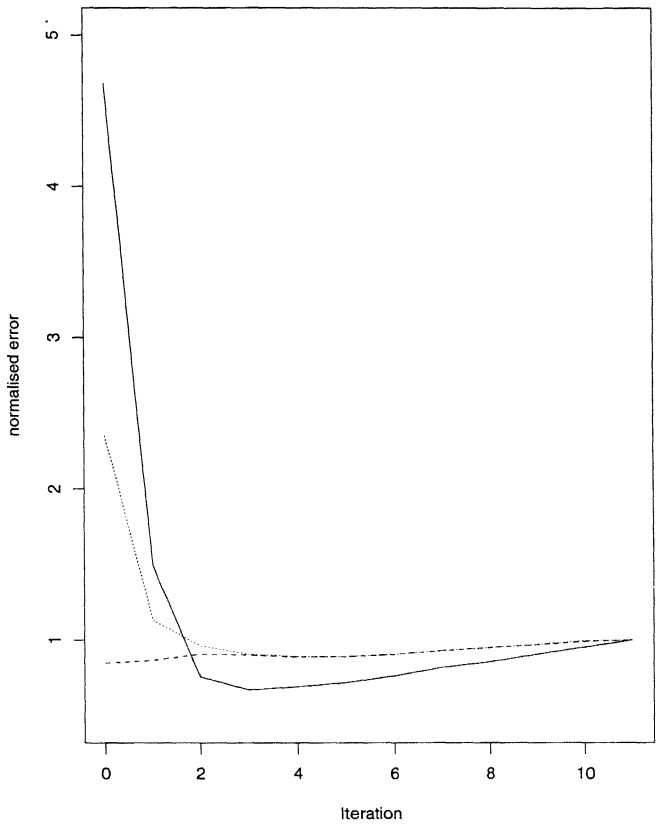

FIGURE 14. Error measures normalised by value at iteration 11