An error metric for binary images [1]

A J Baddeley

Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam
The Netherlands

The discrepancy between two binary images is traditionally measured by either the 'statistical' misclassification error rate, or by Pratt's [1] Figure Of Merit. We discuss weaknesses of these measures, and introduce an improved metric Δ which has both a theoretical basis [2] and intuitive appeal. The error measures are compared on artificial data and on the standard chessboard test for edge detectors.

1 Introduction

A numerical measure of the discrepancy between two binary images is important in studying the performance of image processing algorithms for applications such as edge detection in computer vision, and classification/segmentation in remote sensing. This paper introduces a new error metric for binary images, defined as the p-th order mean difference between thresholded distance transforms of the two images. This has a theoretical justification related to topological ideas in mathematical morphology [14, 27, 39] and random set theory [24, 29, 45, 50]. It also has some intuitive interpretations.

Theoretical development of the metric will be presented in a separate paper [2] (an earlier unsuccessful attempt was in [3]). The theory is also applicable to grey-level images, but here we describe only the implementation for binary images and compare it with standard measures such as the misclassification error rate and Pratt's [1, 31] figure of merit.

[1] Paper published in W. Förstner and S. Ruwiedel (eds.), *Robust Computer Vision: Quality of Vision Algorithms* (Proceedings, International Workshop on Robust Computer Vision, Bonn, 9-11 march 1992), Wichmann Verlag, Karlsruhe 1992. Pages 59–78.

2 Notation

Let X denote the pixel raster, assumed to be a finite set. A binary image is a function $b(x)$ of pixel $x \in X$ with values $b(x) = 0$ or 1. The value 1 will be interpreted as logical 'true' and displayed as black.

Of course, a binary image b can be identified with a subset $B \subseteq X$ by $B = \{x \in X : b(x) = 1\}$ and we shall do this without further mention. Useful set notation includes the set minus operator, $B \setminus A = \{x \in b : x \notin A\}$, and the set difference operator $A \triangle B = (A \setminus B) \cup (B \setminus A)$, the latter corresponding to exclusive-or. We will write $n(B)$ for the number of pixels in B.

3 Existing error measures and their problems

Error measures are frequently used in the design of algorithms for segmentation and classification (particularly for land cover type classification in remote sensing) [3, 7, 8, 25, 28, 33, 42, 43, 44, 47, 49] and edge detection in computer vision [1, 10, 19, 20, 22, 23, 30, 31, 46, 48]. Important general principles for error measurement were enunciated by Canny [10] in the context of edge detection. He argued that a good edge filter should exhibit

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 there is an uncertainty principle balancing good detection against good localization: optimal edge filtering involves a tradeoff between these two criteria.

Error measures for detection and localization were surveyed by Peli & Malah [30] and van Vliet et al [48]; we will now study them.

3.1 Detection performance ("statistical") measures

These measures report the frequency of incorrect classification for individual pixels. 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. The type I error rate is

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

and the type II error rate

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

where again $n(A)$ = number of pixels in A. Note that these rates are relative to pixels of a particular class in the **true** image. Derived quantities discuused in [30] are the binary noise-to-signal ratio

$$NSR = \frac{n(B \setminus A)}{n(B \cap A)} = \frac{\alpha}{1-\beta}\frac{1-r}{r}$$

where $r = n(A)/n(X)$ is the area fraction of ideal edge pixels; and the "mean width of the detected edge"

$$\frac{n(B)}{n(A)} = 1 - \beta + \alpha\frac{1-r}{r}$$

appropriate when $B \supseteq A$. Fram & Deutsch [12, 15] used combinations of row-wise and column-wise error rates (applicable for straight edges only).

Recent statistical research on classification and segmentation algorithms [7, 8, 17, 18, 25, 28, 33, 32] almost exclusively uses the *pixel misclassification error rate*

$$\epsilon(a,b) = \frac{n\{x : a(x) \neq b(x)\}}{n(X)}$$

where the pixel values $a(x), b(x)$ are arbitrary (e.g. labels, grey values). For binary images, this reduces to

$$\epsilon(A,B) = \frac{n(A \triangle B)}{n(X)} = \alpha(1-r) + \beta r.$$

Misclassification error ϵ has some theoretical advantages over α and β because it is symmetric in A and B, and does not call for normalisation with respect to the true image A. It is a special case of of the L^1 metric or mean absolute error favoured in recent developments [16, 26]. On the other hand, α and β can be more informative and understandable.

If the objective of classification or segmentation is merely to estimate the *number* of pixels of each class, then ϵ, α, β are reasonable measures of error. Note that the mean value of $\epsilon(A,B)$ under some stochastic model equals the average over all pixels x of the disagreement probability $\mathbb{P}\{a(x) \neq b(x)\}$; similar statements hold for α and β. In this sense ϵ, α, β deserve to be called 'statistical' measures.

However, if the objective is to produce an image (say, a map of land tenure, or an edge image), then it is widely acknowledged [8, p.299], [33, pages 97,110], [40, 41, 42] 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 spike or filling-in of holes, 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 [8, discussion], [17, 33, 32]. These comments support Canny's observations.

3.2 Localization performance ("distance") measures

We now assume that the distance $\rho(x, y)$ between any two pixels $x, y \in X$ has been defined and satisfies the formal axioms of a metric (see §4). Let $d(x, A)$ denote the shortest distance from pixel $x \in X$ to $A \subseteq X$:

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

with $d(x, \emptyset) \equiv \infty$. For standard pixel distance metrics ρ on a rectangular or hexagonal grid X, the function $d(\cdot, A)$ can be computed rapidly by the distance transform algorithm [9, 35, 36].

Measures of localization performance discussed by Peli and Malah [30] were the mean error distance

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

the mean square error distance

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

and Pratt's [1, 31] "figure of merit"

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

where α 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 $\mathrm{FOM}\,(A, B) \neq \mathrm{FOM}\,(B, A)$.

FOM is the most popular of these and is widely used [1, 4, 19, 30, 31]. The author is not aware of any theoretical justification for it, however. FOM can be appreciated as a kind of average localization error **for the type I errors only**. The normalization is designed so that $0 < \mathrm{FOM}\,(A, B) \leq 1$ and $\mathrm{FOM}\,(A, B) = 1$ iff $A = B$.

The following criticisms should be recorded:

1. FOM is not sensitive to type II errors, except indirectly through the normalising factor in (1). For example if there are no type I errors, $B \subseteq A$, then $\mathrm{FOM}(A,B) = n(B)/n(A) = 1-\beta$ regardless of the positions of the type II errors.

2. FOM is not sensitive to the pattern of error pixels, since it is the *average* of a function $f(x,A)$ over all type I error pixels x. A dramatic example, found by Peli and Malah[30], 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.

3. Peli and Malah [30] and van Vliet et al [48] 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 [48, p. 186 and section 6] in the case of the classical Laplacian operator the FOM-optimal images often had large sections of the true contour missing and there were high-frequency oscillations around the true contour. This behaviour can be explained by noticing that for $x \notin A \cup B$ we have $\mathrm{FOM}(A, B\cup\{x\}) < \mathrm{FOM}(A,B)$ if and only if $d(x,A) > \alpha^{-1/2}\left(\mathrm{FOM}(A,B)^{-1}-1\right)^{1/2}$. If $\mathrm{FOM}(A,B) > 0.9$ and $\alpha = 1/9$ then this happens whenever $d(x,A) \geq 1$. That is, when FOM is large, it will always prefer to commit a type II error than a type I error, however innocuous.

4. The same 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.

FOM seems difficult to interpret because of the normalisation by a variable denominator $\max\{n(A), n(B)\}$. 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.

3.3 Hausdorff metric

Define the Hausdorff distance between two subsets $A, B \subseteq X$ by

$$H(A,B) = \max\left\{\sup_{x\in A} d(x,B), \sup_{x\in B} d(x,A)\right\} \tag{2}$$

using the pixel distance metric ρ and distance transform $d(\cdot, A)$ as above. Thus $H(A,B)$ is the maximum distance from a point in one set to the nearest point in the other set.

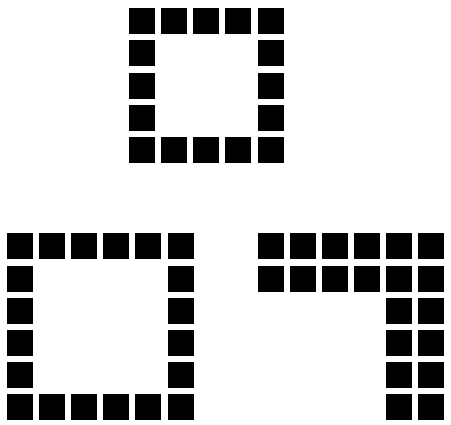

Figure 1: Peli-Malah counterexample. True picture A (top) and two error pictures B_1, B_2 (bottom) with the same value of FOM.

For the empty set $\emptyset$ put $H(\emptyset, \emptyset) = 0$ and $H(\emptyset, B) = H(B, \emptyset) = \infty$ for $B \neq \emptyset$.

The Hausdorff distance is theoretically interesting and important. It is a metric (see §4) on the class of all subsets of X if X is finite, or all compact sets in $X = \mathbb{R}^d$. It has natural connections with the basic operations of mathematical morphology and stochastic geometry. It generates the myopic topology for compact sets in $\mathbb{R}^d$ and a modification generates the hit-or-miss topology for closed sets in $\mathbb{R}^d$. For explanation and discussion see [39, pp. 63-92]; for proofs [27, p. 15], [51, thm 3.1], [5, 6, 11]. The discussion in [39] makes it clear that continuity with respect to H is a very desirable property for image processing algorithms.

However H is never used in practice (to the author's knowledge) as an error measure for images. It has been used to measure differences of sets in functional analysis [37, 38].

The problem is that H is painfully sensitive to noise. A single error pixel can cause elevation of H to its maximum possible value, because of the supremum in the definition (2). Another aspect of this is the "minimax property"

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

The Hausdorff metric is therefore so unstable as to be unusable in this context.

4 On the selection of metrics

This section records some general principles about the construction and selection of error measures. We take stock of the problems encountered in section 3 and suggest solutions.

4.1 Properties required

It was argued in [3] that there are at least three separate uses for error measures: they may provide

(a) a theoretical framework for deriving 'optimal' algorithms;

(b) a numerical benchmark for quantifying the performance of an algorithm in a computer experiment;

(c) a measure of achieved quality.

Here we are interested in (a)–(b), which call for direct comparison of the 'true' image A with the image B resulting from the algorithm. Thus the desired error measure is a quantity $\Delta(A, B) \geq 0$ defined for binary images A, B. We argue that Δ should satisfy the axioms of a **metric:**

- $\Delta(A, B) = 0$ if and only if $A = B$;
- symmetry: $\Delta(A, B) = \Delta(B, A)$;
- triangle inequality: $\Delta(A, B) \leq \Delta(A, C) + \Delta(C, B)$.

See, e.g. [13, chap. XI] for theory, and [39, p.72ff] for a discussion in the context of image processing. The metric property is theoretically important because it generates a topology that defines notions of continuity and convergence. That is, a topology allows us to make sense of statements like "when B is close to A then $f(B)$ is close to $f(A)$" for some derived property $f(A)$ of image A.

A metric is desirable for theoretical optimality (a); optimal Wiener filtering theory [21, 34, 52] is based on the L^2 or root-mean-square metric. However, for practical experiments (b), the metric properties are arguable. 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 (b).

Note that a metric is a specific numerical measure of the 'closeness' of images A, B, while a topology merely determines which functionals f will be continuous, i.e. satisfy statements "when B is close to A, then $f(B)$ is close to $f(A)$". A metric generates a topology; two different metrics may generate the same topology. The most natural procedure is to first decide which functionals f should be continuous in the topology; this determines the topology; then to choose a metric that generates that topology.

It is also important to distinguish topologies and *uniformities* [13, 200–204] (not mentioned in [27, 39]). A functional f is *uniformly continuous* with respect to Δ if we can guarantee $|f(B) - f(A)| < t$ for a given tolerance t by requiring $\Delta(A, B) < s$ where s depends only on t. Two metrics may generate the same topology, yet not generate the same uniformity. Many of the criticisms of specific metrics in section 3 were related to the uniformity generated by the metric. One can change a metric so as to preserve the desired topology but change the undesired uniformity.

4.2 Tempering distances

Problems with over-sensitivity to large error distances in a few pixels, noted in 3.2 and 3.3, are really associated with values of the pixel distance ρ. This can be moderated by transforming ρ.

Lemma 1 *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,

$$w(t) = 0 \text{ iff } t = 0.$$

If ρ is a pixel distance metric on X, then so is $\tau = w \circ \rho$, i.e. the metric

$$\tau(x, y) = w(\rho(x, y)). \tag{3}$$

The metrics τ and ρ generate the same topology and the same uniformity.

Examples include

$$w(t) = \frac{t}{1 + t} \tag{4}$$

$$w(t) = \tan^{-1}(t) \tag{5}$$

$$w(t) = \min\{t, c\} \tag{6}$$

for a fixed $c > 0$. These examples transform an unbounded metric to a bounded one. Choice (6) corresponds to "giving up" distance measurement beyond a cutoff distance c.

The effect of the transform (3) on the Hausdorff metric H is particularly simple: the function w is just applied to the result of H.

Lemma 2 *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.*

$$H_\tau(A, B) = w(H_\rho(A, B)) \tag{7}$$

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

The case (6) is special, for if we regard subsets $A \subseteq X$ of a finite X as mass densities assigning mass c to each point in A, then H_w is identical to the Lévy-Prohorov metric [2] for bounded measures, which plays an important role in probability theory.

Incidentally we note

$$1 - \mathrm{FOM}\,(A,B) = \frac{1}{\max\{n(A),n(B)\}} \sum_{x\in B} \frac{\alpha d(x,A)^2}{1+\alpha d(x,A)^2} \tag{8}$$

so that $1 - \mathrm{FOM}$ is interpretable as an average of transformed distance values, analogous to $\bar{\epsilon}$. However the transform $f(t) = \alpha t^2/(1+\alpha t^2)$ is not concave, so FOM cannot be interpreted as the mean distance in a transformed pixel distance metric.

5 The new metric

We have seen that the Hausdorff metric has the right topological properties, but is far too sensitive because it is based on a supremum of distance values. Our idea is to replace this supremum by a mean or p-th order mean.

If we naively replace the sup and/or max in (2) by a p-th order mean, the result is not a metric. Intuitively this is because (as with FOM) the domain of averaging depends on the sets A, B. The way out is indicated by the following observation:

Lemma 3 *For any ρ and for any $A, B \subseteq X$*

$$H(A,B) = \sup_{x\in X} |d(x,A) - d(x,B)|$$

To prove this we notice that for $x \in A$ the distance to A is zero so $|d(x,A) - d(X,B)| = d(x,B)$. Similarly for $x \in B$, so that $\sup_{x\in X} |d(x,A) - d(x,B)| \geq H(A,B)$. For the converse we use the key property

$$d(x,A) \leq \rho(x,y) + d(y,A). \tag{9}$$

Fix $x \in X$; then by definition of $d(x,B)$ for any $\epsilon > 0$ there must exist $b \in B$ such that $\rho(x,b) < d(x,B) + \epsilon$. Applying (9) gives $d(x,A) \leq \rho(x,b) + d(b,A) < d(x,B) + d(b,A) + \epsilon$ so that $d(x,A) - d(x,B) < d(b,A) + \epsilon$. Interchanging A, B and taking suprema we find $\sup_{x\in X} |d(x,A) - d(x,B)| \leq H(A,B) + \epsilon$ and since ϵ was arbitrary the result follows.

The idea is now to simply replace the supremum by an average.

Definition 4 *For $1 \leq p < \infty$ define*

$$\Delta^p(A,B) = \left[\frac{1}{N} \sum_{x\in X} |d(x,A) - d(x,B)|^p\right]^{1/p} \tag{10}$$

where $N = n(X) =$ total number of pixels in the raster.

It is obvious (see [2]) that Δ^p is an image metric.

The previous lemma holds also for the transformed metrics

$$H_w(A,B) = \sup_{x\in X} |w(d(x,A)) - w(d(x,B))|$$

so we may more generally define

$$\Delta_w^p(A,B) = \left[\frac{1}{N}\sum_{x\in X} |w(d(x,A)) - w(d(x,B))|^p\right]^{1/p} \tag{11}$$

and for a concave continuous function w which is strictly increasing at 0.

Implementation is straightforward: we apply the distance transform algorithm of Rosenfeld and Borgefors [9, 35, 36] to compute $d(\cdot, A)$ and $d(\cdot, B)$, transform the distance values by the function w, then take the pth order mean difference.

Intuitively $\Delta^p(A,B)$ measures the fidelity, or extent to which each image can be used as a 'replacement' for the other, in the sense that replacing A by B will disturb the scene (as expressed by the distance transform) by an amount given by Δ^p.

Since $d(x,A) = 0$ for $x \in A$, the sum in (11) includes contributions $\sum_{x\in A} w(d(x,B))^p$ and $\sum_{x\in B} w(d(x,A))^p$ which are analogous to FOM as seen in (8). However, the sum in (11) also includes other terms for x outside A and B.

In applications we shall always use the cutoff transformation (6),

$$w(t) = \min\{t, c\}$$

for a fixed $c > 0$. In this case the contributions to the sum in (11) 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.

Parameters c and p determine the tradeoff between localization error and misclassification error. The value of c 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 $\frac{1}{c}\Delta_w^p(A,B) \to \epsilon(A,B)^{1/p}$. The value of p determines the relative importance of large localization errors. For large p the effect is similar to the Hausdorff metric; $\Delta_w^\infty(A,B) = H_w(A,B)$.

Counterintuitively, the metric Δ^p is topologically equivalent to the Hausdorff metric [2].

Lemma 5 *If X is a discrete raster, then any sequence of images A_n, $n = 1, 2, \ldots$ converges in the new metric, $\Delta_w^p(A_n, A) \to 0$, if and only if it converges in the Hausdorff metric, $H(A_n, A) \to 0$.*

The reason is essentially that distance functions satisfy a Lipschitz property (9) ensuring that L^p convergence is equivalent to L^∞ convergence.

For continuous space X, theorems in [2, section 7] establish that the analogue of Δ_w^p with $w(t) = \min\{t, c\}$ is topologically equivalent to the Hausdorff metric. Under appropriate conditions Δ_w^p generates either the myopic topology or the hit-or-miss topology [27, 39].

6 Examples

Throughout this section we have compared the figure of merit FOM for $\alpha = 1/9$ with the Δ^2 metric for cutoff $c = 5$.

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 values are 0.323 (left picture) and 0.512 (right picture).

6.2 Artificial data

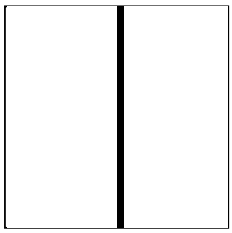

Figure 2: Synthetic true image A

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

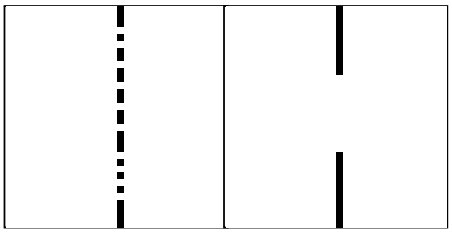

Figure 3: Images `gaps` and `lost`

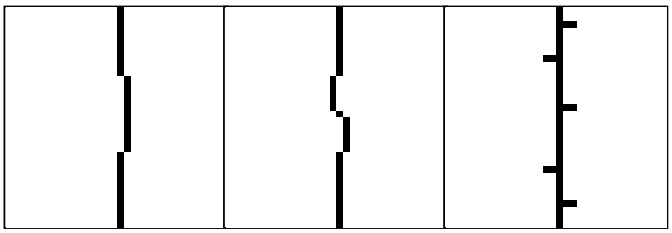

Figure 4: Images `shift`, `bend` and `barbs`

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`, `barbs` and `bend`, while Δ^2 spreads them over a wide range.

Image **B**	α	β	FOM	Δ
`gaps`	0	0.313	0.688	0.149
`lost`	0	0.344	0.656	0.682
`shift`	0.011	0.344	0.966	0.319
`bend`	0.010	0.313	0.969	0.291
`barbs`	0.010	0	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 [22] (see [20, 23, 48]) and compares optimality under FOM and under Δ_2.

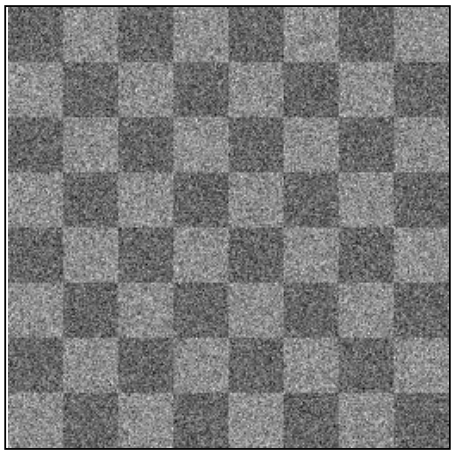

Figure 5: Chessboard image with additive Gaussian noise (SNR = 2)

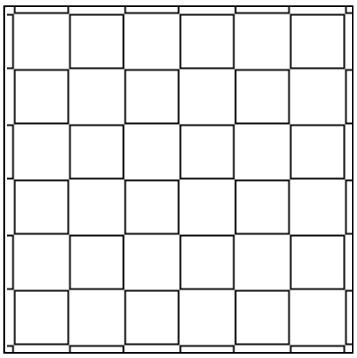

Figure 6: True edges of chessboard

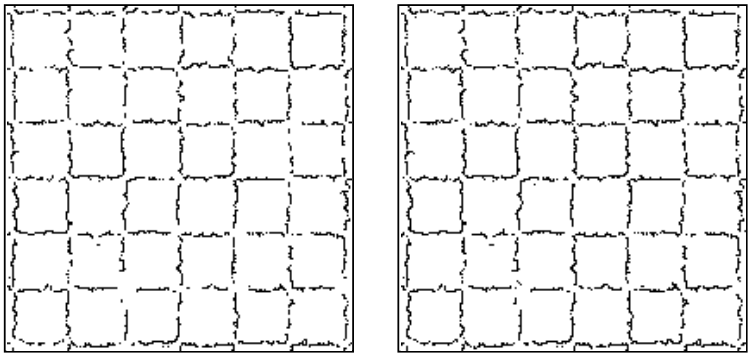

Figure 7: FOM-optimal threshold (left) and Δ-optimal threshold

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 2.0, followed by the classical 4-connected Laplacian, zero-crossing by thresholding and distance transform [48], then the Lee-Haralick morphological edge strength detector with a pseudocircular mask of size 3 [48] was applied to the smoothed data and the result multiplied by the zero-crossing image. The resulting image gives edge positions with edge strengths. We then thresholded this image at all possible levels to obtain binary images B, and compared $\mathrm{FOM}(A, B)$ with $\Delta_c^2(A, B)$. The FOM parameter α was set to the usual $1/9$ and the cutoff parameter of Δ was $c = 5$.

Figure 7 shows the FOM-optimal and Δ-optimal thresholded images, and Figure 8 the difference. The results are similar but the FOM optimal threshold value was higher; the lost pixels (Figure 8) have enlarged the gaps in the edge contour. This is consistent with our theoretical comments about FOM. The plot of FOM and Δ values in Figure 9 shows that FOM is almost indifferent to a wide range of thresholds, near its maximum.

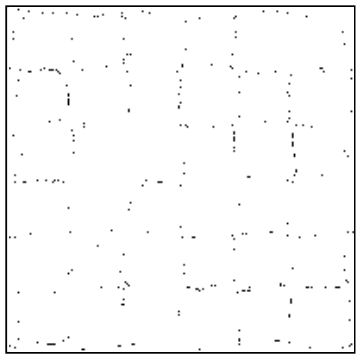

Figure 8: Difference between FOM-optimal and Δ-optimal thresholds

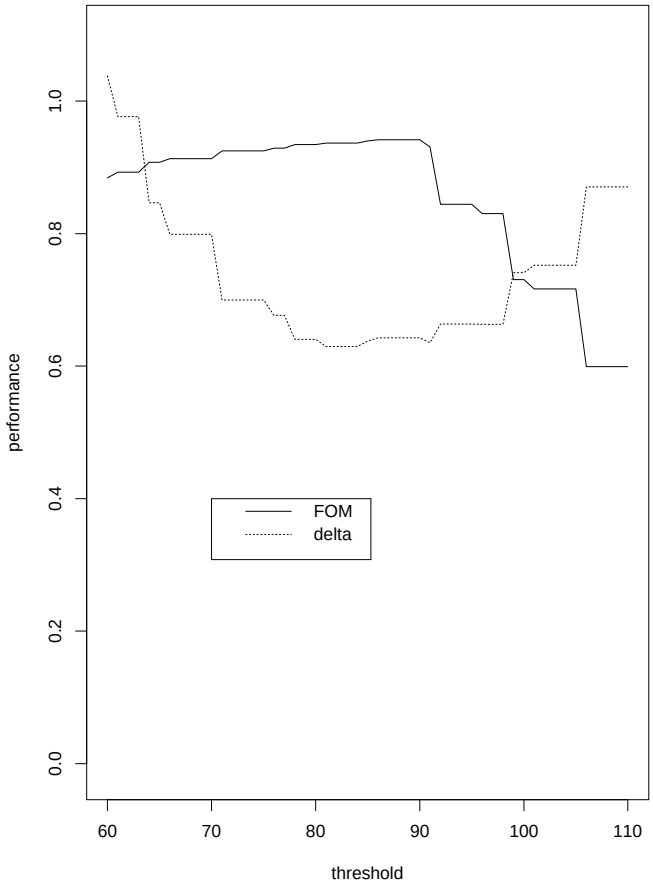

Figure 9: FOM and Δ^2 errors for the thresholding experiment

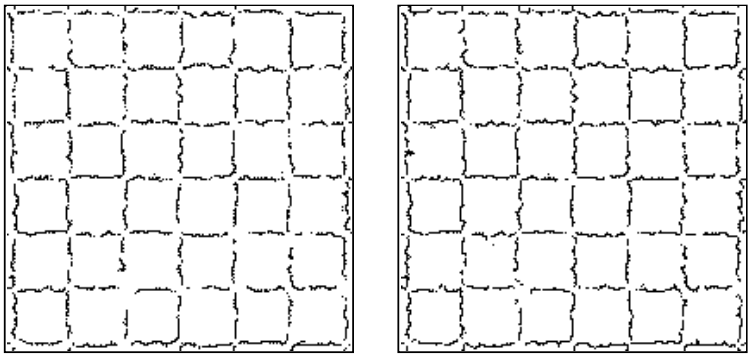

Figure 10: Laplace edge detector with FOM-optimal (left) and Δ-optimal smoothing

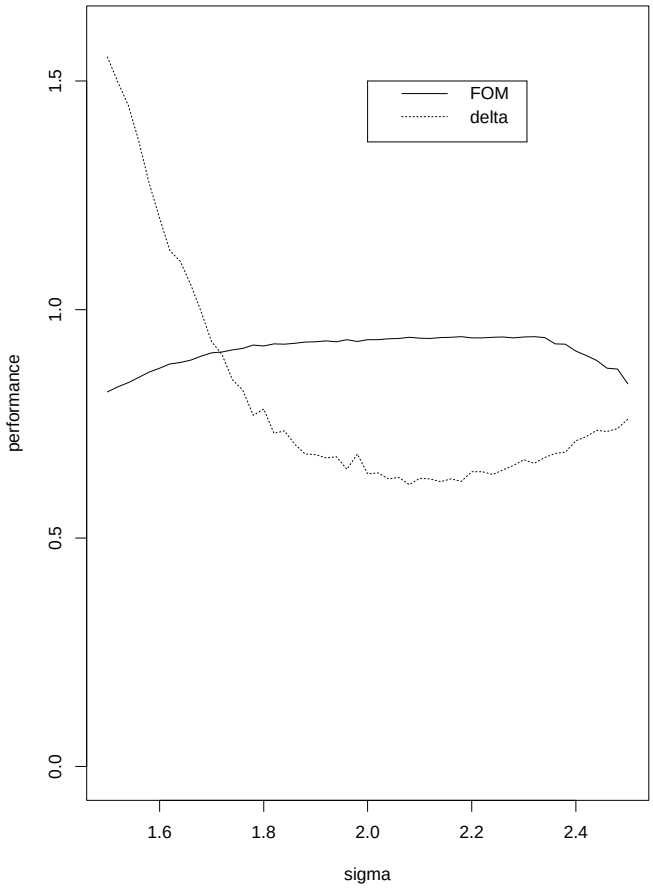

Figure 11: FOM and Δ^2 errors against smoothing parameter

In the second part of this experiment we varied the standard deviation parameter σ of the Gaussian smoothing, and fixed the final threshold level and all other parameters as described above. Again we used FOM and Δ^2 to select optimal values of σ: the results are shown in Figure 10.

This time the differences are quite marked: FOM has chosen $\sigma = 2.32$ and Δ has chosen $\sigma = 2.08$. The plot of FOM and Δ values (Figure 11) again shows that FOM appears less sensitive than Δ to changes around its optimum. The FOM values for Figure 10 were 0.941 and 0.939; the corresponding Δ_2 values were 0.663 and 0.617.

Acknowledgements

The author is grateful to J D Boyd, M N M van Lieshout, P Nacken, A G Steenbeek, W Vervaat and P de Waal for their helpful comments and advice.

References

[1] I. E. Abdou and W. K. Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67:753–763, 1979.

[2] A. J. Baddeley. Hausdorff metric for capacities. Submitted for publication.

[3] A. J. Baddeley. A class of image metrics. In *Australian and New Zealand Association for the Advancement of Science, Proceedings 11th Congress, Townsville 1987*, 1987.

[4] D. G. Bailey and R. M. Hodgson. Range filters: intensity subrange filters and their properties. *Image Vision Computing*, 3:99–110, 1985.

[5] G. Beer. On convergence of closed sets in a metric space and distance functions. *Bulletin of the Australian Mathematical Society*, 31:421–432, 1985.

[6] G. Beer. Metric spaces with nice closed balls and distance functions for closed sets. *Bulletin of the Australian Mathematical Society*, 35:81–96, 1987.

[7] J. Besag. Discussion of paper by P. Switzer. *Bulletin of the International Statistical Institute*, 50:422–425, 1983.

[8] J. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, series B*, 48:259–302, 1986.

[9] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics and Image Processing*, 34:344–371, 1986.

[10] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[11] C. Castaing and M. Valadier. *Convex analysis and measurable multifunctions.* Lecture Notes in Mathematics 580. Springer, 1977.

[12] E. S. Deutsch and J. R. Fram. A quantitative study of the orientation bias of some edge detector schemes. *IEEE Transactions on Computing*, 27:205–213, 1978.

[13] J. Dugundji. *Topology.* Allyn and Bacon, Boston, Mass., 1966.

[14] J. M. G. Fell. A Hausdorff topology for the closed subsets of a locally compact non-Hausdorff space. *Proc. Amer. Math. Soc.*, 13:472–476, 1962.

[15] J.R. Fram and E.S. Deutsch. On the evaluation of edge detector schemes and their comparison with human performance. *IEEE Transactions on Computing*, 24:616–628, 1975.

[16] M. Gabbouj and E. J. Coyle. Minimum mean absolute error stack filtering with structural constraints and goals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:955–968, 1990.

[17] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[18] S. Geman and D.E. McClure. Bayesian image analysis: application to single photon emission tomography. Preprint, Applied Mathematics Dept., Brown University.

[19] J.J. Gerbrands, E. Backer, and W. A. G. van der Hoeven. Quantitative evaluation of edge detection by dynamic programming. In Gelsema and Kanal, editors, *Pattern Recognition in Practice II*, pages 91–99. Elsevier, North-Holland, New York, 1986.

[20] W. E. L. Grimson and E. C. Hildreth. Comments on "digital step edges from zero crossings of second directional derivatives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:121–127, 1985.

[21] R.W. Hamming. *Digital filters.* Signal Processing Series. Prentice-Hall, Englewood Cliffs, N.J., second edition, 1983.

[22] R. M. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:58–68, 1984.

[23] R. M. Haralick. Author's reply. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:127–129, 1985.

[24] O. Kallenberg. *Random measures.* Akademie Verlag/Academic Press, Berlin/New York, third edition, 1983.

[25] H. Kiiveri and N. Campbell. Allocation of remote sensing data using Markov models for spectral variables and pixel labels. Preprint.

[26] J.-H. Lin, T. M. Sellke, and E. J. Coyle. Adaptive stack filtering under the mean absolute error criterion. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:938–954, 1990.

[27] G. Matheron. *Random sets and integral geometry.* John Wiley and Sons, New York, 1975.

[28] E. Mohn, N. Hjort, and G. Storvik. A comparison of some classification methods in remote sensing by a Monte Carlo study. Norwegian Computing Centre Report Kart/03/86, Norsk Regnesentral, Oslo, 1986.

[29] T. Norberg. Random capacities and their distributions. *Probability Theory and Related Fields*, 73:281–297, 1986.

[30] T. Peli and D. Malah. A study on edge detection algorithms. *Computer Graphics and Image Processing*, 20:1–21, 1982.

[31] W.K. Pratt. *Digital image processing.* John Wiley and Sons, New York, 1977.

[32] B. D. Ripley. *Statistical inference for spatial processes.* Cambridge University Press, 1988.

[33] B.D. Ripley. Statistics, images and pattern recognition. *Canad. J. Statist*, 14:83–111, 1986.

[34] A. Rosenfeld and A. C. Kak. *Digital picture processing.* Academic Press, Orlando, 2nd edition, 1982.

[35] A. Rosenfeld and J. L. Pfalz. Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, 13:471, 1966.

[36] A. Rosenfeld and J. L. Pfalz. Distance functions on digital pictures. *Pattern Recognition*, 1:33–61, 1968.

[37] B. Sendov. Hausdorffsche metrik und Approximation. *Numerische Mathematik*, 9:214–266, 1966.

[38] B. Sendov. Some questions on the theory of approximation of functions and sets in the Hausdorff metric. *Russian Mathematical Surveys*, 24:143–183, 1969.

[39] J. Serra. *Image analysis and mathematical morphology.* Academic Press, London, 1982.

[40] L. A. Shepp. Personal communication.

[41] L. A. Shepp, S. K. Hilal, and R. A. Schulz. The tuning fork artifact in computerized tomography. *Computer Graphics and Image Processing*, 10:246–255, 1979.

[42] L. A. Shepp and J. A. Stein. Simulated reconstruction artifacts in computerized X-ray tomography. In M. M. Ter-Pogossian, editor, *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, 1976.

[43] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1:113–122, 1982.

[44] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka. 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*, 52:271–324, 1990.

[45] D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* John Wiley and Sons, Chichester, 1987.

[46] H. D. Tagare and R. J. P. deFigueiredo. On the localization performance measure and optimal edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1186–1190, 1990.

[47] C. C. Taylor. Measure of similarity between two images. Technical report, University of Strathclyde, 1988.

[48] L. J. van Vliet, I. T. Young, and A. L. D. Beckers. A nonlinear Laplace operator as edge detector in noisy images. *Computer vision, graphics and image processing*, 45:167–195, 1989.

[49] Y. Vardi, L.A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.*, 80:8–, 1985.

[50] W. Vervaat. Narrow and vague convergence of set functions. *Statistics and Probability Letters*, 6:295–298, 1988.

[51] R. Wijsman. Convergence of sequences of convex sets, cones and functions II. *Transactions of the American Mathematical Society*, 123:32–45, 1966.

[52] C.K. Yuen and D. Fraser. *Digital spectral analysis.* CSIRO/Pitman, Melbourne, 1979.