A New Normalized Supervised Edge Detection Evaluation

Hasan Abdulrahman, Baptiste Magnier, Philippe Montesinos

A NEW NORMALIZED SUPERVISED EDGE DETECTION EVALUATION

Hasan Abdulrahman, Baptiste Magnier, and Philippe Montesinos

Ecole des Mines d'Alès, Parc scientifique Georges Besse, 30000 Nîmes, France

Abstract. In digital images, edges characterize object boundaries, then their detection remains a crucial stage in numerous applications. To achieve this task, many edge detectors have been designed, producing different results, with different qualities. Evaluating the response obtained by these detectors has become a crucial task. In this paper, several referenced-based boundary detection evaluations are detailed, pointing their advantages and disadvantages through concrete examples of edge images. Then, a new supervised edge map quality measure is proposed, comparing a ground truth contour image, the candidate contour image and their associated spacial nearness. Compared to other boundary detection assessments, this new method has the advantage to be normalized and remains a more reliable edge map quality measure.

Keywords: Edge detection, distance measure, supervised evaluation.

1 IMPORTANCE OF A NEW ERROR MEASURE

In image processing tasks, edge detection remains a key point in many applications. Boundaries include the most important structures of the image, and an efficient boundary detection method should create a contour image containing edges at their correct locations with a minimal of misclassified pixels. Different algorithms have been developed in the past, but few of them give an objective performance comparison. The evaluation process should produce a result that correlates with the perceived quality of the edge image, which is relied on human judgment. In other words, a reliable edge map should characterize all the relevant structures of an image. On the other hand, a minimum of spurious pixels or holes (oversights) must be created by the edge detector at the same time. Therefore, an efficient evaluation can be used to assess and improve an algorithm, or to optimize edge detector parameters [1].

The measurement process can be classified into either an unsupervised or a supervised evaluation criteria. The first class of methods exploits only the input contour image and gives a score of coherence that qualifies the result given by the algorithm [1]. The second one computes a dissimilarity measure between a segmentation result and a ground truth obtained from synthetic data or an expert judgment (i.e. manual segmentation) [2][3][4]. This work focusses on comparisons of supervised assessment of edge detection evaluations. Furthermore, a new supervised edge map quality measure based on the distances of misplaced pixels is presented and compared to the others, using synthetic and real images.

2 SUPERVISED IMAGE CONTOUR EVALUATIONS

As introduced above, a supervised evaluation process estimates scores between a ground truth and a candidate edge map. In image processing, the Structural Similarity Index ($SSIM$) corresponds to an image quality evaluation, which estimates the visual impact of gray scale shifts in an image [5]. Otherwise, contours (binary images) could be evaluated counting the number of erroneous pixels, but also throughout spatial distances of misplaced or oversights contours.

2.1 Error measures involving only the confusion matrix

Let G_t be the reference contour map corresponding to ground truth and D_c the detected contour map of an image I. Comparing pixel per pixel G_t and D_c, common positive or negative presence of points is the first criterion to be assessed. A basic evaluation is compounded of statistics issued of a confusion matrix. To that effect, G_t and D_c are combined. Afterward, denoting $|\cdot|$ the cardinality of a set, all points are partitioned into four sets:

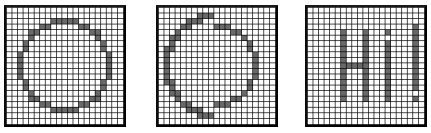
- True Positive points (TPs), common points of G_t and D_c: $TP = |D_c \cap G_t|$,
- False Positive points (FPs), spurious detected edges of D_c: $FP = |D_c \cap \neg G_t|$,
- False Negative points (FNs), missing boundary points of D_c: $FN = |\neg D_c \cap G_t|$,
- True Negative points (TNs), common non-edge points: $TN = |\neg D_c \cap \neg G_t|$.

In one hand, let us consider boundary detection of images, FPs appear in the presence of noise, texture or other contours influencing the filter used by the edge detection operator. In the other hand, FNs represent holes in a contour of D_c. Finally, a wrong threshold of the segmentation could generate both FPs and FNs. Computing only FPs and FNs enables a segmentation assessment [6][7], and a reliable edge detection should minimize the following indicators [3]:

$$\begin{cases} \text{Over-detection error} : Over(G_t, D_c) = \frac{FP}{|I|+|G_t|}, \\ \text{Under-detection error} : Under(G_t, D_c) = \frac{FN}{|G_t|}, \\ \text{Localization-error} : Loc(G_t, D_c) = \frac{FP+FN}{|I|}. \end{cases}$$

Additionally, the $Performance\ measure\ P_m^*$ presented in Table 1 considers directly at the same time the three entities TP, FP and FN to assess an a binary image. The obtained score reflects the percentage of statistical errors.

Another way to display evaluations are Receiver Operating Characteristic (ROC) [8] curves or Precision-Recall (PR) [9], involving *True Positive Rates* ($TPR = \frac{TP}{TP+FN}$) and *False Positive Rates* ($FPR = \frac{FP}{FP+TN}$). Derived from TPR and FPR, the three measures Φ, χ^2 and F_α (see Table 1) are frequently used in edge detection assessment.



(a) G_t, $|G_t| = 68$. (b) D_1, $FP = 54$, $D_1 \cap G_t = \varnothing$. (c) D_2, $FP = 54$, $D_2 \cap G_t = \varnothing$.

$Over(G_t, D_{\{1,2\}}) = 0.1224$
$Under(G_t, D_{\{1,2\}}) = 1$
$Loc(G_t, D_{\{1,2\}}) = 0.2449$
$P_m^*(G_t, D_{\{1,2\}}) = 1$
$\Phi^*(G_t, D_{\{1,2\}}) = 1$
$\chi^{2*}(G_t, D_{\{1,2\}}) = 0.983$
$F^*_{\alpha=0.5}(G_t, D_{\{1,2\}}) = 1$

$FoM(G_t, D_1) = 0.22$
$FoM(G_t, D_2) = 0.60$
$H(G_t, D_1) = 1.41$
$H(G_t, D_2) = 7.67$
$\Delta_w^k(G_t, D_1) = 0.96$
$\Delta_w^k(G_t, D_2) = 2.31$

Fig. 1. Evaluations issued of a confusion matrix can be the same for different D_c. For the two candidate edge images, number of FPs and number of FNs are the same.

Table 1. List of error measures, $k = 1$ or $k = 2$ are the most common values.

Error measure name	Formulation	Parameters
Performance measure [11]	$P_m^*(G_t, D_c) = 1 - \frac{TP}{TP+FP+FN}$	None
Complemented Φ measure [12]	$\Phi^*(G_t, D_c) = 1 - \frac{TPR \cdot TN}{TN+FP}$	None
Complemented χ^2 measure [13]	$\chi^{2*}(G_t, D_c) = 1 - \frac{TPR-TP-FP}{1-TP-FP} \cdot \frac{TP+FP+FPR}{TP+FP}$	None
Complemented F_α measure [9]	$F_\alpha^*(G_t, D_c) = 1 - \frac{PREC \cdot TPR}{\alpha \cdot TPR + (1-\alpha) \cdot PREC}$, with $PREC = \frac{TP}{TP+FP}$	$\alpha \in]0;1]$
Pratt's FoM [14]	$FoM(G_t, D_c) = 1 - \frac{1}{\max(\lvert G_t \rvert, \lvert D_c \rvert)} \cdot \sum_{p \in D_c} \frac{1}{1+\kappa \cdot d_{G_t}^2(p)}$	$\kappa \in]0;1]$
FoM revisited [15]	$F(G_t, D_c) = 1 - \frac{1}{\lvert G_t \rvert + \beta \cdot FP} \cdot \sum_{p \in G_t} \frac{1}{1+\kappa \cdot d_{D_c}^2(p)}$	$\kappa \in]0;1]$ and $\beta \in \mathbb{R}^+$
Combination of *FoM* and statistics [16]	$d_4(G_t, D_c) = \frac{1}{2} \cdot \sqrt{\frac{(TP - \max(\lvert G_t \rvert, \lvert D_c \rvert))^2 + FN^2 + FP^2}{(\max(\lvert G_t \rvert, \lvert D_c \rvert))^2} + FoM(G_t, D_c)}$	$\kappa \in]0;1]$ and $\beta \in \mathbb{R}^+$
Yasnoff measure [17]	$\Upsilon(G_t, D_c) = \frac{100}{\lvert I \rvert} \cdot \sqrt{\sum_{p \in D_c} d_{G_t}^2(p)}$	None
Hausdorff distance [18]	$H(G_t, D_c) = \max\left(\max_{p \in D_c}(d_{G_t}(p)), \max_{p \in G_t}(d_{D_c}(p))\right)$	None
Maximum distance [2]	$f_2 d_6(G_t, D_c) = \max\left(\frac{1}{\lvert D_c \rvert} \cdot \sum_{p \in D_c} d_{G_t}(p), \frac{1}{\lvert G_t \rvert} \cdot \sum_{p \in G_t} d_{D_c}(p)\right)$	None
Distance to G_t [19][4]	$D^k(G_t, D_c) = \frac{1}{\lvert D_c \rvert} \cdot \sqrt[k]{\sum_{p \in D_c} d_{G_t}^k(p)}$, $k = 1$ for [19]	$k \in \mathbb{R}^+$
Oversegmentation [20][21]	$\Theta(G_t, D_c) = \frac{1}{FP} \cdot \sum_{p \in D_c} \left(\frac{d_{G_t}(p)}{\delta_{TH}}\right)^k$, $k = \delta_{TH} = 1$ for [20]	for [21]: $k \in \mathbb{R}^+$ and $\delta_{TH} \in {\mathbb{R}_*}^+$
Undersegmentation [20][21]	$\Omega(G_t, D_c) = \frac{1}{FN} \cdot \sum_{p \in G_t} \left(\frac{d_{D_c}(p)}{\delta_{TH}}\right)^k$, $k = \delta_{TH} = 1$ for [20]	for [21]: $k \in \mathbb{R}^+$ and $\delta_{TH} \in {\mathbb{R}_*}^+$
Symmetric distance [2][4]	$S^k(G_t, D_c) = \sqrt[k]{\frac{\sum_{p \in D_c} d_{G_t}^k(p)) + \sum_{p \in G_t} d_{D_c}^k(p)}{\lvert D_c \cup G_t \rvert}}$, $k = 1$ for [2]	$k \in \mathbb{R}^+$
Baddeley's Delta Metric [22]	$\Delta^k(G_t, D_c) = \sqrt[k]{\frac{1}{\lvert I \rvert} \cdot \sum_{p \in I} \lvert w(d_{G_t}(p)) - w(d_{D_c}(p)) \rvert^k}$	$k \in \mathbb{R}^+$ and a convex function $w : \mathbb{R} \mapsto \mathbb{R}$
Edge map quality measure [23]	$D_p(G_t, D_c) = \frac{1/2}{\lvert I \rvert - \lvert G_t \rvert} \cdot \sum_{p \in D_c} \left(1 - \frac{1}{1+\alpha \cdot d_{G_t}^2(p)}\right) + \frac{1/2}{\lvert G_t \rvert} \cdot \sum_{p \in G_t} \left(1 - \frac{1}{1+\alpha \cdot d_{G_t \cap D_c}^2(p)}\right)$	$\alpha \in]0;1]$
Magnier *et al.* measure [24]	$\Gamma(G_t, D_c) = \frac{FP+FN}{\lvert G_t \rvert^2} \cdot \sqrt{\sum_{p \in D_c} d_{G_t}^2(p)}$	None

P_m^*, Φ, χ^2 and F_α measures are normalized and decrease with the quality of the detection; a score equal to 0 qualifies a perfect segmentation. These measures evaluate the comparison of two edge images, pixel per pixel, tending to penalize severely a misplaced contour (even weak). So they do not indicate significant variations of the desired contour shapes through an evaluation (as illustrated in Fig. 1). As this penalization tends to be too severe, some evaluations issued from the confusion matrix recommend a spatial tolerance, particularly for assimilation of TPs [8] [9]. This inclusion could be carried by a distance threshold or a dilation of D_c and/or G_t. A such strategy of assimilation leads to counting several near contours as parallel stripes to the desired boundary. Tolerating a distance from the true contour and integrating several TPs for one detected contour are opposite to the principle of unicity in edge detection expressed by the 3rd Canny criteria: an optimal edge detector must produce a single response for one contour [10]. Finally, to perform an edge evaluation, the assessment should penalize a misplaced edge point proportionally to the distance to its true location.

2.2 Assessment involving distances of misplaced pixels

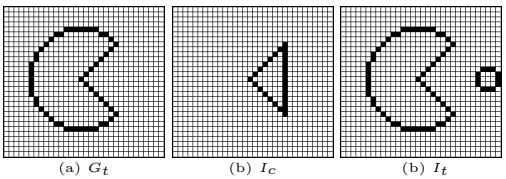
Existing quality measures involving distances: A reference-based edge map quality measure requires that a displaced edge expects to be penalized in function not only of the FPs and/or FNs but also in function of the distance to the position they should be located at. Table 1 reviews the most relevant existing measures. The common feature between these evaluators corresponds to the error distance $d_{G_t}(p)$ or $d_{D_c}(p)$. Indeed, for a pixel $p \in D_c$, $d_{G_t}(p)$ represents the minimal distance between p and G_t, whereas if $p \in G_t$, $d_{D_c}(p)$ corresponds to the minimal distance between p and D_c. This distance refers to the Euclidean distance, even though some authors involve others, see [4]. Thus, a measure computing an error distance only in function of d_{G_t} estimates the divergence of FPs, which corresponds to an over-segmentation (cases Υ, D^k, Θ, FoM and Γ). On the contrary, the sole use of a distance d_{D_c} enables an estimation of the FNs divergence, representing an under-segmentation, as Ω distance measure. A measure widely computed in matching techniques is represented by the Hausdorff distance H which estimates the mismatch of two sets of points [18]. This max-min distance could be strongly deviated by only one pixel which can be positioned sufficiently far from the pattern (illustrated in Fig. 2); so the measured distance becomes that between the pattern and the (erroneous) point, disturbing in that case the score of H. To improve H such that it becomes less sensitive to outliers, an idea is to compute H with a proportion of the maximum distances (for example 5% of the values [18]); let us note $H_{n\%}$ this measurement for $n\%$ of values ($n \in \mathbb{R}_*^+$). One of the most popular descriptor corresponds to Figure of Merit (FoM). This distance measure ranges from 0 to 1, where 0 corresponds to a perfect segmentation [14], but computes only distances of the FPs [22]. Thus, some improvements have been developed as F and d_4. Furthermore, as concluded in [3], a complete and optimum edge detection evaluation measure should combine assessments of both over- and under-segmentation, as in S^k, Δ_w^k and D_p. As an example, inspired by $f_2 d_6$ [2], another way is to consider the combination of both $FoM(G_t, D_c)$ and $FoM(D_c, G_t)$, as the two following formulas:

$$\text{Symmetric } FoM\text{: } SFoM(G_t, D_c) = \frac{1}{2} \cdot FoM(G_t, D_c) + \frac{1}{2} \cdot FoM(D_c, G_t) \quad (1)$$

$$\text{Maximum } FoM\text{: } MFoM(G_t, D_c) = \max\left(FoM(G_t, D_c), FoM(D_c, G_t)\right). \quad (2)$$

Finally, $SFoM$ and $MFoM$ take into account both distances of FNs (i.e. d_{D_c}) and FPs (i.e. d_{G_t}), so they can compute a global evaluation of a contour image.

Another way to compute a global measure is presented in [23] with the normalized edge map quality measure D_p. In fact, this distance measure seems



(a) G_t (b) I_c (b) I_t

Measure	I_t	I_c
FoM	0.63	0.17
F	0.72	0.18
d_4	0.62	0.15
$SFoM$	0.64	0.17
$MFoM$	0.82	0.58
D_P	0.33	0.007
Υ	1.14	3.30
H	109	130
$H_{5\%}$	10.39	11.01

Measure	I_t	I_c
$D^k_{k=2}$	0.39	0.47
$\theta_{\delta_{TH}=1}$	2.74	9.07
$\theta_{\delta_{TH}=5}$	0.37	3.35
$\Omega_{\delta_{TH}=1}$	7.91	0
Δ^k	6.05	1.20
$f2d6$	5.79	1.60
$S^k_{k=2}$	5.97	2.84
Γ	0.19	0.12
Ψ	0.94	0.12

Fig. 2. The scores of the over-segmentation evaluations are higher for I_t whereas I_t is more closer visually to G_t than I_c.

similar to $SFoM$ with different coefficients. However, both the left and the right terms are composed of a $\frac{1}{2}$ coefficient, so in the presence of only a pure under- or over-segmentation, the score of D_p does not attain over $\frac{1}{2}$.

A new edge detection assessment measure: In [24] is developed a normalized measure of the edge detection assessment, denoted Γ. This function represents an over-segmentation measure which depends also of FN and FP. As this measure is not sufficiently efficient concerning FNs because it does not consider d_{D_c} for false negative points (see Fig. 7). Thus, inspired by S^k, the new measure Ψ holds different coefficients changing the behavior of the measure:

$$\Psi(G_t, D_c) = \frac{FP + FN}{|G_t|^2} \cdot \sqrt{\sum_{p \in G_t} d^2_{D_c}(p) + \sum_{p \in D_c} d^2_{G_t}(p)} \tag{3}$$

Compared to Γ, Ψ improves the measurement by combining both d_{G_t} and d_{D_c} (illustrated in Fig. 7). Authors of Γ have studied the influence of the coefficient in different concrete cases [24]. They concluded that such a formulation must take into consideration all observable cases and theoretically observable. In fact, a performing measure has to take into account all the following input parameters $|G_t|$, FN and FP whereas the image dimensions should not be considered. Thus, the parameter $\frac{FP+FN}{|G_t|^2}$ seems a good compromise and has been introduced to the new formula of assessment Ψ.

2.3 Normalization of the edge detection evaluation:

In order to compare each boundary detection assessments, all measures must be normalized, but also indicate the same information: an error measure close to 1 means a poor segmentation whereas a value close to 0 indicates a good segmentation. Thereby, the values of FoM, F, d_4 and D_p belongs to $[0, 1]$. However, concerning other distance measures of table 1, a normalization is required. Introduced in [24], a formula called Key Performance Indicator (KPI) gives value close to 1 for a poor segmentation; alternatively, a KPI value close to 0 translates a good segmentation:

$$\begin{aligned} KPI_u : [0; \infty[&\mapsto [0; 1[\\ u &\to 1 - \frac{1}{1 + u^h}. \end{aligned} \tag{4}$$

where the parameter u represents a distance error and h a constant ($h \in \mathbb{R}^+_*$).

An undeniable parameter of KPI formula is the power of the denominator term called h. Inasmuch as KPI depends on its value, it evolves more or less quickly around 0.5 and embodies a range of observable cases. The advice to choose values between 1 and 2 can be easily checked. Otherwise, the more KPI evolution will be abrupt, the less the transition between 0.5 and 1 will be marked. As a compromise, fixing the power at the *golden ratio* $\phi \simeq 1.6180339887$, the measurement becomes not too strong in the presence a small measure score, but increases to 1 for a high score of the distance measure, see [24].

3 EXPERIMENTAL RESULTS

Experiments realized in this part aim to be the most accomplished, thus the more close and realistic of the reality. In respect of these directives, in a first

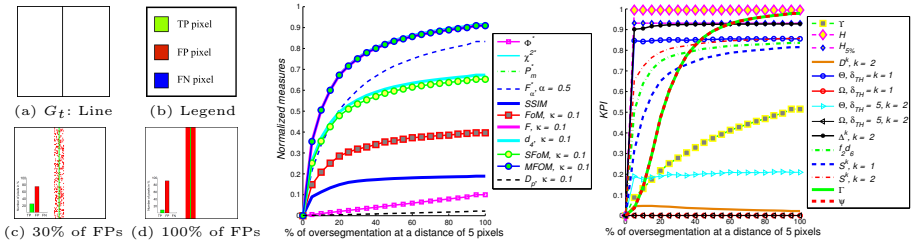


Fig. 3. Measures scores in function of the over-segmentation in the contour area.

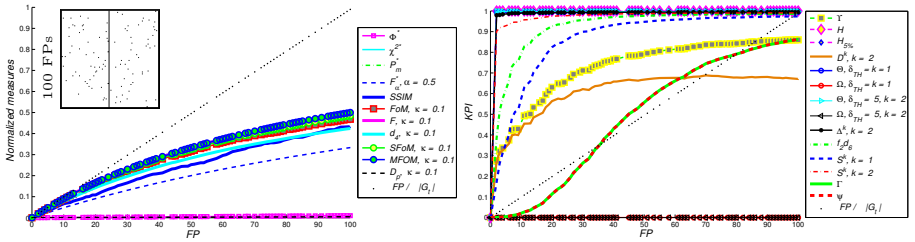


Fig. 4. Evolution of the dissimilarity measures in function of the FPs addition.

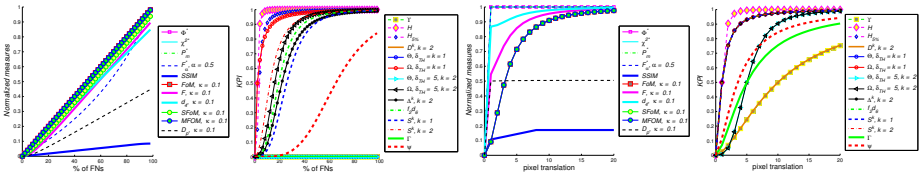


Fig. 5. Measure scores in function of the FNs addition and the edge translation.

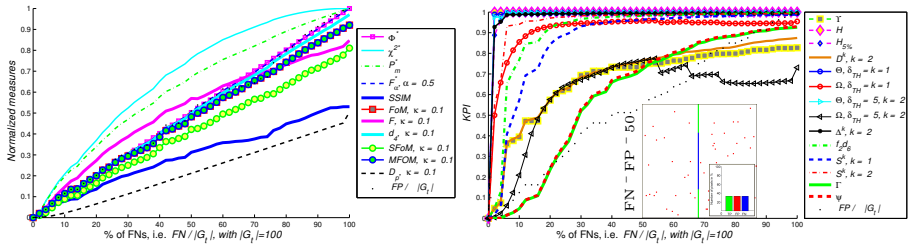


Fig. 6. Dissimilarity measure scores in function of addition of both the FNs and FPs.

time, considering a synthetic edge model (i.e. ground truth) the edge detection evaluation measures are subject to the following studies: addition of false positive points close to the true contour, addition of false negative points (under-segmentation), addition of false positive points (over-segmentation), addition of both false negatives and false positive points, translation of the boundary. Thus, 24 measures and the new proposed method are tested and compared together. The KPI in eq. 4 is computed for the non-normalized algorithms in Table 1.

The first experiment is to create an over-segmentation at a maximal distance of 5 pixels, as illustrated in Fig. 3 (100% of over-segmentation represents a dilatation of the vertical line with a structural element of size 1×6, corresponding of a total saturation of the contour, see Fig. 3(d)). Curves presented here show

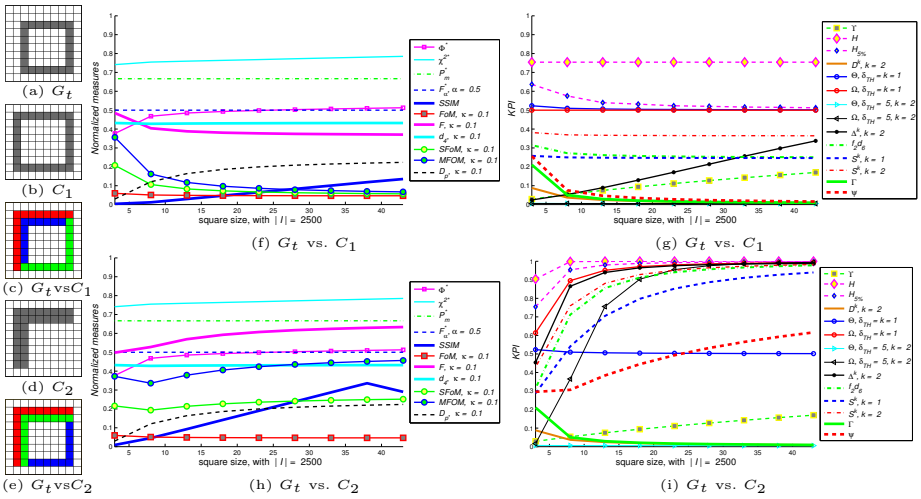


Fig. 7. Dissimilarity measure scores in function of the size of the original square.

that F_α, d_4, F, $MFoM$, H, $H_{5\%}$, Δ_k, D^k, f_2d_6, and S^k are very sensitive to FPs, whereas $SSIM$, D_p, Φ and D^k (which is not homogeneous) do not penalize enough D_c. Ω remains constant at 0 because it corresponds to an under-segmentation measure. Moreover, Υ and FoM are not too abrupt, even though they stagnate, like D^k, $SFoM$, f_2d_6, and S^k. Finally, Γ and Ψ are not too abrupt and penalize strongly D_c in the presence of many FPs.

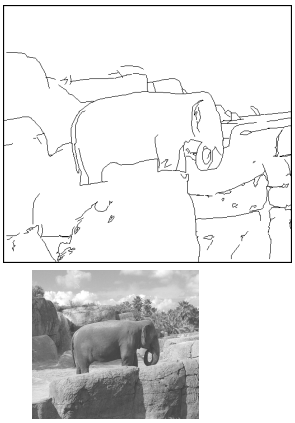
The second test is to add random undesirable pixels to G_t until 100 FPs, as represented in Fig. 4 top left. Globally, the curves in Fig. 4 illustrates that the measures using KPI behave like the previous experiment; only Γ and Ψ are not too sensitive to FPs. The normalized evaluation measures increase correctly, but seem stagnant, excepted Φ and D_p which stay close to 0.

Concerning the addition of FNs, Fig. 5 (left) illustrates that H, $H_{5\%}$ and $\Omega_{\delta_{TH}=1}$ are very sensitive to the presence of FNs. Also, D_p attains only the score of $\frac{1}{2}$. The over-segmentation methods Υ, D^k, Ω and Γ remain constant at 0. On the other hand, the score of the KPI of Ψ attains 0.5 for 50% of FNs. Afterward, contrary to the addition of FPs or FNs, error measures without distance measures obtains a score of 1 after one pixel of translation and the score of D_p stays constant at $\frac{1}{2}$ (Fig. 5 (right)). Only FoM, $SFoM$, $MFoM$, the KPI of $\Omega_{\delta_{TH}=5}$, the KPI of Γ and the KPI of Ψ behave correctly.

Concerning the line, the last experiment corresponds to an addition of both FPs and FNs. Thus, Fig. 6 shows that the normalized measures, excepted D_p and $SSIM$ behave correctly. Concerning other measures, the KPI scores of $\Omega_{\delta_{TH}=5}$, Γ and Ψ are not too abrupt for few number or errors and penalize strongly D_c in the presence of many FPs and FNs (but $\Omega_{\delta_{TH}=5}$ is not homogeneous). For example, Fig. 6 (bottom right) illustrates the line where both 50% of the pixels are missing and 50 FPs are added, corresponding to 33% of TPs. In this precise case, the KPI score of new measure Ψ is close to 0.7, thus, reflecting the reality.

Another experiment in Fig. 7, two different shapes are compared to a square (G_t), illustrating the importance to consider both d_{D_c} and d_{G_t}. Furthermore, all the shapes are growing at the same time, keeping the same percentage of FPs and FNs with G_t. The more G_t grows, the more C_1 is visually closer to G_t whereas FNs deviate strongly in the case of C_2. Despite these tow different evolutions, statistical measures, FoM, F, d_4, D_p, Υ and Γ obtain close the same measurements for C_1 and C_2. On the contrary, the KPI of Ψ grows around 0.5 for C_2, whereas it converges towards 0 for C_1, since C_1 becomes visually closer to G_t with the enlargement (note that $MFoM$ behaves identically).

To conclude experimental evaluations, Table 2 reports different assessments for five edge detection methods on a real image: Sobel, Canny [10], Steerable Filters of order 1 (SF_1) [25], Steerable Filters of order 5 (SF_5) [26] and Half Gaussian Kernels (HK) [6]. Even though the problem of hand-made ground



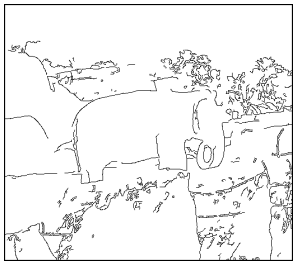
(a) Gt / Original [1]

Measure	Score	Measure	Score	Measure	Score
Φ^*	0.519	$SFoM$	0.702	$\theta_{\delta_{TH}=5}$	0.993
χ^{2*}	0.946	$MFoM$	0.814	$\Omega_{\delta_{TH}=1}$	0.587
P_m^*	0.883	D_P	0.117	Δ^k	0.995
F_α^*	0.790	Υ	0.618	$f_2 d_6$	0.987
$SSIM$	0.355	H	0.999	$S_{k=1}^k$	0.982
FoM	0.629	$H_{5\%}$	0.999	$S_{k=2}^k$	0.992
F	0.800	$D_{k=2}^k$	0.049	Γ	0.812
d_4	0.690	$\theta_{\delta_{TH}=1}$	0.989	Ψ	0.814

(b) Sobel

Measure	Score	Measure	Score	Measure	Score
Φ^*	0.563	$SFoM$	0.601	$\theta_{\delta_{TH}=5}$	0.995
χ^{2*}	0.928	$MFoM$	0.770	$\Omega_{\delta_{TH}=1}$	0.799
P_m^*	0.847	D_P	0.139	Δ^k	0.995
F_α^*	0.775	Υ	0.527	$f_2 d_6$	0.986
$SSIM$	0.282	H	0.999	$S_{k=1}^k$	0.978
FoM	0.540	$H_{5\%}$	0.999	$S_{k=2}^k$	0.992
F	0.723	$D_{k=2}^k$	0.0714	Γ	0.597
d_4	0.645	$\theta_{\delta_{TH}=1}$	0.989	Ψ	0.602

(c) Canny [10]



Measure	Score	Measure	Score	Measure	Score
Φ^*	0.571	$SFoM$	0.574	$\theta_{\delta_{TH}=5}$	0.995
χ^{2*}	0.920	$MFoM$	0.762	$\Omega_{\delta_{TH}=1}$	0.814
P_m^*	0.835	D_P	0.143	Δ^k	0.995
F_α^*	0.717	Υ	0.509	$f_2 d_6$	0.986
$SSIM$	0.274	H	0.999	$S_{k=1}^k$	0.977
FoM	0.523	$H_{5\%}$	0.999	$S_{k=2}^k$	0.992
F	0.701	$D_{k=2}^k$	0.081	Γ	0.532
d_4	0.632	$\theta_{\delta_{TH}=1}$	0.990	Ψ	0.539

(d) Steerable filters [25]

Measure	Score	Measure	Score	Measure	Score
Φ^*	0.625	$SFoM$	0.468	$\theta_{\delta_{TH}=5}$	0.996
χ^{2*}	0.907	$MFoM$	0.715	$\Omega_{\delta_{TH}=1}$	0.979
P_m^*	0.812	D_P	0.187	Δ^k	0.995
F_α^*	0.683	Υ	0.421	$f_2 d_6$	0.983
$SSIM$	0.219	H	0.999	$S_{k=1}^k$	0.974
FoM	0.431	$H_{5\%}$	0.999	$S_{k=2}^k$	0.992
F	0.654	$D_{k=2}^k$	0.108	Γ	0.326
d_4	0.599	$\theta_{\delta_{TH}=1}$	0.989	Ψ	0.363

(e) Steerable filters [26]

Measure	Score	Measure	Score	Measure	Score
Φ^*	0.669	$SFoM$	0.378	$\theta_{\delta_{TH}=5}$	0.995
χ^{2*}	0.903	$MFoM$	0.686	$\Omega_{\delta_{TH}=1}$	0.986
P_m^*	0.805	D_P	0.211	Δ^k	0.996
F_α^*	0.673	Υ	0.338	$f_2 d_6$	0.978
$SSIM$	0.198	H	0.999	$S_{k=1}^k$	0.970
FoM	0.371	$H_{5\%}$	0.999	$S_{k=2}^k$	0.991
F	0.622	$D_{k=2}^k$	0.120	Γ	0.203
d_4	0.604	$\theta_{\delta_{TH}=1}$	0.988	Ψ	0.260

(f) Half Gaussian filters [6]

Table 2. Comparison measures of different edge detections. A score close to 0 indicates a good edge map whereas a score 1 translates a poor segmentation.

truth on real images is discussed by some researchers [27], only the comparison of D_c with a G_t is studied here. Compared to G_t (Table 2 (a)), the well known Sobel edge detector generates more FPs than the other three methods while SF_5 and HK are less sensitive to noise or texture. Furthermore, HK captures easier straight contours and corners closest to their true positions [6]. So, the measurements in tables of Table 2 must be close to 1 for Sobel and a little less for Canny, but decrease with reasonable error for HK (scores involving KPI for non-normalized algorithms, eq. 4). Thus, Γ and Ψ respect this evolution and indicate a good measurement value. FoM, F, d_4, $SFoM$, $MFoM$ and Υ evolve similarly, but the score for the HK remains too elevated. Also, Φ^*, χ^{2*}, P_m^*, F_α^* and D_p do not indicate a significant difference between all the segmentations. Further, other non normalized methods are not adapted to give a score between 0 and 1 using a KPI. Eventually, given the segmented images, Γ and Ψ indicate a good measurement value. Other results involving other edge images are available on the website: `http://hkaljaf.wixsite.com/hasanabdulrahman/edge-detection-and-evaluation.`

4 CONCLUSION AND FUTURE WORKS

In this paper, several referenced-based boundary detection evaluations are detailed, pointing their advantages and disadvantages through concrete examples of edge images. A new normalized supervised edge map quality measure is proposed, comparing a ground truth contour image, the candidate contour image and their associated spacial nearness. The strategy to normalize the evaluation enables to consider a score close to 0 as a good edge map, whereas a score 1 translates a poor segmentation. Eventually, compared to other edge evaluation assessments, the score of the new evaluation indicates confidently the quality of a segmentation. Next on our work program agenda is to compare different edge detectors with their different parameters and binary image matching.

Acknowledgements

The authors wish to thank the Iraqi Ministry of Higher Education and Scientific Research for funding and supporting this work.

References

1. M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," *IEEE TPAMI*, vol. 19, no. 12, pp. 1338–1359, 1997.
2. M-P Dubuisson and Anil K Jain, "A modified hausdorff distance for object matching," in *IEEE ICPR*, 1994, vol. 1, pp. 566–568.
3. S. Chabrier, H. Laurent, C. Rosenberger, and B. Emile, "Comparative study of contour detection evaluation criteria based on dissimilarity measures," *EURASIP J. on Image and Video Processing*, vol. 2008, pp. 2, 2008.
4. C. Lopez-Molina, B. De Baets, and H. Bustince, "Quantitative error measures for edge detection," *Patt. Rec.*, vol. 46, no. 4, pp. 1125–1139, 2013.

5. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE TIP*, vol. 13, no. 4, pp. 600–612, 2004.
6. B. Magnier, P. Montesinos, and D. Diep, "Fast anisotropic edge detection using gamma correction in color images," in *IEEE ISPA*, 2011, pp. 212–217.
7. B. Magnier, A. Aberkane, P. Borianne, P. Montesinos, and C. Jourdan, "Multiscale crest line extraction based on half gaussian kernels," in *IEEE ICASSP*, 2014, pp. 5105–5109.
8. K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical roc curves," in *CVIU*, 2001, pp. 77–103.
9. D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE TPAMI*, vol. 26, no. 5, pp. 530–549, 2004.
10. J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, , no. 6, pp. 679–698, 1986.
11. C. Grigorescu, N. Petkov, and M. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *IEEE TIP*, vol. 12, no. 7, pp. 729–739, 2003.
12. S. Venkatesh and P. L. Rosin, "Dynamic threshold determination by local and global edge evaluation," *CVGIP*, vol. 57, no. 2, pp. 146–160, 1995.
13. Y. Yitzhaky and E. Peli, "A method for objective edge detection evaluation and detector parameter selection," *IEEE TPAMI*, vol. 25, no. 8, pp. 1027–1033, 2003.
14. I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proc. of the IEEE*, vol. 67, pp. 753–763, 1979.
15. A. J. Pinho and L. B. Almeida, "Edge detection filters based on artificial neural networks," in *ICIAP*. Springer, 1995, pp. 159–164.
16. A. G. Boaventura and A. Gonzaga, "Method to evaluate the performance of edge detector," 2009.
17. W.A. Yasnoff, W. Galbraith, and J.W. Bacus, "Error measures for objective assessment of scene segmentation algorithms.," *Analytical and Quantitative Cytology*, vol. 1, no. 2, pp. 107–121, 1978.
18. D.P. Huttenlocher and W.J. Rucklidge, "A multi-resolution technique for comparing images using the hausdorff distance," in *IEEE CVPR*, 1993, pp. 705–706.
19. T. Peli and D. Malah, "A study of edge detection algorithms," *CGIP*, vol. 20, no. 1, pp. 1–21, 1982.
20. R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE TPAMI*, vol. 6, no. 1, pp. 58–68, 1984.
21. C. Odet, B. Belaroussi, and H. Benoit-Cattin, "Scalable discrepancy measures for segmentation evaluation," in *IEEE ICIP*, 2002, vol. 1, pp. 785–788.
22. A. J. Baddeley, "An error metric for binary images," *Robust Computer Vision: Quality of Vision Algorithms*, pp. 59–78, 1992.
23. K. Panetta, C. Gao, S. Agaian, and S. Nercessian, "A new reference-based edge map quality measure," *IEEE Trans. on Systems Man and Cybernetics: Systems*, vol. 46, no. 11, pp. 1505–1517, 2016.
24. B. Magnier, A. Le, and A. Zogo, "A quantitative error measure for the evaluation of roof edge detectors," in *IEEE IST*, 2016, pp. 429–434.
25. W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE TPAMI*, vol. 13, pp. 891–906, 1991.
26. M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE TPAMI*, vol. 26, no. 8, pp. 1007–1019, 2004.
27. X. Hou, A. Yuille, and C. Koch, "Boundary detection benchmarking: Beyond f-measures," in *IEEE CVPR*, 2013, pp. 2123–2130.