HAL
open science

A New Objective Supervised Edge Detection Assessment Using Hysteresis Thresholds

Hasan Abdulrahman, Baptiste Magnier, Philippe Montesinos

To cite this version:

Hasan Abdulrahman, Baptiste Magnier, Philippe Montesinos. A New Objective Supervised Edge Detection Assessment Using Hysteresis Thresholds. ICIAP 2017 - International Conference on Image Analysis and Processing, Sep 2017, Catania, Italy. pp.3-14, 10.1007/978-3-319-70742-6_1 . hal-01940328

HAL Id: hal-01940328

https://hal.science/hal-01940328v1

Submitted on 10 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Objective Supervised Edge Detection Assessment using Hysteresis Thresholds

Hasan Abdulrahman, Baptiste Magnier, and Philippe Montesinos

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

Abstract. Useful for the visual perception of a human, edge detection remains a crucial stage in numerous image processing applications. Therefore, one of the most challenging goals in contour extraction is to operate algorithms that can process visual information as humans need. Hence, to ensure that it is reliable, an edge detection technique needs to be severely assessed before being used it in a computer vision tools. To achieve this task, a supervised evaluation computes a score between a ground truth edge map and a candidate image. Theoretically, by varying the hysteresis thresholds of the thin edges, the minimum score of the measure corresponds to the best edge map, compared to the ground truth. In this study, a new supervised edge map quality measure is proposed, where the minimum score of the measure is associated with an edge map in which the main structures of the desired objects are distinctive.

Keywords: edge detection, supervised evaluation, hysteresis.

1 Introduction on edge detection and thresholding

Edge detection is an important field in image processing because this process frequently attempts to capture the most important structures in the image. Hence, edge detection represents a fundamental step concerning computer vision approaches. Furthermore, edge detection itself could be used to qualify a region segmentation technique. Additionally, the edge detection assessment remains very useful in image segmentation, registration, reconstruction or interpretation. Hence, it is hard to design an edge detector which is able to extract the exact edge with good localization and orientation from an image. In the literature, different techniques have emerged and, due to its importance, edge detection continues to be an active research area [1]. The best-known and useful edge detection methods are based on gradient computing first-order fixed operators [2, 3]. Oriented operators compute the maximum energy in an orientation [4–6] or two directions [7]. Typically, these methods are composed of three steps:

1. Computation of the gradient magnitude and its orientation η, see Fig. 1.
2. Non-maximum suppression to obtain thin edges: the selected pixels are those having gradient magnitude at a local maximum along the gradient direction η which is perpendicular to the edge orientation.
3. Thresholding of the thin contours to obtain an edge map.

Thus, Fig. 1 exposes the different possibilities of gradient and its associated orientations involving several edge detection algorithms compared in this paper.

The final step remains a difficult stage in image processing, however it represents a crucial operation to compare several segmentation algorithms. In edge detection, the hysteresis process uses the connectivity information of the pixels belonging to thin contours and thus remains a more elaborated method than binary thresholding. Simply, this technique determines a contour image that has been thresholded at different levels (low: τ_L and high: τ_H). The low threshold τ_L determines which pixels are considered as edge points if at least one point higher than τ_H exists in a contour chain where all the pixel values are also higher than τ_L, as represented with a signal in Fig. 1. Thus, the lower the thresholds are, the more the undesirable pixels are preserved.

Usually, in order to compare several edge detection methods, the user has to try some thresholds to select the ones that appear visually as the best edge maps in quality. However, this assessment suffers from a main drawback: segmentations are compared using the threshold (deliberately) chosen by the user, this evaluation is very subjective and not reproducible. Hence, the purpose is to use the dissimilarity measures without any user intervention for an objective assessment. Finally, to consider a valuable edge detection assessment, the evaluation process should produce a result that correlates with the perceived quality of the edge image, which relies on human judgment [9, 8, 10]. In other words, a reliable edge map should characterize all the relevant structures of an image as closely as possible, without any disappearance of desired contours. Nevertheless, a minimum of spurious pixels can be created by the edge detector, disturbing at the same time the visibility of the main/desired objects to detect.

In this paper, a novel technique is presented to compare edge detection techniques by using hysteresis thresholds in a supervised way, being consistent with the visual perception of a human. Indeed, by comparing a ground truth contour map with an ideal edge map, several assessments can be compared by varying the parameters of the hysteresis thresholds. This study shows the importance to penalize stronger the false negative points, compared to the false positive points, leading to a new edge detection evaluation algorithm. The experiment using synthetic and real images demonstrated that the proposed method obtains contours maps closer to the ground truth without requiring tuning parameters and outperforms other assessment methods in an objective way.

Type of operator	Fixed operator [2, 3]	Oriented Filters [4–6]	Half Gaussian Kernels [7]
Gradient magnitude	$\lvert\nabla I\rvert = \sqrt{I_0^2 + I_{\pi/2}^2}$	$\lvert\nabla I\rvert = \max\limits_{\theta\in[0,\pi[} \lvert I_\theta\rvert$	$\lvert\nabla I\rvert = \max\limits_{\theta\in[0,2\pi[} I_\theta - \min\limits_{\theta\in[0,2\pi[} I_\theta$
Gradient direction	$\eta = \arctan\left(\frac{I_{\pi/2}}{I_0}\right)$	$\eta = \arg\max\limits_{\theta\in[0,\pi[} \lvert I_\theta\rvert + \frac{\pi}{2}$	$\eta = \left(\arg\max\limits_{\theta\in[0,2\pi[} I_\theta + \arg\min\limits_{\theta\in[0,2\pi[} I_\theta\right)/2$

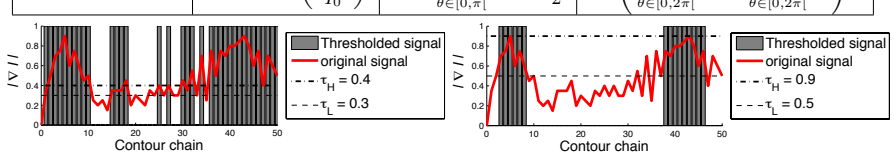


Fig. 1. Gradient magnitude and orientation computation for a scalar image I and example of hysteresis threshold applied along a contour chain. I_θ represents the image derivative using a first-order filter at the θ orientation (in radians).

2 Supervised Measures for Image Contour Evaluations

A supervised evaluation criterion computes a dissimilarity measure between a segmentation result and a ground truth obtained from synthetic data or an expert judgment (i.e. manual segmentation) [11][12][13][14]. In this paper, the closer to 0 the score of the evaluation is, the more the segmentation is qualified as good. This work focusses on comparisons of supervised edge detection evaluations and proposes a new measure, aiming at an objective assessment.

2.1 Error measures involving only statistics

To assess an edge detector, the confusion matrix remains a cornerstone in boundary detection evaluation methods. Let G_t be the reference contour map corresponding to ground truth and D_c the detected contour map of an original image I. Comparing pixel per pixel G_t and D_c, the 1st criterion to be assessed is the common presence of edge/non-edge points. A basic evaluation is composed of statistics; to that end, G_t and D_c are combined. Afterwards, denoting $|\cdot|$ as the cardinality of a set, all points are divided into four sets (see Fig. 3):

- 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: $TN = |\neg D_c \cap \neg G_t|$.

Several edge detection evaluations involving confusion matrix are presented in Table 1. Computing only FPs and FNs [7] or their sum enables a segmentation assessment to be performed. The complemented *Performance measure* P_m^* considers directly and simultaneously the three entities TP, FP and FN to assess a binary image and decreases with improved quality of detection.

Another way to display evaluations is to create Receiver Operating Characteristic (ROC) [19] curves or Precision-Recall (PR) [18], involving *True Positive Rates* (TPR) and *False Positive Rates* (FPR): $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$. Derived from TPR and FPR, the three measures Φ, χ^2 and F_α (detailed in Table 1) are frequently used. The complement of these measures enables to translate a value close to 0 as a good segmentation.

These measures evaluate the comparison of two edge images, pixel per pixel, tending to severely penalize a (even slightly) misplaced contour, as illustrated in Fig. 2. Consequently, some evaluations resulting from the confusion matrix recommend incorporating spatial tolerance. Tolerating a distance from the true contour and integrating several TPs for one detected contour can penalize efficient edge detection methods, or, on the contrary, advantage poor ones (especially for corners or small objects). Thus, from the discussion below, the assessment should penalize a misplaced edge point proportionally to the distance from its true location (some examples in [14], and, as shown in Fig. 2).

Table 1. List of error measures involving only statistics.

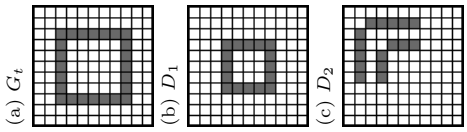
Performance measure [15]	$P_m^*(G_t, D_c) = 1 - \frac{TP}{\lvert G_t \cup D_c \rvert} = 1 - \frac{TP}{TP+FP+FN}$
Complemented Φ measure [16]	$\Phi^*(G_t, D_c) = 1 - \frac{TPR \cdot TN}{TN+FP}$
Complemented χ^2 measure [17]	$\chi^{2*}(G_t, D_c) = 1 - \frac{TPR-TP-FP}{1-TP-FP} \cdot \frac{TP+FP+FPR}{TP+FP}$
Complemented F_α measure [18]	$F_\alpha^*(G_t, D_c) = 1 - \frac{PREC \cdot TPR}{\alpha \cdot TPR+(1-\alpha)\cdot PREC}$, with $PREC = \frac{TP}{TP+FP}$ and $\alpha \in]0;1]$

Table 2. List of error measures involving distances, generally: $k = 1$ or $k = 2$.

Error measure name	Formulation	Parameters
Pratt's FoM [20]	$FoM\left(G_t, D_c\right) = 1 - \frac{1}{\max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)} \cdot \sum_{p \in D_c} \frac{1}{1 + \kappa \cdot d^2_{G_t}(p)}$	$\kappa \in]0;1]$
FoM revisited [21]	$F\left(G_t, D_c\right) = 1 - \frac{1}{\lvert G_t\rvert + \beta \cdot FP} \cdot \sum_{p \in G_t} \frac{1}{1 + \kappa \cdot d^2_{D_c}(p)}$	$\kappa \in]0;1]$ and $\beta \in \mathbb{R}^+$
Combination of FoM and statistics [22]	$d_4\left(G_t, D_c\right) = \frac{1}{2} \cdot \sqrt{\frac{\left(TP - \max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)\right)^2 + FN^2 + FP^2}{\left(\max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)\right)^2} + FoM\left(G_t, D_c\right)}$	$\kappa \in]0;1]$ and $\beta \in \mathbb{R}^+$
Symmetric FoM [14]	$SFoM\left(G_t, D_c\right) = \frac{1}{2} \cdot FoM\left(G_t, D_c\right) + \frac{1}{2} \cdot FoM\left(D_c, G_t\right)$	$\kappa \in]0;1]$
Maximum FoM [14]	$MFoM\left(G_t, D_c\right) = \max\left(FoM\left(G_t, D_c\right), FoM\left(D_c, G_t\right)\right)$	$\kappa \in]0;1]$
Yasnoff measure [23]	$\Upsilon\left(G_t, D_c\right) = \frac{100}{\lvert I\rvert} \cdot \sqrt{\sum_{p \in D_c} d^2_{G_t}(p)}$	None
Hausdorff distance [24]	$H\left(G_t, D_c\right) = \max\left(\max_{p \in D_c}(d_{G_t}(p)), \max_{p \in G_t}(d_{D_c}(p))\right)$	None
Maximum distance [11]	$f_2 d_6\left(G_t, D_c\right) = \max\left(\frac{1}{\lvert D_c\rvert} \cdot \sum_{p \in D_c} d_{G_t}(p), \frac{1}{\lvert G_t\rvert} \cdot \sum_{p \in G_t} d_{D_c}(p)\right)$	None
Distance to G_t [25][13]	$D^k\left(G_t, D_c\right) = \frac{1}{\lvert D_c\rvert} \cdot \sqrt[k]{\sum_{p \in D_c} d^k_{G_t}(p)}, \quad k = 1$ for [25]	$k \in \mathbb{R}^+$
Oversegmentation [26]	$\Theta\left(G_t, D_c\right) = \frac{1}{FP} \cdot \sum_{p \in D_c} \left(\frac{d_{G_t}(p)}{\delta_{TH}}\right)^k$	for [26]: $k \in \mathbb{R}^+$ and $\delta_{TH} \in \mathbb{R}_*{}^+$
Undersegmentation [26]	$\Omega\left(G_t, D_c\right) = \frac{1}{FN} \cdot \sum_{p \in G_t} \left(\frac{d_{D_c}(p)}{\delta_{TH}}\right)^k$	for [26]: $k \in \mathbb{R}^+$ and $\delta_{TH} \in \mathbb{R}_*{}^+$
Symmetric distance [11][13]	$S^k\left(G_t, D_c\right) = \sqrt[k]{\frac{\sum_{p \in D_c} d^k_{G_t}(p)) + \sum_{p \in G_t} d^k_{D_c}(p)}{\lvert D_c \cup G_t\rvert}}, \quad k = 1$ for [11]	$k \in \mathbb{R}^+$
Baddeley's Delta Metric [27]	$\Delta^k(G_t, D_c) = \sqrt[k]{\frac{1}{\lvert I\rvert} \cdot \sum_{p \in I} \lvert w(d_{G_t}(p)) - w(d_{D_c}(p))\rvert^k}$	$k \in \mathbb{R}^+$ and a convex function $w : \mathbb{R} \mapsto \mathbb{R}$
Edge map quality measure [28]	$D_p\left(G_t, D_c\right) = \frac{1/2}{\lvert I\rvert - \lvert G_t\rvert} \cdot \sum_{p \in D_c} \left(1 - \frac{1}{1 + \alpha \cdot d^2_{G_t}(p)}\right) + \frac{1/2}{\lvert G_t\rvert} \cdot \sum_{p \in G_t} \left(1 - \frac{1}{1 + \alpha \cdot d^2_{G_t \cap D_c}(p)}\right)$	$\alpha \in]0;1]$
Magnier *et al.* measure [29]	$\Gamma(G_t, D_c) = \frac{FP + FN}{\lvert G_t\rvert^2} \cdot \sqrt{\sum_{p \in D_c} d^2_{G_t}(p)}$	None
Complete distance measure [14]	$\Psi(G_t, D_c) = \frac{FP + FN}{\lvert G_t\rvert^2} \cdot \sqrt{\sum_{p \in G_t} d^2_{D_c}(p) + \sum_{p \in D_c} d^2_{G_t}(p)}$	None

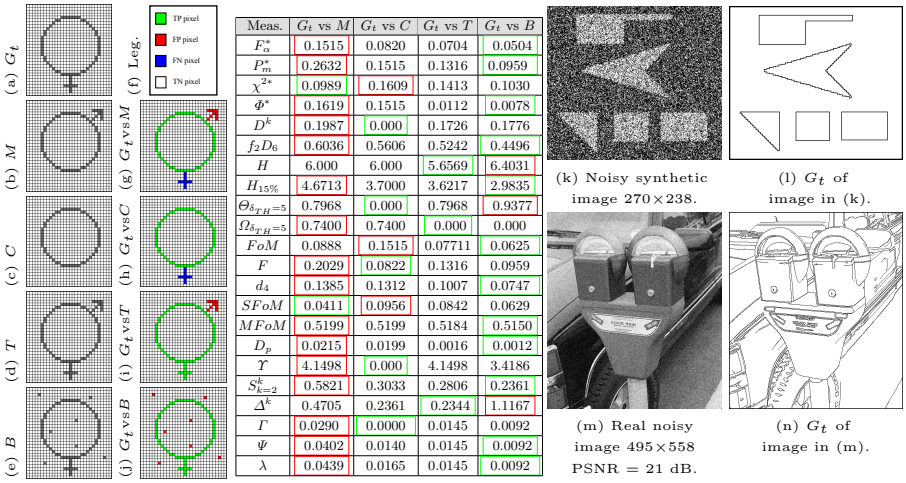
2.2 Assessment involving distances of misplaced pixels

A reference-based edge map quality measure requires that a displaced edge should be penalized in function not only of FPs and/or FNs but also of the distance from the position where it should be located. Table 2 reviews the most relevant measures involving distances. Thus, for a pixel p belonging to the desired contour D_c, $d_{G_t}(p)$ represents the minimal Euclidian distance between p and G_t. If p belongs to the ground truth G_t, $d_{D_c}(p)$ is the minimal distance between p and D_c. On the one hand, some distance measures are specified in the evaluation of over-segmentation (i.e. presence of FPs), like: Υ, D^k, Θ and Γ. On the other hand, Ω measure assesses an edge detection by computing only an under segmentation (FNs). Other edge detection evaluation measures consider both distances



$F^*_\alpha(G_t, D_1)$=1.000	$F^*_\alpha(G_t, D_2)$=1.000
$\Upsilon(G_t, D_1)$=13.223	$\Upsilon(G_t, D_2)$=13.223
$FoM((G_t, D_1)$= 0.3939	$FoM(G_t, D_2)$=0.3939
$H(G_t, D_1)$=1.4142	$H(G_t, D_2)$= 5.3852
$S^k_{k=2}(G_t, D_1)$=1.0414	$S^k_{k=2}(G_t, D_2)$=1.6993
$\lambda(G_t, D_1)$=0.4482	$\lambda(G_t, D_2)$=0.5725

Fig. 2. Different D_c: FPs and number of FNs are the same for D_1 and for D_2.



Meas.	G_t vs M	G_t vs C	G_t vs T	G_t vs B
F_α^*	0.1515	0.0820	0.0704	0.0504
P_m^*	0.2632	0.1515	0.1316	0.0959
χ^{2*}	0.0989	0.1609	0.1413	0.1030
Φ^*	0.1619	0.1515	0.0112	0.0078
D^k	0.1987	0.000	0.1726	0.1776
f_2D_6	0.6036	0.5606	0.5242	0.4496
H	6.000	6.000	5.6569	6.4031
$H_{15\%}$	4.6713	3.7000	3.6217	2.9835
$\Theta_{\delta_{TH}=5}$	0.7968	0.000	0.7968	0.9377
$\Omega_{\delta_{TH}=5}$	0.7400	0.7400	0.000	0.000
FoM	0.0888	0.1515	0.07711	0.0625
F	0.2029	0.0822	0.1316	0.0959
d_4	0.1385	0.1312	0.1007	0.0747
$SFoM$	0.0411	0.0956	0.0842	0.0629
$MFoM$	0.5199	0.5199	0.5184	0.5150
D_p	0.0215	0.0199	0.0016	0.0012
Υ	4.1498	0.000	4.1498	3.4186
$S_{k=2}^k$	0.5821	0.3033	0.2806	0.2361
Δ^k	0.4705	0.2361	0.2344	1.1167
Γ	0.0290	0.0000	0.0145	0.0092
Ψ	0.0402	0.0140	0.0145	0.0092
λ	0.0439	0.0165	0.0145	0.0092

(k) Noisy synthetic image 270×238.

(l) G_t of image in (k).

(m) Real noisy image 495×558 PSNR = 21 dB.

(n) G_t of image in (m).

Fig. 3. Results of evaluation measures and images for the experiments.

of FPs and FNs [8]. A perfect segmentation using an over-segmentation measure could be an image including no edge points and an image having most undesirable edge points (FPs) concerning under-segmentation evaluations (see Fig. 3). Also, another limitation of only over- and under-segmentation evaluations are that several binary images can produce the same result (Fig. 2). Therefore, as demonstrated in [8], a complete and optimum edge detection evaluation measure should combine assessments of both over- and under-segmentation.

Among the distance measures between two contours, one of the most popular descriptors is named the Figure of Merit (FoM). Nonetheless, for FoM, the distance of the FNs is not recorded and are strongly penalized as statistic measures (see above). For example, in Fig.3, $FoM(G_t, C) > FoM(G_t, M)$, whereas M contains both FPs and FNs and C only FNs. Further, for the extreme cases:

- if $FP = 0$: $FoM\,(G_t, D_c) = 1 - TP/|G_t| = 1 - (|G_t| - FN)/|G_t|$,
- if $FN = 0$: $FoM\,(G_t, D_c) = 1 - \frac{1}{\max(|G_t|,|D_c|)} \cdot \sum_{p \in D_c \cap \neg G_t} \frac{1}{1+\kappa \cdot d^2_{G_t}(p)}$.

When FN>0 and FP constant, it behaves like matrix-based error assessments (Fig. 2). Moreover, for FP>0, the FoM penalizes the over-detection very low compared to the under-detection. On the contrary, the F measure computes the distances of FNs but not of the FPs, so F behaves inversely to FoM. Also, d_4 measure depends particularly on TP, FP, FN and FoM but penalizes FNs like the FoM measure. $SFoM$ and $MFoM$ take into account both distances of FNs and FPs, so they can compute a global evaluation of a contour image. However, $MFoM$ does not consider FPs and FNs at the same time, contrary to $SFoM$. Another way to compute a global measure is presented in [28] with the edge map quality measure D_p. The right term computes the distances of the FNs between the closest correctly detected edge pixel, i.e. $G_t \cap D_c$. Finally, D_p is more sensitive to FNs than FPs because of the coefficient $\frac{1}{|I|-|G_t|}$.

A second measure widely computed in matching techniques is represented by the Hausdorff distance H, which measures the mismatch of two sets of points [24].

This max-min distance could be strongly deviated by only one pixel which can be positioned sufficiently far from the pattern (Fig. 3). To improve the measure, one idea is to compute H with a proportion of the maximum distances; let us note $H_{15\%}$ this measure for 15% of the values [24]. Nevertheless, as pointed out in [11], an average distance from the edge pixels in the candidate image to those in the ground truth is more appropriate, like S^k or Ψ. Eventually, Delta Metric (Δ^k) [27] intends to estimate the dissimilarity between each element of two binary images, but is highly sensitive to distances of misplaced points [8][14].

A new objective edge detection assessment measure: In [14] a measure of the edge detection assessment is developed: it is denoted Ψ (Tab. 2) and improvemes the over-segmentation measure Γ, by combining both d_{G_t} and d_{D_c}, see Fig. 3. Ψ gives the same weight for d_{G_t} and d_{D_c} in its assessment of errors. Thus, using Ψ, a missing edge remains not enough penalized contrary to the distance of FPs which could be too important. Another example, in Fig. 3, $\Psi(G_t, C) < \Psi(G_t, T)$ whereas C must be more penalized because of FNs which does not allow to identify the object (also Fig. 5). The solution proposed here is to penalize stronger the distances of the FNs depending on the number of TPs:

$$\lambda(G_t, D_c) = \frac{FP + FN}{|G_t|^2} \cdot \sqrt{\sum_{p \in D_c} d^2_{G_t}(p) + \min\left(|G_t|^2,\ \frac{|G_t|^2}{TP^2}\right) \cdot \sum_{p \in G_t} d^2_{D_c}(p)} \quad (1)$$

The term influencing the penalization of FN distances can be rewritten as: $\frac{|G_t|^2}{TP^2} = \left(\frac{FN+TP}{TP}\right)^2 = \left(1 + \frac{FN}{TP}\right)^2 \geqslant 1$, ensuring a stronger penalty for $d^2_{D_c}$, compared to $d^2_{G_t}$. When $TP = 0$, the min function avoids the multiplication by infinity; moreover, the number of FNs is large, corresponding to a strong penalty with the weight term $|G_t|^2$ (see Fig. 4 left). When $|G_t| = TP$, λ is equivalent to Ψ and Γ (see Fig. 3, image T). Also, compared to Ψ , λ penalizes more D_c having FNs, than D_c with only FPs, as illustrated in Fig. 3 (images C and T). Finally, the weight $\frac{|G_t|^2}{TP^2}$ tunes the λ measure by considering an edge map of better quality when FNs points are localized close to the desired contours D_c.

The next subsection details the way to evaluate an edge detector in an objective way. Results presented in this communication show the importance to penalize stronger the false negative points, compared to the false positive points because the desired objects are not always completely visible by using ill-suited evaluation measure, and, λ provides a reliable edge detection assessment.

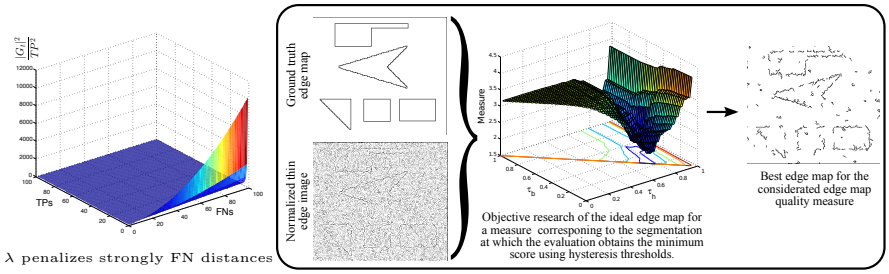


Fig. 4. Number of FNs penalizes λ and computation of a measure minimum score.

2.3 Minimum of the measure and ground truth edge image

Dissimilarity measures are used for an objective assessment using binary images. Instead of choosing manually a threshold to obtain a binary image (see Fig. 3 in [8]), the purpose is to compute the minimal value of a dissimilarity measure by varying the thresholds (τ_L, τ_H) of the thin edges (see Table 1). Thus, compared to a ground truth contour map, the ideal edge map for a measure corresponds to the desired contour at which the evaluation obtains the minimum score for the considered measure among the thresholded (binary) images. Theoretically, this score corresponds to the thresholds at which the edge detection represents the best edge map, compared to the ground truth contour map [30, 12, 8]. Fig. 4 right illustrates the choice of a contour map in function of τ_L and τ_H. Since small thresholds lead to heavy over-segmentation and strong thresholds may create numerous false negative pixels, the minimum score of an edge detection evaluation should be a compromise between under- and over-segmentation (detailed in [8]).

As demonstrated in [8], the significance of the ground truth map choice influences on the dissimilarity evaluations. Indeed, if not reliable [31], an inaccurate ground truth contour map in terms of localization penalizes precise edge detectors and/or advantages the rough algorithms as edge maps presented in [10, 9]. For these reasons, the ground truth edge map concerning the real image in our experiments is built in a semi-automatic way detailed in [8].

3 Experimental results

In these experiments, the importance of an assessment to penalize stronger the false negative points is enlightened, compared to the false positive points. In order to study the performance of the contour detection evaluation measures, the hysteresis thresholds vary and the minimum score of the studied measure corresponds to the best edge map. The thin edges of both synthetic and real noisy images are computed by five or six edge detectors: Sobel [2], Canny [3], Steerable Filters of order 1 (SF_1)[4] or 5 (SF_5)[5], Anisotropic Gaussian Kernels (AGK)[6] and Half Gaussian Kernels (H-K)[7]. Fig. 5 presents the results for 14 measures with their associated scores (bars) according to the hysteresis parameters. In the one hand, we must take into account the obtained edge map, and on the other hand the measure score. Generally, the optimal edge map for FoM, $SFoM$, f_2d_6, Ψ and λ measures allows to distinct the majority of the desired edges for each contour detection operator (except Sobel), whereas for the other assessments, contours are too disturbed by undesirable points or distinguished with high difficulty (especially Ψ which does not penalizes enough FNs). Note that $SFoM$ measure does not classify the Sobel algorithm as less efficient. Concerning the experiment with a real image in Fig. 6, 8 measures are compared together. For FoM, H, Δ^k and S^k, the ideal edge maps concerning Sobel edge detector are highly corrupted by undesirable contours, the main objects are not recognizable. The other segmentations are also disturbed by undesirable pixels for FoM, H and Δ^k. Moreover, the higher score for Δ^k (AGK) does not represent the more disturbed map. Ultimately, using λ, the essential structures are visible in the optimal contour map for each edge detector (objects are easily recognizable).

Moreover, contrary to H, FoM, d_4, Δ^k and S^k measures, the scores of λ are coherent, in relation to the obtained segmentations (Sobel and H-K results).

4 Conclusion and future works

This study presents a new supervised edge detection assessment method λ which enables to assess a contour map in an objective way. Based on the theory of the dissimilarity evaluation measures, the objective evaluation allows to evaluate 1st-order edge detectors. Indeed, the segmentation which obtains the minimum score of a measure is considered as the best one. Theory and experiments prove that the minimum score of the new dissimilarity measure λ corresponds to the best edge quality map evaluations, which is similarly closer to the ground truth, compared to the other methods. On the one hand, this new measure takes into account the distances of false positive points, in the other hand, it considers the distance of false negative points tuned by a weight. This weight depends on the number of false negative points: the more it is elevated, the more the segmentation is penalized. Thus, this enables to obtain objectively an edge map containing the main structures, similar to the ground truth, concerning a reliable edge detector. Finally, the computation of the minimum score of a measure does not require tuning parameters, which represents a huge advantage. For this purpose, we plan in a future study to deeply compare the robustness of several edge detection algorithms and use the new measure in object recognition.

Acknowledgements

The authors thank the Iraqi Ministry of Higher Education and Scientific Research for funding and supporting this work and reviewers for their remarks.

References

1. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE TPAMI*, vol. 33, no. 5, pp. 898–916, 2011.
2. I.E. Sobel, "Camera Models and Machine Perception," *PhD Thesis, Stanford University*, 1970.
3. J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, , no. 6, pp. 679–698, 1986.
4. W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE TPAMI*, vol. 13, pp. 891–906, 1991.
5. 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.
6. J.M. Geusebroek, A. Smeulders, and J. van de Weijer, "Fast anisotropic Gauss filtering," *ECCV 2002*, pp. 99–112, 2002.
7. B. Magnier, P. Montesinos, and D. Diep, "Fast anisotropic edge detection using Gamma correction in color images," in *IEEE ISPA*, 2011, pp. 212–217.
8. H. Abdulrahman, B. Magnier, and P. Montesinos, "From contours to ground truth: How to evaluate edge detectors by filtering," in *WSCG*, 2017.
9. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *IEEE ICCV*. IEEE, 2001, vol. 2, pp. 416–423.

10. 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.
11. M-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *IEEE ICPR*, 1994, vol. 1, pp. 566–568.
12. 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.
13. 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.
14. H. Abdulrahman, B. Magnier, and P. Montesinos, "A new normalized supervised edge detection evaluation," in *IbPRIA*, 2017, pp. 203–213.
15. 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.
16. S. Venkatesh and P. L. Rosin, "Dynamic threshold determination by local and global edge evaluation," *CVGIP*, vol. 57, no. 2, pp. 146–160, 1995.
17. 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.
18. 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.
19. K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical ROC curves," in *CVIU*, 2001, pp. 77–103.
20. 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.
21. A. J. Pinho and L. B. Almeida, "Edge detection filters based on artificial neural networks," in *ICIAP*. Springer, 1995, pp. 159–164.
22. A. G. Boaventura and A. Gonzaga, "Method to evaluate the performance of edge detector," 2009.
23. 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.
24. 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.
25. T. Peli and D. Malah, "A study of edge detection algorithms," *CGIP*, vol. 20, no. 1, pp. 1–21, 1982.
26. C. Odet, B. Belaroussi, and H. Benoit-Cattin, "Scalable discrepancy measures for segmentation evaluation," in *IEEE ICIP*, 2002, vol. 1, pp. 785–788.
27. A. J. Baddeley, "An error metric for binary images," *Robust Computer Vision: Quality of Vision Algorithms*, pp. 59–78, 1992.
28. 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.
29. 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.
30. N.L. Fernández-Garcıa, R. Medina-Carnicer, A. Carmona-Poyato, F.J. Madrid-Cuevas, and M. Prieto-Villegas, "Characterization of empirical discrepancy evaluation measures," *Patt. Rec. Lett.*, vol. 25, no. 1, pp. 35–47, 2004.
31. X. Hou, A. Yuille, and C. Koch, "Boundary detection benchmarking: Beyond F-measures," in *IEEE CVPR*, 2013, pp. 2123–2130.

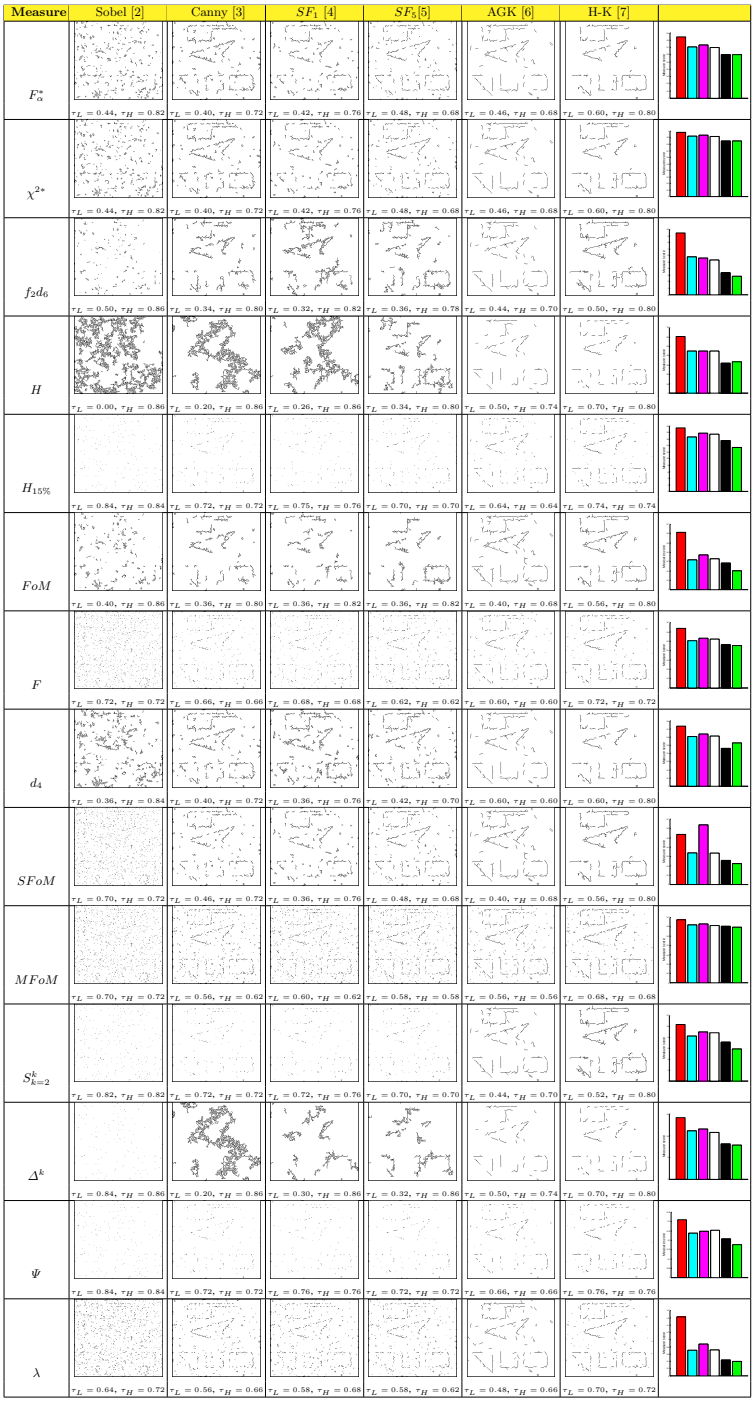


Fig. 5. Comparison of best maps and minimum scores for different evaluation measures. The bars legend is presented in Fig. 6. G_t and original image are available in Fig. 3.

Legend for the bars in Figs. 5 and 6.
The standard deviation for Canny, SFs, AGK and HK is equal to 1.
In Fig. 6, the contour images and scores of Canny are similar to SF_l.

Measure	Sobel [2]	Canny [3]	SF_5 [5]	AGK [6]	H-K [7]	
F_α^*	$\tau_L = 0.16$, $\tau_H = 0.30$	$\tau_L = 0.12$, $\tau_H = 0.20$	$\tau_L = 0.14$, $\tau_H = 0.22$	$\tau_L = 0.10$, $\tau_H = 0.16$	$\tau_L = 0.16$, $\tau_H = 0.24$	
H	$\tau_L = 0.0$, $\tau_H = 0.38$	$\tau_L = 0.06$, $\tau_H = 0.26$	$\tau_L = 0.08$, $\tau_H = 0.5366$	$\tau_L = 0.0$, $\tau_H = 0.26$	$\tau_L = 0.0$, $\tau_H = 0.24$	
FoM	$\tau_L = 0.10$, $\tau_H = 0.34$	$\tau_L = 0.08$, $\tau_H = 0.20$	$\tau_L = 0.08$, $\tau_H = 0.22$	$\tau_L = 0.08$, $\tau_H = 0.12$	$\tau_L = 0.10$, $\tau_H = 0.20$	
d_4	$\tau_L = 0.14$, $\tau_H = 0.28$	$\tau_L = 0.10$, $\tau_H = 0.20$	$\tau_L = 0.10$, $\tau_H = 0.20$	$\tau_L = 0.08$, $\tau_H = 0.14$	$\tau_L = 0.14$, $\tau_H = 0.20$	
Δ^k	$\tau_L = 0.06$, $\tau_H = 0.38$	$\tau_L = 0.08$, $\tau_H = 0.26$	$\tau_L = 0.08$, $\tau_H = 0.28$	$\tau_L = 0.06$, $\tau_H = 0.20$	$\tau_L = 0.12$, $\tau_H = 0.24$	
$S_{k=2}^k$	$\tau_L = 0.06$, $\tau_H = 0.38$	$\tau_L = 0.18$, $\tau_H = 0.18$	$\tau_L = 0.18$, $\tau_H = 0.18$	$\tau_L = 0.14$, $\tau_H = 0.14$	$\tau_L = 0.18$, $\tau_H = 0.18$	
ψ	$\tau_L = 0.28$, $\tau_H = 0.28$	$\tau_L = 0.18$, $\tau_H = 0.18$	$\tau_L = 0.18$, $\tau_H = 0.18$	$\tau_L = 0.14$, $\tau_H = 0.14$	$\tau_L = 0.18$, $\tau_H = 0.18$	
λ	$\tau_L = 0.24$, $\tau_H = 0.24$	$\tau_L = 0.16$, $\tau_H = 0.16$	$\tau_L = 0.16$, $\tau_H = 0.16$	$\tau_L = 0.12$, $\tau_H = 0.12$	$\tau_L = 0.18$, $\tau_H = 0.18$	

Fig. 6. Comparison of best maps and minimum scores for different evaluation measures. G_t and the original real image are presented in Fig. 3.