EDGE DETECTION EVALUATION: A NEW NORMALIZED FIGURE OF MERIT

Baptiste Magnier

LGI2P, IMT Mines Alès, 6. avenue de Clavière, 30319 Alès, France

ABSTRACT

Figure of merit represents an expression characterizing the performance of an algorithm. In edge detection assessment, it corresponds to a supervised evaluation by quantifying differences between a reference edge map and a candidate, computed by a performance measure/criterion. This paper introduces a new normalized supervised edge detection evaluation measure which provides an overall evaluation of the quality of a contour map, by taking into account the amount of false positives, false negatives and degrees of shifting. Finally, an objective assessment performed by varying the hysteresis thresholds on contours of real images shows that the new measure outperforms six other compared normalized methods, in term of evaluation and visualization of the detected contours.

Index Terms— Edge detection, objective evaluation.

1. INTRODUCTION

Edge detection is a fundamental process and is widely used in image analysis and computer vision applications. Hence, an edge extractor needs to perform efficiently despite different perturbations in the data which might be created by the image acquisition. Consequently, a large number of edge detection algorithms have been developed and many methods continue to be proposed proposed [1, 2]. So, edge detectors have to be carefully tested and assessed in order to study the influence of the input parameters. However, most contour extraction methods suffer from the lack of effective, reliable and automated edge quality evaluation measurement/metric [3]. Moreover, an objective evaluation of the segmentation remains vital for many applications: medical, aerial, satellite, tracking an so on. This study focusses on supervised edge detection evaluations with respect to the binary representation of the boundaries. A supervised evaluation process estimates scores between a ground truth and a candidate edge map (both binary images). These scores could be evaluated by counting the number of erroneous pixels, but also through the spatial distances of misplaced or undetected contours. Therefore, a lot of works have been put into developping efficient metrics. Even though some of these metrics are widely studied in the literature, they are seldom applied objectively and few of them are normalized.

This paper seeks to propose a new supervised edge detection evaluation which would allow us to compare several edge detectors by filtering in an objective way by varying the hysteresis thresholds of the thin gradient image. The proposed measure computes a normalized score in function of weights of both false positive and false negative points. Rather, by varying the parameters of the hysteresis thresholds, the ideal edge map can be objectively determined and appears visually as the best edge maps in quality when compared to other edge detection assessments. The next section presents a review of the most relevant normalized metrics present in the literature.

2. ON EXISTING NORMALIZED MEASURES

Various supervised methods have been proposed in the literature to assess edge detectors [8, 9, 10, 11]. The normalization remains valuable to compare a set of algorithms more easily. The most frequently normalized measures of dissimilarity are described here. Each measure computes a score of quality; the closer to 1 the score of the evaluation is, the more the segmentation is qualified as suitable. On the contrary, a score close to 0 corresponds to a poor edge detection.

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, a basic evaluation is composed of statistics:

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

where $|\cdot|$ denotes the cardinality of a set. Several edge detection evaluations involving only statistics have been developed, see [9, 10, 11]. Among them, the *Performance measure* P_m simultaneously considers the three entities TP, FP and FN [12, 13] and is currently used in segmentation:

$$P_m(G_t, D_c) = \frac{TP}{|G_t \cup D_c|} = \frac{TP}{TP + FP + FN} \qquad (1)$$

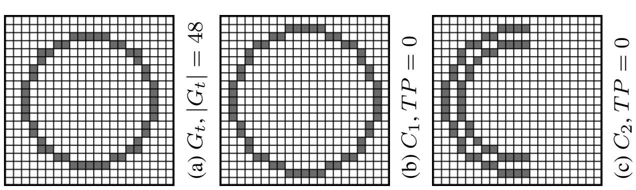
Other statistical measures are similar to P_m [14, 15, 16] or worse in objective evaluation, see [17, 18], so P_m is the basis for the comparison in this paper. Also, statistical measures such as ROC [3] or PR [16] curves evaluate the comparison of two edge images, pixel per pixel, but do not detect when the

Table 1. List of normalized dissimilarity measures involving distances, generally: $\kappa = 0.1$ or $1/9$.

Error measure name	Formulation	Parameters
Pratt's FoM [4]	$FoM\left(G_t, D_c\right) = \frac{1}{\max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)} \cdot \sum_{p \in D_c} \frac{1}{1 + \kappa \cdot d^2_{G_t}(p)}$	$\kappa \in]0;1]$
FoM revisited [5]	$F\left(G_t, D_c\right) = \frac{1}{\lvert G_t\rvert + \beta \cdot FP} \cdot \sum_{p \in G_t} \frac{1}{1 + \kappa \cdot d^2_{D_c}(p)}$	$\kappa \in]0;1]$ and $\beta \in \mathbb{R}^+$
Combination of FoM and statistics [6]	$d_4\left(G_t, D_c\right) = 1 - \frac{1}{2} \cdot \sqrt{\frac{\left(TP - \max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)\right)^2 + FN^2 + FP^2}{\left(\max\left(\lvert G_t\rvert, \lvert D_c\rvert\right)\right)^2} + \left(1 - FoM\left(G_t, D_c\right)\right)^2}$	$\kappa \in]0;1]$
Edge map quality measure [7]	$D_p\left(G_t, D_c\right) = 1 - \frac{1/2}{\lvert I\rvert - \lvert G_t\rvert} \sum_{p \in FP}\left(1 - \frac{1}{1 + \kappa \cdot d^2_{G_t}(p)}\right) - \frac{1/2}{\lvert G_t\rvert} \sum_{p \in FN}\left(1 - \frac{1}{1 + \kappa \cdot d^2_{TP}(p)}\right)$	$\kappa \in]0;1]$

edge is correctly identified with a small displacement, which tends to severely penalize a (even slightly) misplaced contour (cf. Fig.1). Consequently, some evaluations resulting from the confusion matrix recommend incorporating spatial tolerance as the *Performance value* P_v in [19]. Despite P_v being normalized, the same measurement is obtained for two different shapes (see [20]). Note that the approach presented in [21] gives interesting results; unfortunately, it is not normalized.

As a result, a reference-based edge map quality measure requires that a displaced edge should be penalized not just according to FPs and/or FNs but also according to the distance from the position where it should be located [8, 10, 11, 20, 22]. A review of normalized edge evaluation measures involving the distances of errors is presented in this paragraph. Thus, for a pixel p belonging to a contour in an image, $d_E(p)$ represents the minimal Euclidian distance between p and another edge E in a compared image. Table 1 reviews the most relevant normalized criteria/measures involving distances. Alternative dissimilarity measures, inspired from the Hausdorff distance [23], but non-normalized, have been proposed in the literature, see [8, 9, 10, 18, 17, 11]. In [11], a normalization for the measure of distances is proposed, but it is not really practical for real images. In edge detection evaluation, a widely used similarity measure refers to FoM [4]. The parameter κ plays the role of scale parameter, the more κ is close to 1, the more FoM tackles FPs [11]. Nonetheless, the distance of the FNs is not recorded and are strongly penalized as statistic measures (see [20] and Fig.1): $FoM(G_t, D_c) = \frac{1}{\max(|G_t|, |D_c|)} \cdot \left(TP + \sum_{p \in FP} \frac{1}{1+\kappa \cdot d^2_{G_t}(p)}\right)$. Thereby, different shapes have the same interpretation, as in Fig.1. Further, if $FP = 0$: $FoM\left(G_t, D_c\right) = TP/|G_t|$. Lastly, for $FP{>}0$, FoM penalizes the over-detection much less than it does under-detection [11]. Several evaluation measures are derived from FoM: FoM_e, F, d_4 and D_p. Firstly, FoM_e represents a modified version FoM by counting only FPs, corresponding to an over-segmentation evaluation: $FoM_e\left(G_t, D_c\right) = \frac{1}{\max\left(\mathrm{e}^{-FP}, FP\right)} \cdot \sum_{p \in FP} \frac{1}{1+\kappa \cdot d^2_{G_t}(p)}$, with $\kappa \in]0;1]$. Consequently, FoM_e does not penalize FNs, without FPs, a contour image is considered to be correctly segmented (cf. black curve in Fig.2 middle). Otherwise, contrary to FoM, the F measure computes the distances of FNs but not of FPs. Thus, FPs are strongly penalized, as in Fig.1 with the low score of 0.38. Furthermore, different G_t leads to the same score for the same D_c: in Fig.1, $F(C_1, G_t){=}F(C_2, G_t){=}0.42$. On the other hand, d_4 measurement depends particularly on TP, FP, FN and $\approx$1/4 on FoM, but d_4 penalizes FNs like the FoM measure (cf. Fig.1). Otherwise, the right term of the dissimilarity measure D_p [7] computes the distances of the FNs between the closest correctly detected edge pixel, i.e., TPs (FNs are huge tackled when TPs are far from FPs, or when $G_t \cap D_c = \varnothing$ as in Fig.1). Finally, D_p is more sensitive to FNs than FPs because of the huge coefficient $\frac{1}{|I|-|G_t|}$ for the left term (detailed in [20]).



$P_m(G_t, C_1) = 0$	$P_m(G_t, C_2) = 0$
$FoM(G_t, C_1) = 0.8$	$FoM(G_t, C_2) = 0.8$
$FoM_e(G_t, C_1) = 0.2$	$FoM_e(G_t, C_2) = 0.2$
$F(G_t, C_1){=}0.38$	$F(G_t, C_2) = 0.26$
$d_4(G_t, C_1){=}\ 0.15$	$d_4(G_t, C_2) = 0.15$
$D_p((G_t, C_1){=}\ 0.48$	$D_p(G_t, C_2) = 0.46$
$\mathcal{N}^{\kappa_{FP}=0.1}_{\kappa_{FN}=0.4}(G_t, C_1) = 0.81$	$\mathcal{N}^{\kappa_{FP}=0.1}_{\kappa_{FN}=0.4}(G_t, C_2) = 0.64$
$\mathcal{N}^{\kappa_{FP}=0.1}_{\kappa_{FN}=0.2}(G_t, C_1) = 0.87$	$\mathcal{N}^{\kappa_{FP}=0.1}_{\kappa_{FN}=0.2}(G_t, C_2) = 0.70$

Fig. 1. Different D_c: FPs and number of FNs are the same for C_1 and for C_2 ($FN = 48$, $FP = 52$), but the distances of FNs and the shapes of the two D_c are different.

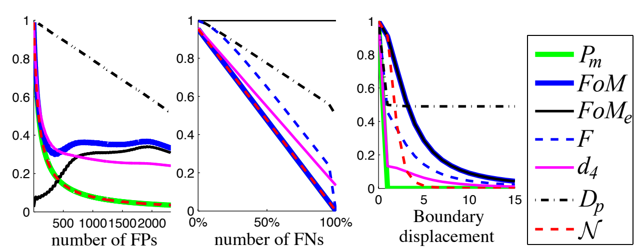


Fig. 2. Evolution of normalized measures in function of addition of FPs or FNs, or, edge displacement. On the left, P_m, F and $\mathcal{N}$ overlap. In the middle, P_m, FoM and $\mathcal{N}$ overlap. On the right, FoM and FoM_e overlap. Here $\kappa_{FP}{=}\kappa_{FN}{=}\kappa{=}0.1$.

3. A NEW NORMALIZED MEASURE

Three curves in Fig.2 illustrate the motivations and the drawbacks of the measures presented above as a function of the number of FPs, FNs or of the boundary displacements. For the first curves, FPs are added to D_c until they recover completely the image. Concerning the second curve, FNs are created until D_c disappears completely. The third curve displays the measurement scores as a function of the translation of the desired edges. The curve must start from 1 and converge monotonously toward 0. Nevertheless, FoM is not monotonous for FPs addition and does not penalize severely boundary displacements, contrary to P_m and d_4 which are sensitive to displacement. Also, D_p obtains a score not lower than 0.5 because it corresponds to a measure which separates under- and over-segmentation; (the score becomes close to 0 only with addition of both FPs and FNs). Finally, F does not penalize displacements enough whereas d_4 obtains a score of 0.25 for 100% of FPs in the D_c image and FoM_e considers only FPs distances. These curves illustrate that a dissimilarity measure must take into account both FP and FN distances ($d_{G_t}(p)$ and $d_{D_c}(p)$ respectively) for an objective assessment.

The main motivation is that there exists no normalized measure which considers both FPs and FNs distances capable of obtaining a desired evolution like in Fig.2. Secondly, as demonstrated in [20], the evaluation of FPs and FNs distances must not be symmetrical, because a symmetrical assessment could alter the visibility of desired objects in an objective evaluation [18]: some measures compute a strong mistake for a single FP point at a sufficiently high distance (as Hausdorff distance [8, 20]), whereas numerous desired contours are missing, but unfortunately are not penalized enough (see next section). Therefore, inspired by the *Relative Distance Error* in [32] and demonstrations in [20], for $FN>0$ or $FP>0$, the new edge detection evaluation formula becomes:

$$\mathcal{N}(G_t, D_c) = \frac{1}{FP+FN} \cdot \left[\frac{FP}{|D_c|} \cdot \sum_{p \in D_c} \frac{1}{1+\kappa_{FP} \cdot d^2_{G_t}(p)} + \frac{FN}{|G_t|} \cdot \sum_{p \in G_t} \frac{1}{1+\kappa_{FN} \cdot d^2_{D_c}(p)} \right], \quad (2)$$

where $(\kappa_{FP}, \kappa_{FN}) \in]0,1]^2$ represent two scale parameters and the coefficient $\frac{1}{FP+FN}$ normalizes the $\mathcal{N}$ function. If FP=FN=0, then $\mathcal{N}$=1. Therefore, to become as fair as possible, FPs and FNs distances are penalized in function of the relationship between FPs and $|D_c|$ and between FNs and $|G_t|$ respectively, ensuring an equal distribution of mistakes, without symmetry of penalties. However, when $\kappa_{FP}<\kappa_{FN}$, $\mathcal{N}$ penalizes the FNs more, compared to the FPs (see Fig.1 and 4). Results presented below show the importance of the weights given for FNs because the desired objects are not always completely visible by using ill-suited evaluation measures. Eventually, $\mathcal{N}$ has a normalized desired behavior for strong mistakes: FPs, FNs addition and boundary displacement (Fig.2), becoming very close to 0 for a large number of FPs and displacement. Note that the Matlab code is available.

4. EXPERIMENTS: OBJECTIVE EVALUATION

The aim of the experiments is to obtain the best edge map in a supervised way. In the experiments, 9 edge detection methods based on filtering gradient computation are compared: Sobel [24], Shen [25], Bourennane [26], Deriche [27], Canny [28], Steerable filter of order 1 (SF_1) [29], of order 5 (SF_5) [30], Anisotropic Gaussian Kernels (AGK) [31], and Half Gaussian Kernels (H-K) [32]. The parameters of the filters are chosen to keep the same spatial support for the derivative information [32]. Finally, after a non-maximum suppression, a hysteresis threshold is applied on thin edges to obtain a binary edge map [28]. Theoretically, to be objectively compared, the ideal edge map for a measure must correspond to a D_c at which the evaluation obtains the maximum score [33, 9, 11, 17, 20]. These scores of the different measures are recorded by varying the thresholds of the normalized thin edges computed by an edge detector and plotted as a function of the PSNR, as presented in Fig.3 (top). Hence, a plotted curve must decrease monotonously with the noise level (white Gaussian noise).

As demonstrated in [34, 17], the significance of the ground truth map choice influences the dissimilarity evaluations but is not discussed here. Among all the edge detectors, box [24] and exponential [25, 26] filters do not delocalize contour points [35] whereas they are sensitive to noise (i.e., addition of FPs). The Deriche [27] and Gaussian filters [28] are less sensitive to noise, but suffer from rounding corners and junctions (see [35]) as the oriented filters SF_1 [29], SF_5 [30] and AGK [31], but the more the 2D filter is elongated, the more the segmentation remains robust against noise. At last, as a compromise, H-K correctly detects contour points that have corners and is robust against noise [32]. Consequently, the scores of the evaluation measures for the first 3 filters must be lower than the three last ones. Furthermore, as SF_5, AGK and H-K are less sensitive to noise than other filters, the ideal segmented image for these 3 algorithms should be visually closer to G_t (Fig.3(j)). The presented segmentations correspond to the original image for a PSNR=14dB. Concerning the P_m, F and d_4 measures, the curves are confused, it is not easy to notice which edge detector is better. For the segmented images, Canny and AGK give almost the same score for P_m but the contours of AGK are straighter and less noisy. For F and d_4, the scores of the Canny or Deriche filter are either better than or similar to H-K whereas the segmentation remains clearly better for H-K. Also, the curves of FoM are not monotonous, segmented images are too noisy and the Shen segmentation where objects are not visually distinctive is considered better than H-K. Otherwise, the D_p scores are monotonous, but the worse segmentation corresponds to the more robust filter against noise, whereas the segmentations are heavily corrupted by FPs. On the other hand, values of $\mathcal{N}$ evolve monotonously, respecting the orders of the filters: box, exponentials at the bottom, Deriche and Gaussians in the middle and robust against noise on top. The presented segmentations are consistent with the scores, and the main structures

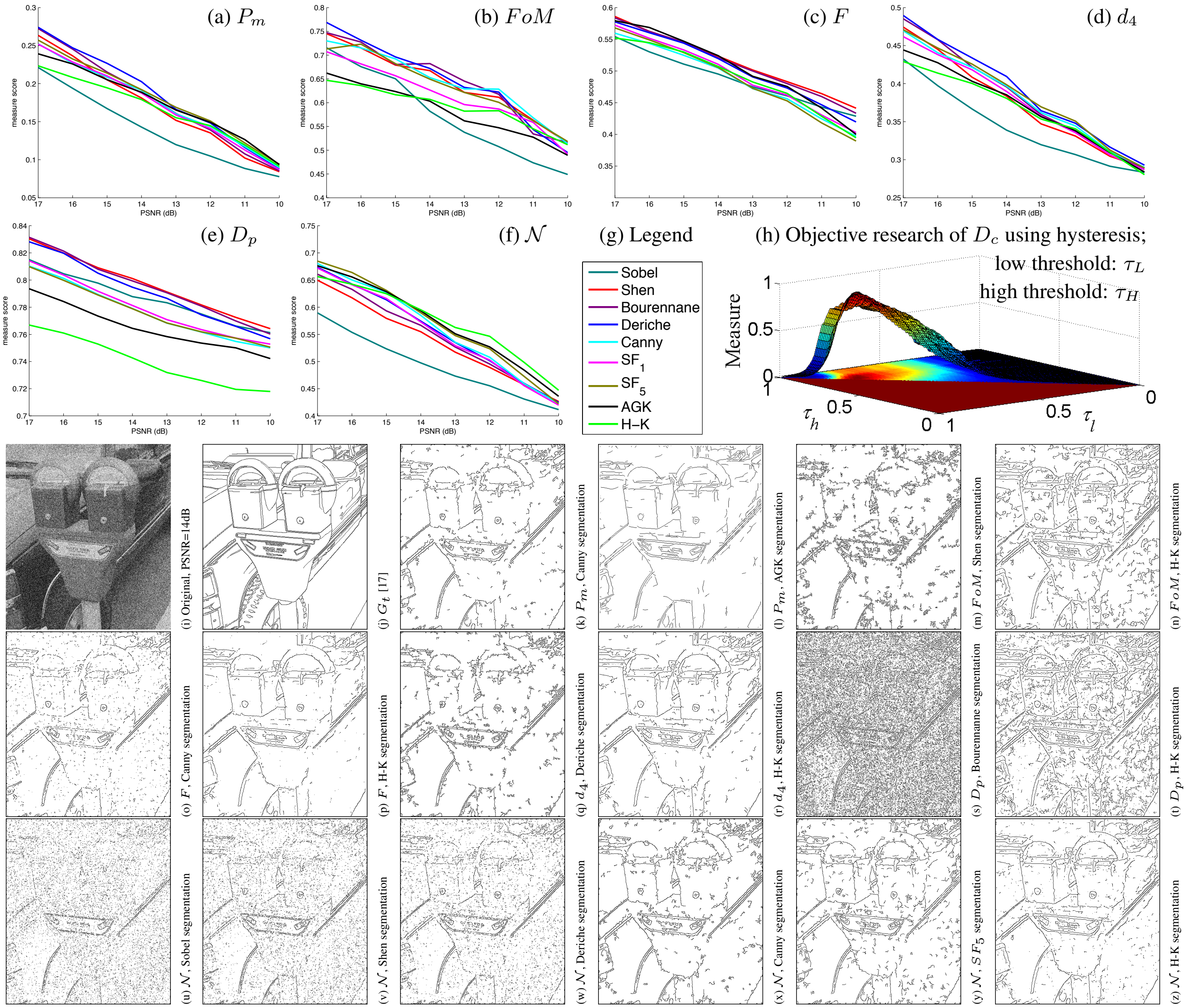


Fig. 3. Comparison of best maps and maximum scores for different evaluation measures, with κ=0.1, κ_{FN}=0.2 and κ_{FP}=0.1.

are visible, without being submerged by undesirable points. Otherwise, Fig. 4 illustrates when $\kappa_{FP}>\kappa_{FN}$, extracted edges create filaments. However, when $\kappa_{FP}<<\kappa_{FN}$, the edge map contains a lot of undesirable isolated points (FPs). Note that curves of the scores have the same gait with a constant gap. Generally, results on dozens of noisy images at different noise levels are similar for all the dissimilarity measures.

(a) Shen [25], PSNR=14dB (b) Shen [25], PSNR=14dB (c) Scores of $\mathcal{N}$

Fig. 4. Best maps with different κ_{FP} and κ_{FN} parameters.

5. CONCLUSION

For a computer vision application, edge detection evaluation remains an element in the choice of an edge detector. This paper describes a supervised evaluation $\mathcal{N}$ of the quality of a desired contour. Theoretically, $\mathcal{N}$ ranges the shifting and deformation error between a desired contour and a ground truth as a function of FPs and FNs distances. Experimental results are presented in an objective way. Finally, when $\kappa_{FN}>\kappa_{FP}$ (parameters of $\mathcal{N}$ penalizing FPs and FNs), the maximum score of the new dissimilarity measure corresponds to the best edge quality map evaluation, which, visually, is similarly closer to the ground truth (i.e., containing the main structures), compared to the other methods. Moreover, $\mathcal{N}$ is normalized and ranged scores between 0 and 1 are reliable and consistent with both the noise level and the edge detector quality. For this purpose, future works concern the objective assessment of image restoration methods through edge detection evaluation.

6. 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] G. Papari and N. Petkov, "Edge and line oriented contour detection: State of the art," *IVC*, vol. 29, no. 2, pp. 79–103, 2011.

[3] K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical ROC curves," in *CVIU*, 2001, pp. 77–103.

[4] 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.

[5] A.J. Pinho and L.B. Almeida, "Edge detection filters based on artificial neural networks," in *ICIAP*. Springer, 1995, pp. 159–164.

[6] A. G. Boaventura and A. Gonzaga, "Method to evaluate the performance of edge detector," in *Brazlian Symp. on Comput. Graph. Image Process.* 2006, pp. 234–236, Citeseer.

[7] 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.

[8] M-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *IEEE ICPR*, 1994, vol. 1, pp. 566–568.

[9] 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.

[10] C. Lopez-Molina, B. De Baets, and H. Bustince, "Quantitative error measures for edge detection," *Pat. Rec.*, vol. 46, no. 4, pp. 1125–1139, 2013.

[11] B. Magnier, "Edge detection: a review of dissimilarity evaluations and a proposed normalized measure," *Multimedia Tools and Applications*, pp. 1–45, 2017.

[12] P. Sneath and R. Sokal, *Numerical taxonomy. The principles and practice of numerical classification*, 1973.

[13] S. Wang, F. Ge, and T. Liu, "Evaluating edge detection through boundary detection," *EURASIP J. on App. Signal Proc.*, vol. 2006, pp. 213–213, 2006.

[14] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[15] R. Usamentiaga, D. F. García, C. López, and D. González, "A method for assessment of segmentation success considering uncertainty in the edge positions," *EURASIP J. on Applied Signal Proc.*, vol. 2006, pp. 207–207, 2006.

[16] 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.

[17] H. Abdulrahman, B. Magnier, and P. Montesinos, "From contours to ground truth: How to evaluate edge detectors by filtering," *Journal of WSCG*, vol. 25, no. 2, pp. 133–142, 2017.

[18] H. Abdulrahman, B. Magnier, and P. Montesinos, "A new objective supervised edge detection assessment using hysteresis thresholds," in *International Workshop on Brain-Inspired Computer Vision, held as part of ICIAP*, 2017, pp. 3–14.

[19] F. L. Valverde, N. Guil, J. Munoz, R. Nishikawa, and K. Doi, "An evaluation criterion for edge detection techniques in noisy images," in *IEEE ICIP*. IEEE, 2001, vol. 1, pp. 766–769.

[20] B. Magnier, H. Abdulrahman, and P. Montesinos, "A review of supervised edge detection evaluation methods and an objective comparison of filtering gradient computations using hysteresis thresholds," *Journal of Imaging*, vol. 4, no. 6, pp. 74, 2018.

[21] R. Román-Roldán, J.F. Gómez-Lopera, C. Atae-Allah, J. Martınez-Aroza, and P. Luque-Escamilla, "A measure of quality for evaluating methods of segmentation and edge detection," *Pat. Rec.*, vol. 34, no. 5, pp. 969–980, 2001.

[22] H. Abdulrahman and B. Magnier, "A study of measures for contour-based recognition and localization of known objects in digital images," in *IEEE IPTA - to appear-*, 2018.

[23] 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.

[24] B. Magnier, "An objective evaluation of edge detection methods based on oriented half kernels," *ICISP*, pp. 80–89, 2019.

[25] I.E. Sobel, "Camera Models and Machine Perception," *PhD Thesis, Stanford University*, 1970.

[26] J. Shen and S. Castan, "An optimal linear operator for step edge detection," *CVGIP*, vol. 54, no. 2, pp. 112–133, 1992.

[27] E. Bourennane, P. Gouton, M. Paindavoine, and F. Truchetet, "Generalization of canny–deriche filter for detection of noisy exponential edge," *Signal Processing*, vol. 82, no. 10, pp. 1317–1328, 2002.

[28] R. Deriche, "Using canny's criteria to derive a recursively implemented optimal edge detector," *IJCV*, vol. 1, no. 2, pp. 167–187, 1987.

[29] J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, , no. 6, pp. 679–698, 1986.

[30] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE TPAMI*, vol. 13, pp. 891–906, 1991.

[31] 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.

[32] J.M. Geusebroek, A. Smeulders, and J. van de Weijer, "Fast anisotropic Gauss filtering," *ECCV*, pp. 99–112, 2002.

[33] N.L. Fernández-Garcıa, R. Medina-Carnicer, A. Carmona-Poyato, F.J. Madrid-Cuevas, and M. Prieto-Villegas, "Characterization of empirical discrepancy evaluation measures," *Pat. Rec. Let.*, vol. 25, no. 1, pp. 35–47, 2004.

[34] X. Hou, A. Yuille, and C. Koch, "Boundary detection benchmarking: Beyond F-measures," in *IEEE CVPR*, 2013, pp. 2123–2130.

[35] O. Laligant, F. Truchetet, and F. Meriaudeau, "Regularization preserving localization of close edges," *IEEE Sig. Proc. Let.*, vol. 14, no. 3, pp. 185–188, 2007.