



HAL
open science

A New Normalized Method of Object Shape-based Recognition and Localization

Behrang Moradi, Hasan Abdulrahman+, Baptiste Magnier

► **To cite this version:**

Behrang Moradi, Hasan Abdulrahman+, Baptiste Magnier. A New Normalized Method of Object Shape-based Recognition and Localization. Conference: The International Conference on Pattern Recognition Systems (ICRPS-19)At: Tours, France, Jul 2019, Tours, France. 10.1049/cp.2019.0246 . hal-02188474

HAL Id: hal-02188474

<https://hal.science/hal-02188474>

Submitted on 18 Jul 2019

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 Normalized Method of Object Shape-based Recognition and Localization

Behrang MORADI*, Hasan ABDULRAHMAN+, Baptiste MAGNIER*

* IMT Mines Alès, LGI2P, 6. avenue de Clavières 30100 Alès, France

+ Northern Technical University, Department of Technical Computer Systems, Kirkuk 36001, Iraq

e-mail: behrang.moradi@mines-ales.fr; hasan.abdulrahman@mines-ales.fr; baptiste.magnier@mines-ales.fr

Keywords: Distance measures, contours, shape.

Abstract

This paper introduces a new normalized measure for the assessment of a contour-based object pose. This algorithm allows a supervised assessment of recognition and localization of known objects, the differences between a reference edge map and a candidate image are quantified by computing a performance measure. This measure is normalized and is able to determine the degree to which an object shape differs from a desired position. Compared to 6 other approaches, experiments on real images at different sizes/scales exhibit the suitability of the new method for an object pose or shape matching estimation.

1 Introduction and motivation

The representation of an object shape is particularly useful for specific industrial inspection tasks. When this shape is aligned with a reference model, many different manipulations might arise. An edge-based representation remains a class of methods and exploits only shape-boundary information. This represents a supervised evaluation of the shape depiction. This paper presents a new approach for the measurement of a contour-based object pose. The proposed measurement estimates a supervised normalized score evaluation of the shape representation based on the weights created by both false positives and false negatives edge pixels. The normalization is especially appropriate to interpret an algorithm result. Indeed, normalization represents a technical operator which is able to qualify when a score is suitable as a function of the desired operation: if the score becomes close to 1, thus the action is considered as good, whereas a score close to 0 corresponds to a wrong initiative. The new method uses different strategies in order to normalize and obtain a reliable assessment of the contour-based object localization. Eventually, compared to 6 other normalized measures, the new approach computes a coherence score, which qualifies the correct object pose possibility.

2 On existing normalized measures

Various evaluation methods have been proposed in the literature to assess different shapes of edges using pixel-based ground truth (see reviews in [1],[2],[3],[4]). Indeed, a supervised evaluation criterion computes a dissimilarity measure be-

tween a ground truth (G_t) and a detected contour map (D_c) of an original image I . In this paper, the closer to 1 the score of the evaluation is, the more the object localization is qualified as suitable. On the contrary, a score close to 0 corresponds to a poor object positioning. To assess a known shape, the confusion matrix remains a cornerstone in evaluation methods. 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 by combining G_t and D_c . Afterwards, denoting $|\cdot|$ as the cardinality of a set (e.g. $|G_t|$ represents the number of edge pixels in G_t), all points are divided into four sets, as illustrated Fig.1:

- True Positive points (TPs): $TP = |G_t \cap D_c|$,
- False Positive points (FPs): $FP = |\neg G_t \cap D_c|$,
- False Negative points (FNs): $FN = |G_t \cap \neg D_c|$,
- True Negative points (TNs): $TN = |\neg G_t \cap \neg D_c|$.

Various edge detection evaluations involving confusion matrices have been developed, cf. [3][4]. The *Dice* measure [10] is one well known example: $Dice(G_t, D_c) = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}$. This type of assessment is useful for region segmentation evaluation [10], but, a reference-based edge map quality measure requires that a displaced edge should be penalized in function of FPs and/or FNs and of the distance from the position where it should be located [3][4], as illustrated in Fig. 1 with arrows.

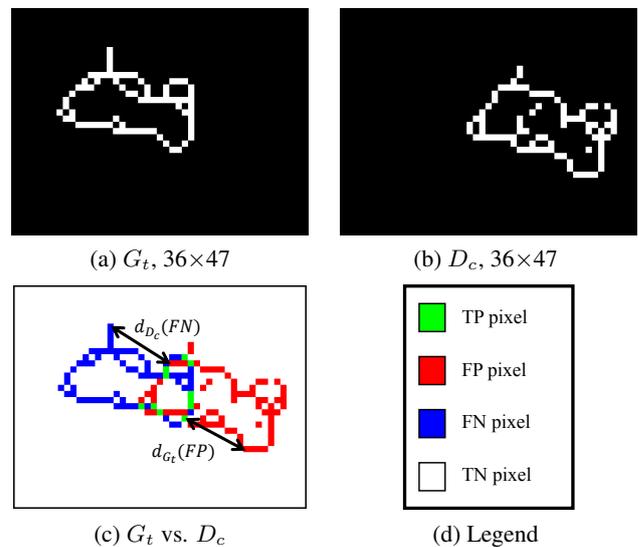


Figure 1. Example a ground truth G_t and a desired contour D_c .

Error measure name	Formulation	Parameters
Pratt's Figure of Merit [5]	$FoM(G_t, D_c) = \frac{1}{\max(G_t , D_c)} \cdot \sum_{p \in D_c} \frac{1}{1 + \kappa \cdot d_{G_t}^2(p)}$	$\kappa \in]0; 1]$
FoM revisited [6]	$F(G_t, D_c) = \frac{1}{ G_t + \beta \cdot FP} \cdot \sum_{p \in G_t} \frac{1}{1 + \kappa \cdot d_{D_c}^2(p)}$	$\kappa \in]0; 1]$ and β a real positive
Combination of FoM and statistics [7]	$d_4(G_t, D_c) = 1 - \frac{1}{2} \cdot \sqrt{\frac{(TP - \max(G_t , D_c))^2 + FN^2 + FP^2}{(\max(G_t , D_c))^2} + (1 - FoM(G_t, D_c))^2}$	$\kappa \in]0; 1]$
Edge map quality measure [8]	$D_p(G_t, D_c) = 1 - \frac{1/2}{ I - G_t } \sum_{p \in FP} \left(1 - \frac{1}{1 + \kappa \cdot d_{G_t}^2(p)}\right) - \frac{1/2}{ G_t } \sum_{p \in FN} \left(1 - \frac{1}{1 + \kappa \cdot d_{TP}^2(p)}\right)$	$\kappa \in]0; 1]$
Edge Mismatch Measure (EMM) [9]	$EMM(G_t, D_c) = \frac{TP}{TP + \omega \cdot \left[\sum_{p \in FN} \delta_{D_c}(p) + \epsilon \cdot \sum_{p \in FP} \delta_{G_t}(p) \right]}$	$M_{dist}, D_{max}, \omega$ and ϵ are real positive.
	$\delta_{D_c}(p) = \begin{cases} d_{D_c}(p), & \text{if } d_{D_c}(p) < M_{dist} \\ D_{max}, & \text{otherwise} \end{cases}$ and $\delta_{G_t}(p) = \begin{cases} d_{G_t}(p), & \text{if } d_{G_t}(p) < M_{dist} \\ D_{max}, & \text{otherwise.} \end{cases}$	$M_{dist} = I /40,$ $D_{max} = I /10,$ $\omega = 10/ I ,$ $\epsilon = 2,$ see [9].

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

Table 1 reviews the most relevant normalized measures involving distances. Thus, for a pixel p belonging to the candidate contour D_c , $d_{G_t}(p)$ represents the minimal Euclidian distance between p and G_t . These types of distance measures play an important role in image matching and may be used to determine the degree of resemblance between two objects shapes [1]. To achieve this, if p belongs to G_t , $d_{D_c}(p)$ corresponds to the minimal distance between p and D_c , Fig. 1 illustrates the difference between $d_{G_t}(p)$ and $d_{D_c}(p)$. Mathematically, denoting (x_p, y_p) and (x_t, y_t) the pixel coordinates of two points p and t respectively, thus $d_{G_t}(p)$ and $d_{D_c}(p)$ are described by:

$$\begin{cases} \text{for } p \in D_c: d_{G_t}(p) = \text{Inf} \left\{ \sqrt{(x_p - x_t)^2 + (y_p - y_t)^2}, t \in G_t \right\}, \\ \text{for } p \in G_t: d_{D_c}(p) = \text{Inf} \left\{ \sqrt{(x_p - x_t)^2 + (y_p - y_t)^2}, t \in D_c \right\}. \end{cases}$$

In the domain of shape positioning area, alternative dissimilarity measures, inspired from the Hausdorff distance [1], have been proposed in the literature, see [1][2][11][3]. The majority represents non-normalized measures. In [3], 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 normalized similarity measure refers to FoM [5]. The parameter κ plays the role of scale parameter, the more κ is close to 1, the more FoM tackles FPs [3]. Nonetheless, the distance of the FNs is not recorded and are strongly penalized as statistic measures (detailed in [4]):

$$FoM(G_t, D_c) = \frac{1}{\max(|G_t|, |D_c|)} \cdot \left(TP + \sum_{p \in FP} \frac{1}{1 + \kappa \cdot d_{G_t}^2(p)} \right).$$

Thereby, different shapes have the same interpretation [3]. Further, if $FP = 0$: $FoM(G_t, D_c) = TP/|G_t|$. When $FN > 0$ and $d_{G_t}^2(FP)$ constant, it behaves like matrix-based error assessments (detailed in [3]). Lastly, for $FP > 0$, FoM penalizes the over-detection much less than it does under-detection [3]. Several evaluation measures are derived from FoM : F , d_4 , D_p and EMM . Firstly, contrary to FoM , the F measure computes the distances of FNs but not of FPs, thus, FPs are strongly penalized. 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. Otherwise, the right term of the dissimilarity measure D_p [8] 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 $G_t \cap D_c = \emptyset$). Also, 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 [4]). On another note, the Edge Mismatch Measure (EMM) depending on TPs and both d_{D_c} and d_{G_t} . Thus, $\delta_{D_c/G_t}(p)$ is a threshold distance function penalizing high distances (exceeding a maximum distance $maxdist$). Note that the suggested parameters depend on $|I|$, the total number of pixels in I . Moreover, EMM computes a score different from 0 only if there exists at least one TP.

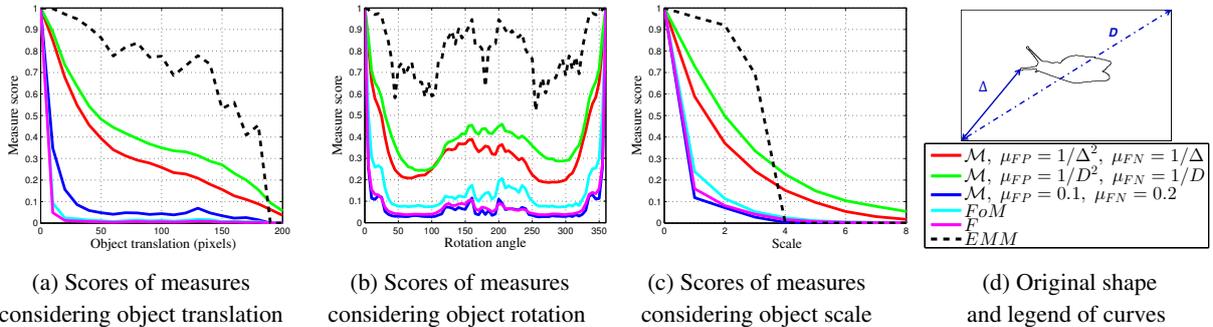


Figure 2. Examples of localization metrics behaviors for a translation, rotation and scale alterations. Several parameters for \mathcal{M} are tested: Δ represents the maximum distance between a pixel in D_c with G_t (usually a pixel of an image corner), whereas D is the diagonal length of the considered image. Parameters D and Δ are computed automatically and remark that $D > \Delta$.

3 A new normalized measure

The main motivation is that there exists no normalized shape-based measure which considers both FPs and FNs distances capable of recording a desired evolution of the object localization. As detailed in [12], the evaluation of FPs and FNs distances must not be symmetrical. Clearly, a measure involving false negative distances remains more accurate than other techniques. However, using only under-segmentation measure, when some parts of the candidate image are missing, but detected close to their desired positions, they are not taken into account (by F for example), the object is not well localized. Missing edges must be penalized higher than spurious ones because some isolated points may disturb the shape location, as the majority of the measures, cf. experiments. To summarize, a measure must penalize highly FNs compared to FPs, because the more FNs are present in D_c , the more the shape of the desirable object becomes unrecognizable and thus localizable.

Therefore, inspired by the *Relative Distance Error* in [13] and demonstrations in [4, 12], for $FN > 0$ or $FP > 0$, the new normalized distance measure formula becomes:

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

where (μ_{FP}, μ_{FN}) are real positives representing the two scale parameters and the coefficient $\frac{1}{FP+FN}$ normalizes the \mathcal{M} function. If $FP=FN=0$, then $\mathcal{M}=1$. Thereafter, 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 $\mu_{FP} < \mu_{FN}$, \mathcal{M} penalizes the FNs more, compared to the FPs. Results presented below show the importance of the weights given for FNs because some isolated points may disturb the shape location. The next section illustrates the interest to choose the optimum values for the parameters (μ_{FP}, μ_{FN}) tied to the maximum Euclidian distance between G_t and any pixel in the image.

4 Evaluation and results

Firstly, in order to test different parameters and verify if the proposed measure has the required properties, several alterations are applied to create synthetic localization results simulating real ones. Thereby, to quantify the reliability of a measure of dissimilarity, an edge map of a synthetic shape is affected by various alterations: rotation, translation and scale change (in Fig.2). Thus, it allows to verify if the evolution of the provided score obtained by a measure is in line with the expected behavior: small error for close shapes (score close to 1) and large penalty for shapes that differ (score close to 0).

In the first test, the contour shape is gradually translated by moving the shape away from its original location along a horizontal straight line. Fig.2(a) reports the scores of FoM , F , EMM and \mathcal{M} . Scores of $Dice$ and d_4 are not reported because they have obvious discontinuities and are very sensitive

to small displacements (see [12]). Also, FoM and F measures are weak to small displacements, as \mathcal{M} with $\mu_{FP}=0.1$ and $\mu_{FN}=0.2$; moreover, as EMM , they are not monotonous (contrary to \mathcal{M} with automatic parameters tied to D and Δ).

The second test is performed by rotating the control shape incrementally. The shape of the curve of the measure scores is expected to be roughly symmetric around 180° . Thus, FoM and F measures are very sensitive to small rotations and EMM does not penalize enough movements, whereas \mathcal{M} considering Δ or D parameters obtains consistent scores.

The last experiment concerning synthetic data concerns scaling up of the object shape with the maximal scale 8 times the original size. However, the curve tied to EMM has sharp discontinuities that exhibit its unstable response to scaling, because its responses are 0 without TP. Once no TP exists, EMM drops down to 0, with no score evolution for other scales. The scores of the FoM and F are very sensitive as soon the first change is applied with scores close to 0.2. Eventually, \mathcal{M} with automatic parameters Δ or D obtains desirable scores.

Experiments on color real images were also performed. In Figs. 3 (a) and (b), thin edges are extracted using the Canny edge detector ($\sigma = 1$) [14]. Here, by moving the camera, the aim is to determine when the object is at the desired position in the image using thin binary edges as features, the scores must converge to 1. During the movement, each frame may be corrupted by numerous FPs and the candidate object may contain FNs when the object is well positioned, as illustrated in Figs. 3 (e) and (f). The first video contains 289 frames, whereas the second video is composed of 116 frames (cf. Fig. 3). The ground truth corresponds to binary boundaries of the desired position of the object, represented in blue pixels in Figs. 3 (a) and (b). Green pixels represent TPs, red points are FPs, whereas blue pixels which are also G_t are tied to FNs. These features are dilated with a structural element of size 3×3 for a

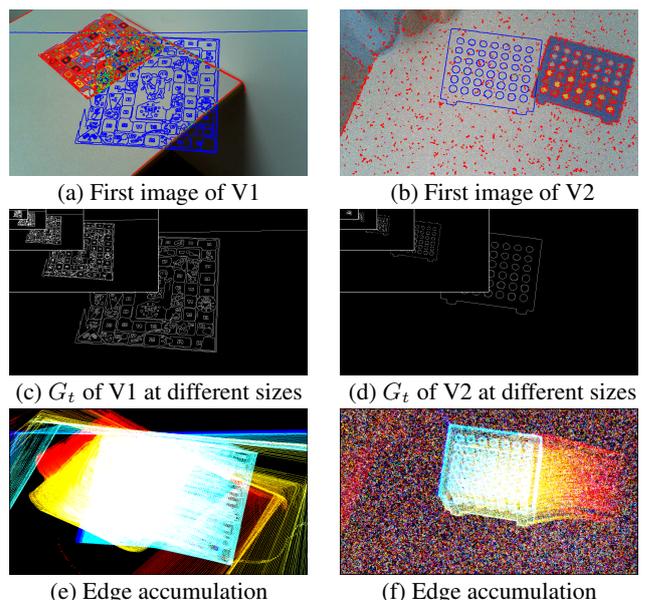


Figure 3. First images of videos 1 and 2 (V1 and V2) with their G_t at different sizes: original size (1280×720), $4 \times$, $16 \times$, $64 \times$ and $256 \times$ reduced (640×360 , 320×180 , 160×90 and 80×45).

better visualization. To test the robustness of the measure normalization, the same experiments are led using the parameters of Tab.1 and different sizes of the images see Figs.3 (c) and (d).

For the first video V1, object contours are easily extracted, with spurious undesirable points created by the table border and the camera rotation. Consequently, only edge points detected out of the candidate shape may disturb the contour-based localization. The object is moving to its desired position, until 150 frames (creating a bump curve), then it moves away before coming back to the final positioning, superposing G_t . Scores of the different measures are reported in Fig. 5 as a function of the image size. Curves tied to EMM and D_p are not significant. Also, $Dice$, F and d_4 scores converge only to 1 when the candidate object is close to the desired location. Only FoM and \mathcal{M} measures have the intended behavior for this sequence, even though FoM scores are globally noisy for small images.

The second video V2 is heavily corrupted by a random noise on the each color plane ($SNR \approx 11dB$). These disturbances create spurious pixels in the edge detection process, but especially, the edges of the candidate object are not well localized or not present. Therefore, in Fig. 6, the majority of measures does not evolve monotonously, but constantly for each image size. For the FoM , its scores increase at the end of the video, but scores are stochastic for small videos. The EMM measure converges rapidly, but, remains constant until the end.

The last experiment presented in Fig. 4 shows the importance of the parameters choice. FoM , F and d_4 measures are tested with $\kappa = \frac{1}{\Delta^2}$ which is similar to \mathcal{M} parameters. Such a value is totally unsuitable for these approaches in shape detection: FoM scores are close to 1 or 0 concerning d_4 . Also, F is decreasing using this parameter (excepted for the last frames), which is contrary to the researched assessment. Concerning \mathcal{M} , it behaves as expected, converging to 1 for each scale (Figs. 5 and 6). Eventually, the use of $\mu_{FP} = \frac{1}{\Delta^2}$ and $\mu_{FN} = \frac{1}{\Delta}$ parameters (instead of $\frac{1}{D}$ concerning particular shapes) allows to obtain $\mu_{FP} < \mu_{FN}$ for each scale, penalizing stronger the FNs compared to FPs in eq. 1, as demonstrated in [12].

5 Conclusion

A new approach for the measurement of a contour-based object pose is presented in this paper. This algorithm allows a supervised assessment of recognition and localization of known objects as a function of FPs and FNs distances. Experiments on real images exhibit the suitability of the approach for an object pose or shape matching estimation. Finally, the new measure is normalized and does not need any tuning parameter; this local-

ization measure may be useful for visual servoing processes.

References

- [1] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *IEEE ICPR*, vol. 1, pp. 566–568, 1994.
- [2] 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, p. 2, 2008.
- [3] B. Magnier, "Edge detection: a review of dissimilarity evaluations and a proposed normalized measure," *Multimedia Tools and Applications*, pp. 1–45, 2017.
- [4] 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, p. 74, 2018.
- [5] 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.
- [6] A. Pinho and L. Almeida, "Edge detection filters based on artificial neural networks," in *ICIAP*, pp. 159–164, Springer, 1995.
- [7] A. G. Boaventura and A. Gonzaga, "Method to evaluate the performance of edge detector," in *Brazilian Symp. on Comput. Graph. Image Process.*, pp. 234–236, 2006.
- [8] 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.
- [9] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. of Elec. imaging*, vol. 13, no. 1, pp. 146–166, 2004.
- [10] W. Crum, O. Camara, and D. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. on Med. Imaging*, vol. 25, no. 11, pp. 1451–1461, 2006.
- [11] H. Abdulrahman, B. Magnier, and P. Montesinos, "From contours to ground truth: How to evaluate edge detectors by filtering," *J. WSCG*, vol. 25, no. 2, pp. 133–142, 2017.
- [12] B. Magnier and H. Abdulrahman, "A study of measures for contour-based recognition and localization of known objects in digital images," in *IEEE IPTA*, pp. 1–6, 2018.
- [13] B. Magnier, "An objective evaluation of edge detection methods based on oriented half kernels," *ICISP*, pp. 80–89, 2018.
- [14] J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, no. 6, pp. 679–698, 1986.

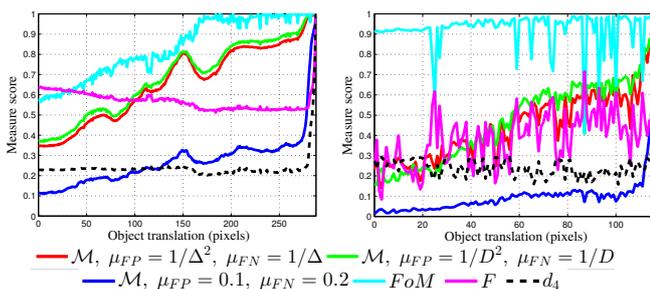


Figure 4. Score evolution in V1 and V2 for $\kappa = 1/\Delta^2$.

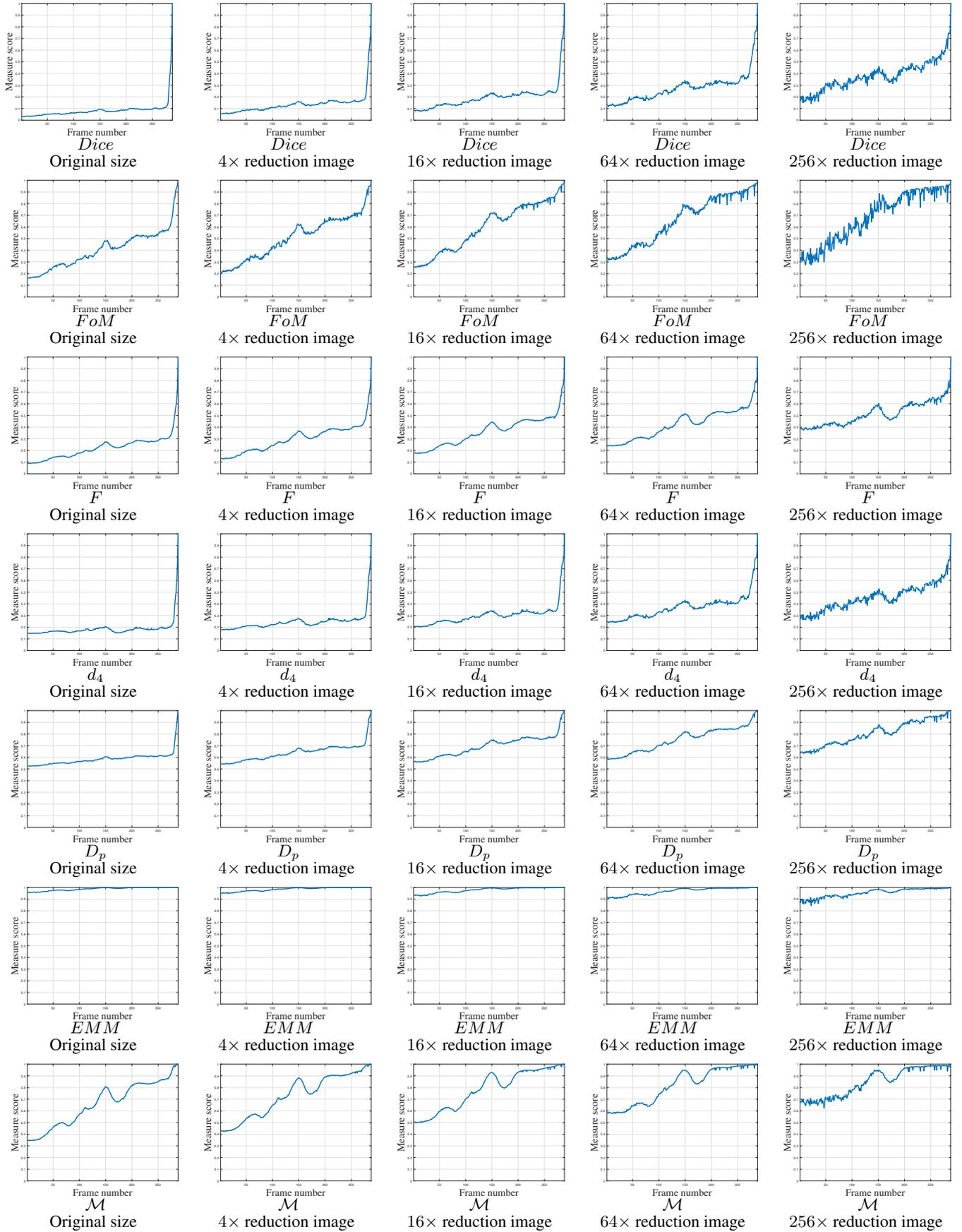


Figure 5. Localization metrics behaviors on real experiment concerning video 1 (V1) of 189 frames. The parameter concerning *FoM*, *F*, *d₄* and *D_p* is $\kappa = 1/9$. Concerning *M*, the parameters are $\mu_{FP} = 1/\Delta^2$ and $\mu_{FN} = 1/\Delta$, so $\mu_{FP} < \mu_{FN}$.

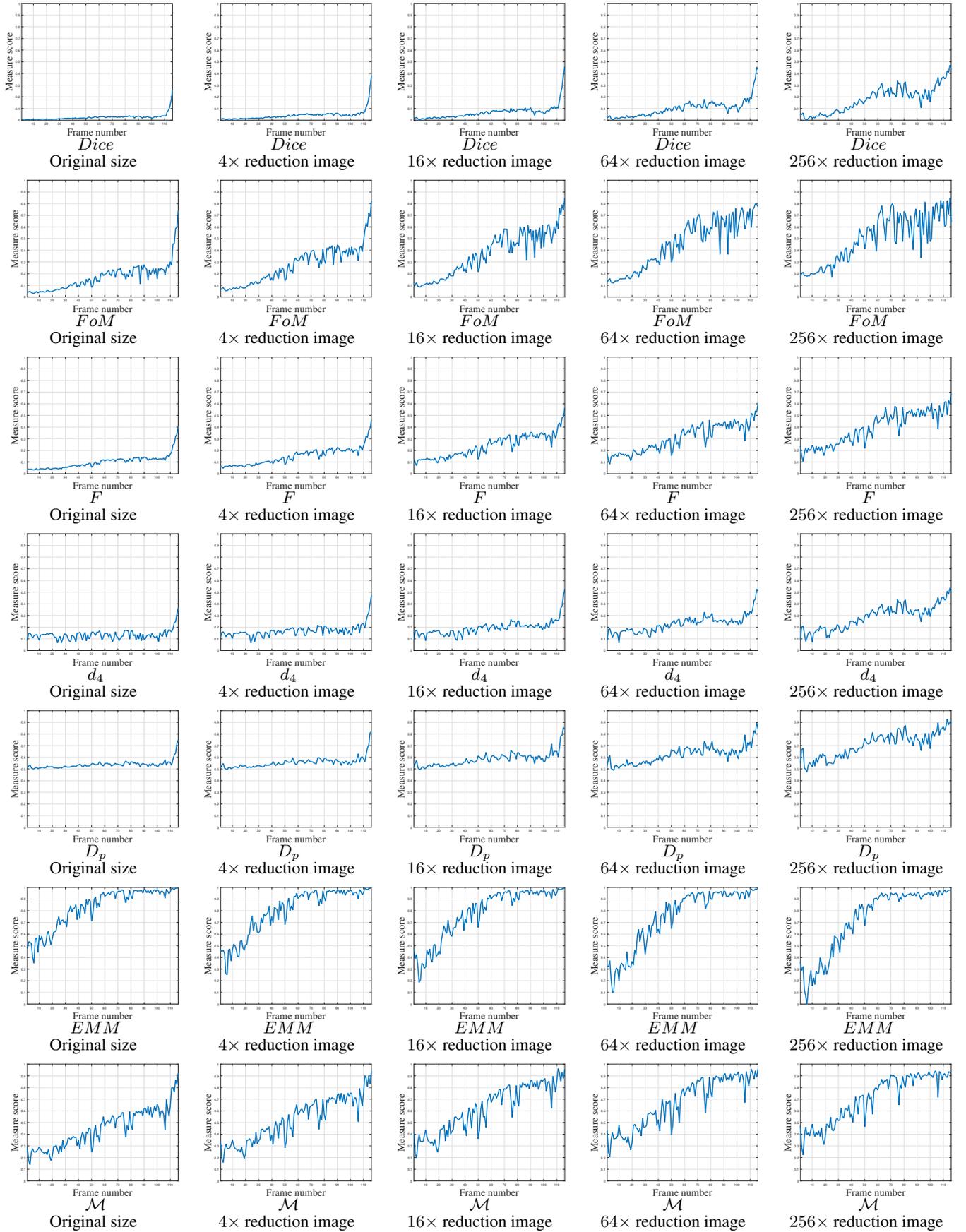


Figure 6. Localization metrics behaviors on real experiment concerning video 2 (V2) of 116 frames. The parameter concerning *FoM*, *F*, *d₄* and *D_p* is $\kappa = 1/9$. Concerning *M*, the parameters are $\mu_{FP} = 1/\Delta^2$ and $\mu_{FN} = 1/\Delta$.