

Étude comparative de métriques pour l'évaluation de la localisation d'objets par des boîtes englobantes.

Baptiste HEMERY, Hélène LAURENT, Christophe ROSENBERGER

Laboratoire Vision et Robotique - UPRES EA 2078
 ENSI de Bourges - Université d'Orléans
 88 boulevard Lahitolle, 18020 Bourges Cedex France
 baptiste.hemery@ensi-bourges.fr

Résumé – Ce papier traite d'une étude comparative de métriques permettant l'évaluation de résultats fournis par des algorithmes de localisation d'objets basée sur des boîtes englobantes. Nous avons étudié 26 métriques dans ce papier. Un protocole est présenté pour la création des vérités terrain et des résultats synthétiques simulant plusieurs erreurs de localisation (translation, changement d'échelle, rotation). Cela nous permet d'étudier le comportement des métriques face aux altérations considérées. Les résultats expérimentaux illustrent la pertinence des différentes métriques face à différents types d'erreur.

Abstract – This paper deals with a comparative study of metrics allowing the evaluation of results provided by object localization algorithms. We particularly focus on localization by the bounding box representation. 26 metrics are studied in this paper. A protocol is presented for the creation of ground truths and synthetic results of localization algorithms. These synthetic results permit to simulate several errors (translation, scale errors...) and to study the metrics behaviors face to the considered alterations. Experimental results illustrate the reliability of the different metrics face to different kinds of error.

1 Introduction

Les algorithmes de traitement d'images regroupent un ensemble de méthodes qui vont traiter l'image depuis son acquisition par un capteur (webcam, satellite, échographe...) jusqu'à l'extraction de l'information utile pour l'application (détection d'un objet particulier, mesure quantitative...). Parmi ces algorithmes, certains ont pour but de détecter et localiser un objet dans une image. Du fait des enjeux liés à ces informations, il est particulièrement important que les résultats fournis par les algorithmes d'interprétation d'images soient le plus juste possible.

Des compétitions tels que le Pascal VOC Challenge [1] ou bien le projet Robin [2] ont alors vu le jour afin de déterminer, notamment, les meilleurs algorithmes de localisation pour des applications données. Pour cela, il est nécessaire de disposer de méthodes d'évaluation afin de juger de l'efficacité de ces algorithmes.

Cette localisation peut être effectuée par les algorithmes de plusieurs façons : centre de l'objet, boîte englobante, contour de l'objet ou bien masque. Si l'évaluation de la localisation du centre d'un objet peut être faite simplement par le calcul d'une distance euclidienne, il n'en est pas de même pour les autres méthodes de localisation. La méthode de localisation la plus fréquemment utilisée étant les boîtes englobantes, ce papier traite de l'évaluation de la localisation d'objets par ce type de méthode.

L'objectif de ce papier est d'observer le comportement de différentes métriques de localisation d'objets face à des situations simulant le résultat d'un algorithme de détection d'objets. Ce papier est divisé en trois parties. Tout d'abord, la première partie traite du protocole utilisé pour notre étude comparative. Ensuite, nous présentons l'inter-

prétation des résultats que nous avons obtenus dans la seconde partie. Enfin, nous concluons et proposons différentes perspectives.

2 Protocole

Cette étude comparative s'appuie sur l'utilisation d'images synthétiques. Nous avons donc créé un ensemble d'images correspondant à différentes configurations de boîtes englobantes (boîte carrée, rectangulaire, limitrophe à un bord de l'image...). Nous avons ensuite appliqué un ensemble d'altérations à ces vérités terrain afin de simuler le résultat d'un algorithme. On évalue ensuite pour 26 critères dans quelle mesure ils pénalisent ces altérations.

2.1 Métriques

Les algorithmes de localisation d'objets peuvent donner des résultats sous différentes formes. Les résultats peuvent tout d'abord être donnés sous la forme d'une boîte englobant l'objet localisé, représentée sous la forme de deux couples de coordonnées $\{x_{min}, y_{min}\}, \{x_{max}, y_{max}\}$. Le projet Robin [2] a pour but l'évaluation d'algorithmes fournissant ce type de données. Afin de permettre ce type d'évaluation, trois métriques ont été développées :

$$RobLoc(Z_l, Z_{gt}) = \frac{2}{\pi} \arctan(\max(\frac{|x_l - x_{gt}|}{w_{gt}}, \frac{|y_l - y_{gt}|}{h_{gt}})) \quad (1)$$

$$RobCor(Z_l, Z_{gt}) = \frac{2}{\pi} \arctan(|\frac{h_l}{w_l} - \frac{h_{gt}}{w_{gt}}|) \quad (2)$$

$$RobCom(Z_l, Z_{gt}) = \frac{|A_l - A_{gt}|}{\max(A_l, A_{gt})} \quad (3)$$

où Z_l est la sortie de l'algorithme de localisation, $\{x_l, y_l\}$ sont les coordonnées du centre de la boîte englobante, A_l

est l'aire couverte par la boîte englobante et $\{h_l, w_l\}$ sont la hauteur et la largeur de la boîte englobante. Les variables indicées par $_{gt}$ correspondent aux mêmes variables pour la vérité terrain. Ces trois métriques évaluent des caractéristiques différentes des boîtes englobantes : RobLoc évalue la localisation du centre de la boîte englobante, RobCor évalue le rapport hauteur/largeur de la boîte englobante et RobCom évalue la taille de la boîte englobante.

Il y a deux autres types de localisation d'objets possibles, une représentant le contour de l'objet localisé, l'autre représentant la région occupée par cet objet. L'aspect région est utilisé, par exemple, par la métrique utilisée dans le Pascal VOC Challenge [1]. La métrique utilisée ici est assez simple :

$$PAS(I_l, I_{gt}) = \frac{Card(I_{gt}^R \cap I_l^R)}{Card(I_{gt}^R \cup I_l^R)} \quad (4)$$

avec $I_{gt}^R \cap I_l^R$ qui correspond aux pixels correctement localisés et $I_{gt}^R \cup I_l^R$ à l'ensemble des pixels appartenant soit à la vérité terrain soit à l'objet localisé. Il est possible d'étendre cette métrique au cas particulier de l'évaluation de la localisation par des boîtes englobantes. Dans ce cas, la région aura une forme rectangulaire.

L'aspect contour est utilisé dans la mesure proposée par Pratt [3] (Figure of Merit). Cette mesure, établie empiriquement, est une distance entre le contour de l'objet dans la vérité terrain I_{gt} et celui de l'objet localisé I_l .

$$FOM(I_l, I_{gt}) = \frac{1}{MP} \sum_{k=1}^{Card(I_l^{Cont})} \frac{1}{1 + \alpha * D(k)} \quad (5)$$

avec

$$D(k) = d^2(Card(I_l^{Cont}(k)), Card(I_{gt}^{Cont}))$$

où I_l^{Cont} correspond aux pixels du contour de l'objet localisé, MP correspond à $Max(Card(I_{gt}^{Cont}), Card(I_l^{Cont}))$, α est une constante de normalisation afin que le résultat soit compris entre 0 et 1, et $d(x, I) = \min_{y \in I} d(x, y)$ la distance euclidienne. Ce type de métrique est indépendant de la forme du contour utilisé. Dans le cadre de cette évaluation, le contour sera un rectangle.

Nous avons testé en tout 26 critères, venant pour la plupart de l'évaluation de la segmentation, listés dans le tableau 1.

TAB. 1 – Liste des métriques utilisées.

Nom	Type	Réf.	Nom	Type	Réf.
ErrLoc	Contour	[4]	Hamming	Région	[4, 9]
ErrSur	Contour	[4]	Yasnoff1	Région	[4, 10]
ErrSous	Contour	[4]	Yasnoff2	Région	[4, 10]
RMS	Contour	[4, 5]	Martin1	Région	[4, 11]
Lq	Contour	[4, 5]	Martin2	Région	[4, 11]
DBh	Contour	[4, 6]	Vinet	Région	[4, 12]
DKu	Contour	[4, 6]	Hafiane	Région	[13]
DJe	Contour	[4, 6]	Pascal	Région	[1]
ND	Contour	[4, 7]	Hen1	Région	[14]
ODI	Contour	[4, 7]	Hen2	Région	[14]
UDI	Contour	[4, 7]	RobLoc	Boîtes	[2]
Hausdorff	Contour	[4, 8]	RobCor	Boîtes	[2]
Pratt	Contour	[4, 3]	RobCom	Boîtes	[2]

2.2 Vérités terrain

Les images que nous avons créées sont des images de 256*256 pixels. Nous avons placé dans ces images une boîte englobante qui va correspondre à notre vérité terrain. Plusieurs cas sont étudiés. Tout d'abord, nous nous sommes intéressés à des boîtes englobantes de différentes tailles et de différentes formes. Nous avons alors obtenu les 8 vérités terrain présentées dans la figure 1.

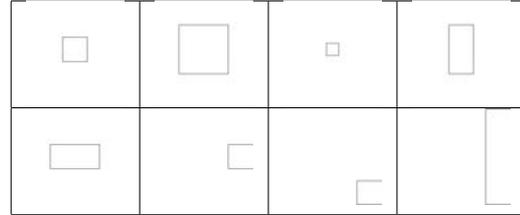


FIG. 1 – Vérités terrain

2.3 Résultats synthétiques

Nous avons ensuite appliqué un ensemble d'altérations aux vérités terrain : une translation, un changement d'échelle et une rotation.

La translation a été effectuée suivant trois directions : l'axe vertical, l'axe horizontal ou bien sur la diagonale. Nous avons appliqué une translation allant de -16 à +16 pixels. De même, le changement d'échelle a été imposé soit suivant l'axe vertical, l'axe horizontal ou bien les deux à la fois (cas de l'homothétie). Nous avons également appliqué un changement d'échelle allant de -16 pixels (réduction) à +16 pixels (agrandissement). Enfin, nous avons simulé une rotation de la boîte englobante de la vérité terrain puis nous avons calculé la nouvelle boîte englobante. Nous avons fait varier l'angle de la rotation de -10° à +90° par pas de 5°. Nous pouvons voir un exemple de chacune de ces transformations dans la figure 2. Les lignes en pointillés correspondent aux vérités terrain et les lignes pleines correspondent aux boîtes englobantes altérées. Nous avons alors obtenu, pour ces 8 vérités terrain, 1696 résultats synthétiques.

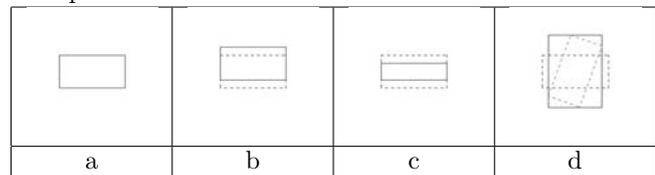


FIG. 2 – Altérations appliquées à la 5^evérité terrain (a : Vérité terrain, b : Translation selon l'axe vertical, c : Changement d'échelle selon l'axe vertical, d : Rotation)

3 Résultats expérimentaux

Nous avons utilisé les critères ci-dessus pour évaluer les résultats synthétiques par rapport aux vérités terrain que nous avons créées. Les critères ont été normalisés afin de pouvoir les comparer : le résultat optimal obtient une note de 0 et les altérations doivent produire des notes d'autant plus grandes que l'altération est importante.

La figure 3 montre les résultats obtenus pour une translation sur l'axe vertical sur la 5^evérité terrain. Nous pouvons tout d'abord voir que ces résultats sont symétriques, ce qui indique que les critères utilisés notent identiquement

une translation vers le haut ou vers le bas de l'image. On peut voir que ErrLoc, ErrSous, ErrSur, Root Mean Square (RMS) et les distances L_q , de Bhattacharyya (DBh), de Küllback (DKu) et de Jensen (DJe) donnent de mauvais résultats. En effet, ces métriques pénalisent très fortement et indifféremment une translation d'un seul ou de 16 pixels. Le critère de Pratt évolue de façon assez similaire mais relâche toutefois les contraintes imposées. Au contraire les métriques de Odet (ODI, UDI, ND) donnent des résultats qui évoluent linéairement par rapport à la translation : une translation de 10 pixels obtient une note deux fois moins bonne qu'une translation de 5 pixels. Enfin, la distance de Hausdorff permet quant à elle d'avoir une tolérance de quelques pixels tout en pénalisant fortement les translations plus importantes. Les critères basés sur la région évoluent tous linéairement comme les métriques de Odet, à l'exception de la première métrique de Henricsson (Hen1) qui ne pénalise pas du tout la translation. Les 3 critères de la compétition Robin évoluent comme prévu : RobLoc pénalise linéairement la translation tandis que RobCor et RobCom ne pénalisent pas ce type d'altération.

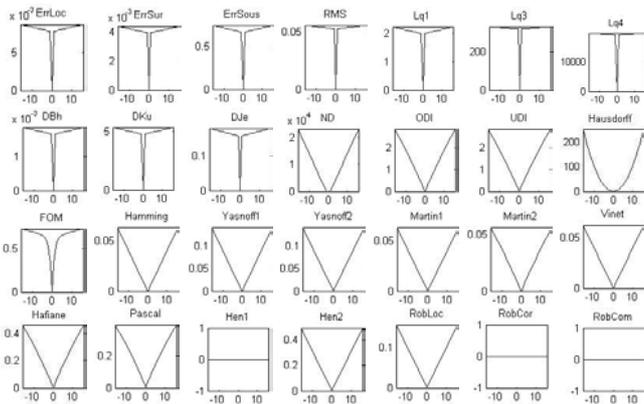


FIG. 3 – Résultats des critères sur la 5^e vérité terrain suivant une translation sur l'axe vertical

Les résultats obtenus sur l'horizontale et la diagonale sont similaires aux résultats précédents à l'exception des métriques suivantes : ErrLoc, ErrSous, ErrSur, RMS et les distances L_q , Bhattacharyya, Küllback et Jensen. Nous pouvons voir sur la figure 4, dans le cas de la distance L_1 , que les translations sur la diagonale sont pénalisées de manière identique pour 1 ou 16 pixels. De plus, la distance L_1 pénalise plus fortement une translation d'un pixel sur la diagonale qu'une translation de 16 pixels sur l'horizontale ou la verticale. Les résultats obtenus sur les autres vérités terrain sont semblables à un facteur d'échelle près.

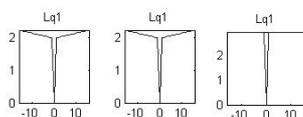


FIG. 4 – Résultat de la distance L_1 sur la 5^e vérité terrain pour une translation selon l'horizontale, la verticale et la diagonale

Nous pouvons voir sur la figure 5 les résultats obtenus dans le cas d'une mise à l'échelle. Ces résultats sont as-

sez similaires aux précédents. Les métriques basées sur le contour, ErrLoc, ErrSous, ErrSur, RMS et les distances L_q , Bhattacharyya, Küllback et Jensen, obtiennent de mauvais résultats en pénalisant fortement de petites altérations. Nous pouvons voir que le critère de Pratt pénalise plus la réduction que l'agrandissement. Les métriques d'Odet évoluent linéairement et symétriquement, c'est-à-dire pénalisent autant la réduction que l'agrandissement. Les métriques basées sur la région obtiennent des résultats globalement meilleurs. La distance de Hamming, la première métrique de Martin, le critère de Vinet et les métriques de Henricsson évoluent linéairement et symétriquement. La seconde métrique de Martin, tout comme la distance de Hausdorff, évolue symétriquement mais non linéairement. Nous pouvons voir que les deux métriques de Yasnoff évoluent différemment, l'une pénalise fortement la réduction tandis que l'autre pénalise fortement l'agrandissement. La métrique de Hafiane et celle utilisée dans la compétition Pascal pénalisent légèrement plus la réduction que l'agrandissement. Enfin, les trois métriques du projet Robin évoluent comme attendu : RobLoc ne pénalise pas du tout les changements d'échelle, alors que RobCor et RobCom les pénalisent. Cependant, la réduction est plus pénalisée que l'agrandissement.

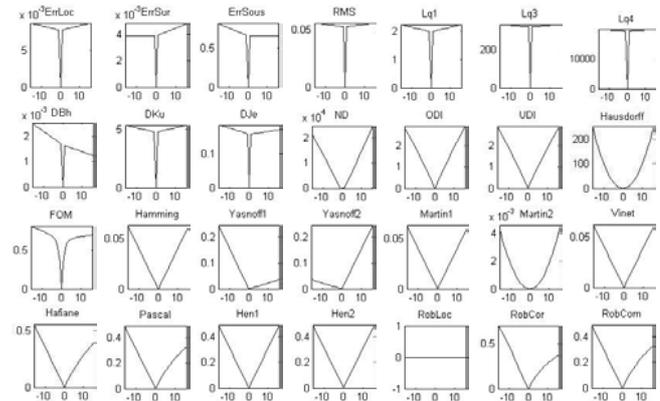


FIG. 5 – Résultats des critères sur la 5^e vérité terrain suivant un changement d'échelle sur l'axe vertical

Nous pouvons voir sur la figure 6 les résultats lorsque l'altération est une rotation. Là encore, les métriques basées sur le contour ne donnent pas de bons résultats, à l'exception de deux métriques d'Odet et de la distance de Hausdorff. Nous pouvons voir que les métriques du projet Robin réagissent toutes différemment. Robloc ne pénalise presque pas ce type d'altération, le centre de la boîte englobante ne bougeant presque pas. RobCor évalue le rapport hauteur/largeur et donne le plus mauvais résultat pour 90° tandis que RobCom, qui pénalise la superficie occupée par la boîte, donne le plus mauvais résultat pour 45°. Les métriques basées région semblent être un compromis entre ces deux métriques. Certaines, Yasnoff1 et Martin2, évoluent plutôt comme le critère RobCor tandis que d'autres, la distance de Hamming ou bien Yasnoff2, évoluent plutôt comme RobCom. La métrique de Hafiane et celle utilisée dans la compétition Pascal semblent être le meilleur compromis pénalisant aussi bien une rotation de 45° que de 90°.

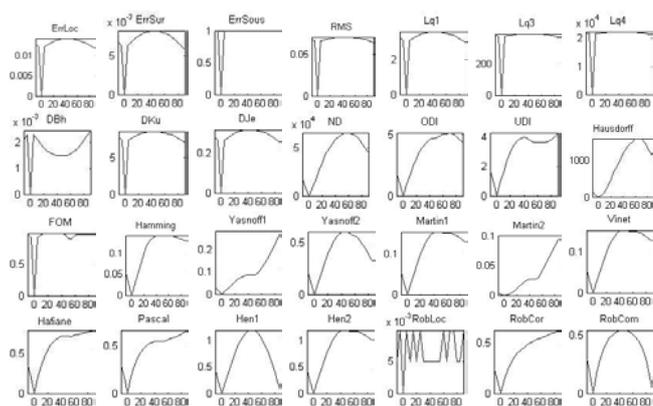


FIG. 6 – Résultats des critères sur la 5^e vérité terrain suivant une rotation

Les résultats obtenus sont récapitulés dans le tableau 2. Plus le nombre d'étoiles est important, plus le critère évolue de façon pertinente.

TAB. 2 – Synthèses des résultats obtenus : le nombre d'étoiles indique l'efficacité de la métrique

Alteration	Translation	Changement d'échelle	Rotation
ErrLoc	*	*	*
ErrSur	*	*	*
ErrSous	*	*	*
RMS	*	*	*
Lq	*	*	*
DBh	*	*	*
DKu	*	*	*
DJe	*	*	*
ND	***	***	**
ODI	***	***	**
UDI	***	***	**
Hausdorff	***	***	**
Pratt	**	**	**
Hamming	***	***	**
Yasnoff1	***	**	**
Yasnoff2	***	**	**
Martin1	***	***	**
Martin2	***	***	**
Vinet	***	***	**
Hafiane	***	***	***
Pascal	***	***	***
Hen1	*	***	**
Hen2	***	***	**
RobLoc	***	*	*
RobCor	*	**	***
RobCom	*	**	***

4 Conclusion et perspectives

Cette étude nous a permis de voir qu'un certain nombre de métriques n'étaient clairement pas utilisables dans le cas de l'évaluation de la localisation d'objets. Mise à part les critères d'Odet, de Pratt et la distance de Hausdorff, les métriques basées sur le contour ne donnent pas des résultats satisfaisants. Les métriques basées région donnent globalement de meilleurs résultats. Ainsi, en dépit de sa simplicité, le critère utilisé dans la compétition Pascal donne de bons résultats, tout comme la métrique de Hafiane,

dont elle est assez proche. Les trois métriques du projet Robin, utilisées conjointement, donnent de bons résultats car chaque métrique pénalise un attribut différent. Cependant, la manière dont sont combinées ces trois métriques peut poser problème.

Nos perspectives sont de procéder au même type d'étude comparative pour des vérités terrain et des altérations plus élaborées, par exemple lorsqu'il y a plusieurs objets complexes localisés non pas par des boîtes englobantes, mais par le contour réel de l'objet.

Références

- [1] M. Everingham, A. Zisserman, C. Williams, L. Van Gool, M. Allan, C. Bishop, O. Chapelle, N. Dalal, T. Deselaers, G. Dorko, et al. The 2005 pascal visual object classes challenge. 2005.
- [2] E. D'Angelo, S. Herbin, and M. Ratiéville. Robin challenge evaluation principles and metrics. 2006.
- [3] W. Pratt, O. D. Faugeras, and A. Gagalowicz. Visual discrimination of stochastic texture fields. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(11) :796–804, 1978.
- [4] C. Rosenberger, S. Chabrier, H. Laurent, and B. Emile. *Advances in Image and Video Segmentation*, chapter Un-supervised and supervised image segmentation evaluation, pages 365–393. Idea group Publishing, 2006.
- [5] D. Coquin, P. Bolon, and Y. Chehadeh. Evaluation quantitative d'images filtrées. *GRETSI97*, 2 :1351–1354, 1997.
- [6] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4) :349–369, 1989.
- [7] C. Odet, B. Belaroussi, and H. Benoit-Cattin, editors. *Scalable Discrepancy Measures for Segmentation Evaluation*, volume 1, septembre 2002.
- [8] M. Beauchemin, KP. B. Thomson, and G. Edwards. On the hausdorff distance used for the evaluation of segmentation results. *Canadian Journal of Remote Sensing*, 24(1) :3–8, 1998.
- [9] Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. *International Conference on Image Processing (ICIP'95) Washington DC, USA*, 3 :53–56, 1995.
- [10] W. A. Yasnoff, J. K. Mui, and J. W. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9 :217–231, 1977.
- [11] 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. *8th Int'l Conf. Computer Vision*, pages 416–423, July 2001.
- [12] L. Vinet. *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques*. PhD thesis, Université de Paris IX Dauphine, Juillet 1991.
- [13] Adel Hafiane. *Caractérisation de textures et segmentation pour la recherche d'images par le contenu*. PhD thesis, Université de Paris-Sud XI, December 2005.
- [14] O. Henricsson and E. Baltsavias. 3-d building reconstruction with aruba : A qualitative and quantitative evaluation. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pages 65–76, 1997.