

Reconnaissance par blocs : une alternative aux descripteurs locaux

Yannick BENEZETH¹, Bruno EMILE², H el ene LAURENT¹, Adel HAFIANE¹,

¹ENSI de Bourges, Institut PRISME (UPRES EA 4229), Universit e d'Orl eans
88 Boulevard Lahitolle, 18020 Bourges Cedex, France

²Institut PRISME (UPRES EA 4229), IUT de l'Indre, Universit e d'Orl eans
2 avenue F. Mitterrand, 36000 Ch ateauroux, France

yannick.benezeth@ensi-bourges.fr, bruno.emile@univ-orleans.fr
helene.laurent@ensi-bourges.fr, adel.hafiane@ensi-bourges.fr

R esum e – La probl ematique abord ee dans cet article est la reconnaissance d'objets dans une image. De nombreux travaux utilisent des descripteurs locaux afin de caract eriser un objet. Cela permet notamment de g erer les occultations partielles, les changements de fond et de pouvoir travailler avec une base d'apprentissage r eduite. Or, dans des contextes perturb es et notamment bruit es, les performances de ces descripteurs locaux restent faibles par rapport aux descripteurs globaux (qui travaillent sur l'image enti ere) comme les moments de Zernike. L'application des descripteurs globaux sur des imagerie (d ecoupage par blocs de l'image) doit alors permettre de conserver  a la fois les avantages des m ethodes globales et locales. L'objectif de cet article est de comparer cette technique aux approches locale et globale classiques. Les performances obtenues pour diff erentes tailles de la base d'apprentissage et lors de l'application de diff erentes perturbations sur l'image sont  egalement pr esent ees. La m ethode propos ee r ealise un bon compromis entre les deux approches en terme de simplicit e de calcul et performance de classification.

Abstract – This paper addresses the issue of the objects recognition in an image. Many works use the local descriptors to characterize an object. This allows to handle with occlusions, background changes, and to have a reduced learning database. However, in contexts including noise disturbance, the performances of these local descriptors remain low compared with global descriptors like the Zernike moments. In order to maintain the advantages of global and local approaches we split the image into regular blocks. Then, the global descriptor is applied locally to each block of the image. We compared this technique to the classical local and global approaches using different size of the learning database, and various disturbances on the images. The proposed methods realize a good trade off between computational complexity and classification performances.

1 Introduction

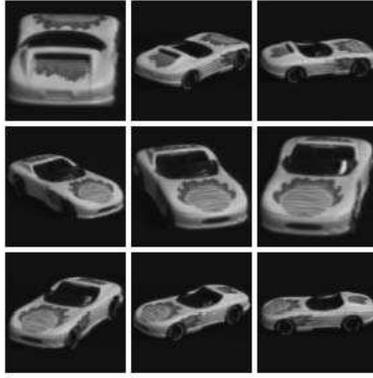
La reconnaissance d'objets dans une image concerne de nombreuses applications (vid eo surveillance, biom etrie, robotique mobile, ...) et a fait l'objet de concours scientifiques (Pascal, Robin [9]). Cette  tape a pour but de ranger l'objet d etect e dans une classe connue ou inconnue afin d'interpr eter une image. Classiquement, des descripteurs permettent de donner une signature  a un objet ou  a une classe d'objets et sont coupl es  a une m ethode de classification avec apprentissage qui permet de classer l'objet d etect e. Dans cet article, nous nous int eressons aux choix des descripteurs et  a la d efinition d'une m ethode de reconnaissance par blocs qui sera coupl ee dans notre  tude aux S eparateurs  a Vaste Marge (SVM) [8] pour la classification.

Les descripteurs actuels sont principalement utilis es localement autour de points d'int er et pr ealablement d etect es. On peut citer les m ethodes SIFT [4], SURF [10], HOG [11]. Dans une pr ecedente  tude [6], nous nous sommes int eress es  a diff erents descripteurs invariants par transformations g eom etriques et  a leurs performances respectives lorsqu'ils  taient utilis es dans un contexte global (travaillant sur des images contenant l'objet

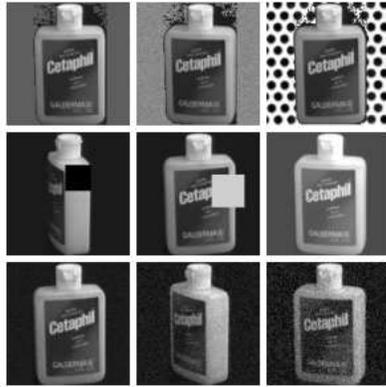
 a reconna tre en entier) ou dans un contexte local (travaillant autour de points d'int er ets). L'efficacit e des descripteurs locaux a  t e d emonstr ee, mais dans certains contextes (tel que la pr esence d'un bruit important), les descripteurs globaux  taient encore les meilleurs.

Afin de conserver les avantages des descripteurs locaux et globaux, nous proposons dans cet article de d ecouper l'image de l'objet en blocs r eguliers et de calculer sur chaque bloc les descripteurs. Cette m ethode par blocs, qui sera qualifi ee par la suite de m ethode semi-locale, sera compar ee aux m ethodes globale et locale en consid erant diff erentes tailles de la base d'apprentissage et diff erentes perturbations appliqu ees aux images.

Nous rappelons dans le paragraphe 2 le principe des m ethodes globale et locale d evelopp ees dans [6] ainsi que les principaux r esultats obtenus. Le paragraphe 3 est quant  a lui d edi e  a la m ethode semi-locale que nous proposons comme alternative aux deux approches pr ecedentes et  a la confrontation des performances obtenues dans diff erents contextes. Enfin, nous concluons cette  tude et nous int eressons aux perspectives en-



(a)



(b)

FIG. 3 – (a) : exemples de rotations et changements d'échelle, (b) : Quelques altérations

TAB. 1 – Taux de reconnaissance pour différentes tailles de la base d'apprentissage

Taille de la base d'apprentissage	25%	50%	75%
Zernike global	70,7	86,9	95,8
Zernike local	93,9	98,9	100

amène donc à rendre minoritaires ceux utilisés pour l'apprentissage.

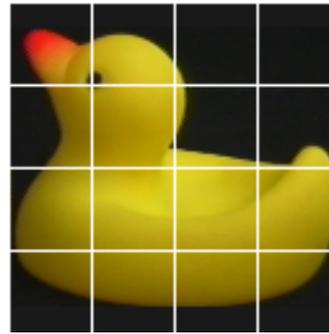
3 Méthode par blocs ou semi-locale

3.1 Principe

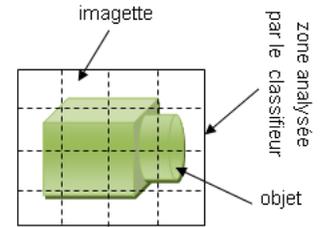
Afin de ne pas perdre l'apport de l'aspect local du calcul des descripteurs et d'y ajouter une certaine insensibilité aux calculs des points d'intérêt qui peuvent être très perturbés par du bruit, nous avons décidé de découper chaque image en blocs (imagettes) et de calculer les descripteurs de Zernike sur chaque imagette. La figure 4 présente le principe de la méthode semi-locale. Les imagettes auront des tailles variables et pour l'étude

TAB. 2 – Taux de reconnaissance face à différentes altérations

Altérations	Zernike global	Zernike local
fond uniforme	31,8%	86,2%
fond bruité	34,9%	63,5%
fond texturé	7,5%	55,2%
occultation noire	74,7%	79,1%
occultation niveau de gris	71,2%	81%
variance de luminance	95,9%	89%
bruit variance 5	100%	73,4%
bruit variance 10	100%	70,1%
bruit variance 20	100%	62,9%



(a)



(b)

FIG. 4 – Principe de la méthode semi-locale

faite dans cet article, il n'y a pas de chevauchement entre les imagettes.

3.2 Protocole d'étude et résultats

Un premier test se déroule de la façon suivante : un pourcentage de la base d'images est utilisé pour l'apprentissage, le reste est utilisé pour la phase de test. Chaque image de la base d'apprentissage est découpée en imagettes de taille identique (nous avons uniquement travaillé sur des imagettes carrées). Sur chaque imagette, on calcule les moments de Zernike. Pour la reconnaissance, un vote majoritaire est calculé sur chaque image : une image est attribuée à la classe désignée par le maximum d'imagettes la constituant. Comme pour l'étude exposée au 2.2, un classifieur SVM est utilisé avec les mêmes paramètres, le tirage aléatoire de la base d'apprentissage a été reproduit 10 fois puis les performances ont été moyennées. Les résultats du taux de reconnaissance de la méthode de Zernike semi-locale obtenus pour différentes tailles de la base d'apprentissage et pour différentes tailles d'imagettes sont présentés dans le tableau 3.

On constate que le taux de reconnaissance (en fonction de la taille de la base d'apprentissage) de la méthode semi-locale est proche des résultats de la méthode locale. La taille des imagettes ne semblent que peu influencer les résultats. On constate

TAB. 3 – Taux de reconnaissance de la méthode semi-locale pour différentes tailles de la base d'apprentissage et différentes tailles d'images

Taille de la base d'apprentissage	25%	50%	75%
Image 8x8	85,23	88,7	92,8
Image 16x16	89,87	95,3	95,8
Image 32x32	90,14	95,26	96,70
Image 64x64	90,32	96,37	97,01

toutefois qu'avec une petite taille des images, pour des objets petits, il y a de nombreuses images où l'objet n'est pas présent. Ceci peut expliquer les résultats moins bons obtenus pour la plus petite taille d'images, même en utilisant un vote majoritaire.

Une seconde série d'essais teste la robustesse de la méthode semi-locale face à différentes altérations de l'image. La taille des images est maintenant fixe (32x32). On regarde comment réagit la méthode semi-locale par rapport à différentes perturbations : différents fonds, des changements de luminosité, du bruit et des occultations. La base non altérée sert d'apprentissage et la base altérée sert de test. Le tableau 4 présente les taux de reconnaissance de la méthode semi-locale par rapport à différentes altérations et les compare par rapport aux meilleurs résultats obtenus soit par la méthode globale soit par la méthode locale.

TAB. 4 – Taux de reconnaissance de la méthode semi-locale face à différentes altérations

Altérations	Méthode semi-locale	Meilleurs résultats précédents
fond uniforme	83,73%	86,2%
fond bruité	53,8%	63,5%
fond texturé	45,3%	55,2%
occultation noire	85,3%	79,1%
occultation niveau de gris	86,3%	81%
variance de luminosité	87,7%	95,9%
bruit variance 5	98,25%	100%
bruit variance 10	96,7%	100%
bruit variance 20	95,5%	100%

Les résultats concernant la robustesse de la méthode semi-locale face à différentes altérations sont très intéressants. En effet la méthode semi-locale, même si elle ne donne pas toujours les meilleurs résultats, s'en approche fortement. Elle permet notamment d'améliorer les taux de reconnaissance des images bruitées obtenus par la méthode locale.

4 Conclusion

L'article a comparé trois approches différentes pour la reconnaissance d'objets dans une image, une approche globale, une approche semi-locale et une approche locale. Même si l'approche semi-locale ne donne pas toujours les meilleurs résultats, elle est un bon compromis permettant de conserver les avantages des approches globales et locales. En effet, elle permet d'obtenir des résultats proches des meilleurs résultats dans pratiquement tous les cas de figure étudiés.

Le calcul des points d'intérêt est une étape non négligeable en temps de calcul. De plus le nombre de points varie d'une image à une autre. Le découpage en images est quasi instantané avec un nombre de blocs toujours fixe. Ceci rend la technique proposée plus adaptée aux applications temps réel.

Références

- [1] F. Ghorbel, *A complete invariant description for gray-level images by the harmonic analysis*, In Pattern Recognition letters, vol. 15, 1994.
- [2] M.K. Hu, *Visual pattern recognition by moments invariants*, In IEEE Transaction Information Theory, vol. 8, 1962.
- [3] A. Khotanzad et Y.H. Hong, *Invariant image recognition by Zernike moments*. In IEEE Transaction PAMI, 12(5), 1990.
- [4] D. Lowe, *Distinctive image features from scale-invariant keypoints*. In international Journal of Computer Vision, 60(2), 2004.
- [5] C. Harris et M. Stephens, *A combined corner and edge detector*. In Proc. 4th Alvey Vision Conf., Manchester, pages 147-151, 1988.
- [6] A. Choksuriwong, B. Emile, C. Rosenberger, H. Laurent, *Comparative study of invariant descriptors for object recognition*. In Journal of Electronic imaging, 17(2), 2008.
- [7] <http://www1.cs.columbia.edu/cave/research/softlib/coil-100.html>
- [8] C. Cortes et V. Vapnik, *Support-Vector Networks*. In Machine Learning, 1995
- [9] <http://robin.inrialpes.fr/>
- [10] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, *SURF : Speeded Up Robust Features*. In Computer Vision and Image Understanding, 110(3), 2008
- [11] B. Verma, M. Blumenstein, *Pattern Recognition Technologies and Applications : Recent Advances*. Edition : illustrated, 2008.