

Représentation éparse d'images par ondelettes fovéales : principes et applications

Julien ROS, Christophe LAURENT

France Telecom R&D - TECH/IRIS/CIM
4, rue du Clos Courtel, 35512 Cesson Sévigné Cedex, FRANCE
{julien.ros|christophe2.laurent}@francetelecom.com

Résumé – Aujourd'hui, les méthodes de représentation locale d'images ont montré leur efficacité dans de nombreux problèmes de vision par ordinateur (indexation d'images, reconnaissance de visages, catégorisation d'objets, etc.). Dans ces méthodes, les images ne sont plus considérées dans leur globalité mais seules quelques zones saillantes (i.e. perceptuellement importantes) sont prises en compte pour représenter le contenu visuel. L'extraction de caractéristiques pertinentes et discriminantes au sein de ces zones est alors essentielle. Dans le cas particulier où ces zones sont assimilées à des points d'intérêt et en observant que ces points d'intérêt sont situés sur les singularités de l'image, il peut être intéressant de décrire de façon fiable et robuste ces singularités. Dans ce papier, nous présentons un nouveau descripteur local utilisant la décomposition sur une base d'ondelettes fovéales permettant de caractériser l'orientation et la régularité des singularités présentes dans une zone saillante. Nous montrons, par le biais de deux applications phares (classification supervisée d'images naturelles et reconnaissance de visages) que ce descripteur obtient des résultats encourageants et est même capable de supplanter les descripteurs de l'état de l'art généralement utilisés pour caractériser les singularités.

Abstract – Nowadays, methods based on a local image representation have shown their efficiency in many computer vision tasks such as image indexing, face recognition or object categorization. In these methods, the images are no more considered as a whole but only a limited set of salient zones are used to represent the visual content. Feature extraction in these regions are then a key step in order to provide a robust and discriminative image signature. In the particular case where the salient zones are represented by points of interest and by observing that these points are located on image singularities, it can be interesting to describe these singularities. In this paper, we propose a new local descriptor that uses the decomposition of the signal on a foveal wavelet basis allowing to describe the orientation and regularity of singularities appearing in a salient region. We then show, by implementing two key applications (natural image classification and face recognition) that the proposed descriptor reaches good results and is even able to outperform classical descriptors generally used to describe singularities.

1 Introduction

Les travaux psychovisuels de Biederman [1] ont montré que le système visuel humain n'utilise pas toute l'information contenue dans une image pour l'analyser. S'inspirant de ces observations, de nombreux problèmes de vision par ordinateur (indexation d'images, reconnaissance d'objets et de visages, etc.) ont été traités en représentant les images de façon éparse c'est-à-dire en se concentrant uniquement sur quelques zones considérées comme perceptuellement importantes [5, 8, 9, 15]. Contrairement aux approches globales pour lesquelles une signature unique est calculée en considérant tous les pixels de l'image avec la même importance, ces approches, qualifiées d'*approches locales*, représentent le contenu de l'image par un ensemble de signatures locales calculées au voisinage de points saillants (i.e. perceptuellement importants).

La détection de ces points a fait l'objet de nombreuses recherches et de nombreuses solutions existent [2, 6, 8]. Comme le montre la figure 1, les points saillants détectés ont la particularité d'être généralement localisés dans les zones de fort contraste et donc sur les singularités de l'image.

A partir de cette observation, il semble alors intéressant de décrire et d'utiliser l'information portée par ces singularités dans le cadre d'un système de reconnaissance basé sur les points saillants. En effet, l'information essentielle d'un signal mono-

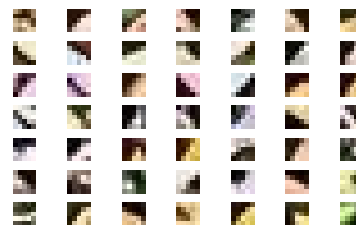


FIG. 1 – Régions de taille 7x7 centrées sur les points saillants

dimensionnel ou bidimensionnel se trouve souvent contenue dans les structures singulières et la description de celles-ci a déjà été envisagée dans [12, 13]. De plus, les ondelettes de Gabor [4], les jets locaux [15] ainsi que le descripteur SIFT [10] s'intéressent également à la caractérisation des singularités. Dans cet article, nous proposons d'utiliser la théorie des ondelettes fovéales développée dans [11] afin de caractériser l'orientation et la régularité des singularités présentes dans une région d'intérêt centré sur un point saillant. Cette description sera ensuite utilisée dans deux types d'applications : la classification supervisée d'images naturelles et la reconnaissance de visages.

2 Description fovéale des singularités

2.1 Rappels sur les singularités

L'étude de la régularité d'une fonction est très importante puisque c'est souvent les singularités de celle-ci qui portent l'information essentielle (les contours dans une image par exemple). Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est dite singulière en $x_0 \in \mathbb{R}$ si elle n'est pas dérivable en x_0 . Afin de mesurer précisément la régularité d'une fonction en tout point de son domaine de définition, on mesure sa régularité Lipschitzienne en chacun de ces points. Cette notion est appelée dans la définition suivante :

Définition 1 $f : [a, b] \rightarrow \mathbb{R}$ est Lipschitzienne $\alpha > 0$ en $x_0 \in [a, b]$ si $\exists K > 0, \delta > 0$ et un polynôme P de degré $m = \lfloor \alpha \rfloor$ tel que :

$$\forall x, x_0 - \delta \leq x \leq x_0 + \delta, |f(x) - P(x - x_0)| \leq K|x - x_0|^\alpha$$

Définition 2 L'exposant de Lipschitz $h_f(x_0)$ de f en x_0 est la borne supérieure des α ($h_f(x_0) = \sup\{\alpha, f \text{ est Lipschitzienne } \alpha \text{ en } x_0\}$).

La régularité locale d'une fonction en un point x_0 est donc caractérisé par son exposant de Lipschitz noté $h_f(x_0)$. De plus, f est dite singulière en x_0 si f n'est pas Lipschitzienne d'ordre 1 en x_0 .

2.2 Les ondelettes fovéales

Les ondelettes fovéales ont été introduites récemment par Mallat dans [11] et sont constituées des dilatations dyadiques d'une ondelette symétrique $\psi_{x_0}^1$ et d'une ondelette antisymétrique $\psi_{x_0}^2$ centrées sur un point x_0 que l'on nomme la fovea. Elles fournissent une base orthonormale aux espaces fovéaux qui sont constitués de fonctions dont la résolution décroît linéairement lorsque l'on s'éloigne de la fovea. Leur théorie s'inspire du fait que la vision humaine est plus précise au niveau de la fovea. Ces bases d'ondelettes fovéales peuvent être obtenues à partir de fonctions splines comme illustré sur la figure 2.

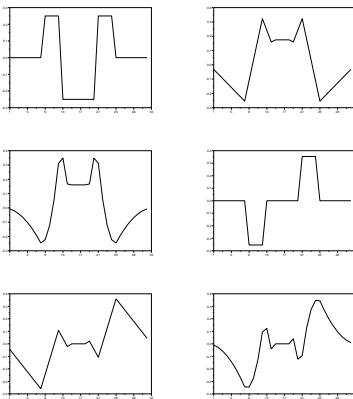


FIG. 2 – Bases d'ondelettes fovéales obtenues à partir de splines d'ordre 0, d'ordre 1 et d'ordre 2.

Il a été montré dans [11] que ces ondelettes sont particulièrement bien adaptées pour reconstruire et caractériser les singularités d'une fonction monodimensionnelle. En effet, le théorème suivant a été démontré dans [11] :

Théorème 1 La fonction $f : [a, b] \rightarrow \mathbb{R}$ est Lipschitzienne α en $x_0 \in [a, b]$ si et seulement si $|\langle f, \psi_{j,x_0}^k \rangle| = O(2^{j(\alpha+1/2)})$ où ψ_{j,x_0}^k désigne l'ondelette ψ^k ($k \in \{1, 2\}$) dilatée d'un facteur j et translatée en x_0 ($\psi_{j,x_0}^k(x) = 2^{-\frac{j}{2}}\psi^k(2^{-j}(x - x_0))$).

La régularité Lipschitzienne α de f en x_0 est alors obtenue par régression linéaire sur j de :

$$\frac{1}{2} \log_2(|\langle f, \psi_{j,x_0}^1 \rangle|^2 + |\langle f, \psi_{j,x_0}^2 \rangle|^2) \quad (1)$$

L'évolution de la quantité $|\langle f, \psi_{j,x_0}^1 \rangle|^2 + |\langle f, \psi_{j,x_0}^2 \rangle|^2$ avec les échelles donne donc une information importante sur la régularité Lipschitzienne locale d'une fonction f en x_0 .

Nous avons donc choisi d'utiliser cette quantité afin de caractériser une région d'intérêt. Toutefois, la théorie des ondelettes fovéales étant applicable dans le cadre d'un signal monodimensionnel, il convient de l'adapter pour l'utiliser dans le cadre de la caractérisation de l'orientation et de la régularité des singularités présentes dans une région d'intérêt, qui est par construction bi-dimensionnelle.

2.3 Le descripteur fovéal

Le descripteur fovéal élaboré dans ce papier permet de caractériser l'orientation et la régularité des singularités présentes dans le voisinage d'un point saillant. Pour ce faire, N signaux 1D f_θ^i ($i \in \{1, \dots, N\}$) d'orientation θ sont extraits au voisinage de chaque point saillant grâce à l'algorithme de Bresenham [3]. Cette procédure est illustrée sur la figure 3.

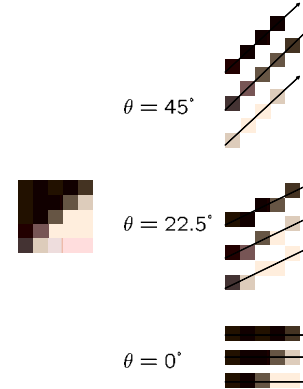


FIG. 3 – Extraction de lignes d'orientation θ par l'algorithme de Bresenham.

Chaque point x_0 de ce voisinage est alors représenté par un ensemble d'énergies fovéales $E_{\theta,j}^i(x_0)$ obtenues en translatant les ondelettes fovéales en x_0 et en les dilatant d'un facteur j . Ces énergies sont définies par :

$$E_{\theta,j}^i(x_0) = |\langle f_\theta^i, \psi_{j,x_0}^1 \rangle|^2 + |\langle f_\theta^i, \psi_{j,x_0}^2 \rangle|^2 \forall i \in \{1, \dots, N\} \quad (2)$$

Le processus décrit ci-dessus est répété pour τ orientations θ_i ($i \in \{1, \dots, \tau\}$) et ν échelles j_k ($k \in \{1, \dots, \nu\}$). On obtient alors $\tau \cdot \nu$ imagerie d'énergie fovéale centrées sur chaque point saillant.

Chaque région support (i.e. région centrée sur un point saillant)

comprenant P pixels est alors représentée par la moyenne μ et l'écart type σ de ses énergies fovéales :

$$\mu_{\theta,j} = \frac{\sum_x E_{\theta,j}^i(x)}{P} \quad (3)$$

$$\sigma_{\theta,j} = \sqrt{\frac{\sum_x (E_{\theta,j}^i(x) - \mu_{\theta,j})^2}{P}}. \quad (4)$$

On obtient alors une signature S de taille $\tau \cdot \nu$ par région support définie par :

$$S = [\mu_{0,0}, \sigma_{0,0}, \dots, \mu_{\tau,\nu}, \sigma_{\tau,\nu}]. \quad (5)$$

Il est intéressant de constater que la construction de ce descripteur est proche de celle proposée dans [14] pour la caractérisation des textures. Toutefois, il utilise de nouvelles bases d'ondelettes qui s'avèrent être plus précises que les ondelettes de Gabor comme nous le verrons dans la prochaine section.

3 Expérimentations

Dans cette section, nous proposons d'étudier le pouvoir descriptif et discriminant de notre descripteur fovéal dans le cadre de deux applications : la classification supervisée d'images naturelles et la reconnaissance de visages.

3.1 Classification supervisée d'images naturelles

Dans cette application, nous avons choisi d'utiliser le détecteur de points saillants mis en place dans nos précédents travaux [8] afin de détecter les zones sur lesquelles le descripteur fovéal doit être calculé. Comme mentionné dans [8], ce détecteur de points saillants est basé sur une décomposition en ondelettes et place donc les points sur les contours abrupts (i.e. les singularités marquées) des images.

En ce qui concerne le paramétrage du descripteur fovéal, nous avons choisi $\tau = 8$ orientations $\nu = 3$ échelles, nous donnant ainsi une signature locale de taille 48 par point saillant. La distance L_2 a été utilisée pour comparer deux signatures fovéales. Pour tester notre approche, nous avons choisi d'utiliser un échantillon de la base d'images SIMPLICITY¹ comprenant cinq classes représentant des catégories sémantiques proches (plages, bus, éléphants, bâtiments, fleurs). Quelques images issues de cette base sont présentées en figure 4.

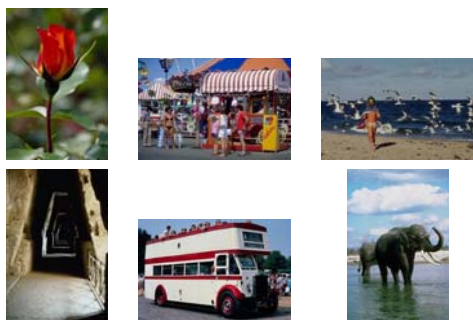


FIG. 4 – Images issues de la base SIMPLICITY

Les résultats présentés dans cette section ont été obtenues en utilisant 250 images d'apprentissage (i.e. 50 images par classe) et 250 images de tests. La classification a été effectuée au k plus proches voisins avec $k = 5$.

Les paramètres variables du descripteur concernent :

- la taille des régions support (i.e. du voisinage des points saillants utilisé pour calculer la signature fovéale) ;
- l'ordre des splines utilisées pour déterminer la base d'ondelettes (cf. section 2).

La figure 5(a) montre que la taille des régions supports influence peu le taux de classification et la figure 5(b) montre des résultats de classification équivalents quelque soit l'ordre de splines utilisées.

Comme nous l'avons mentionné précédemment, les ondelettes de Gabor sont souvent utilisées dans la littérature pour caractériser les singularités. De même, des ondelettes orientées ont récemment été proposées [7] pour caractériser efficacement les textures. La figure 5(c) compare notre descripteur fovéal à ces deux autres types de descripteurs et montre que quelque soit le nombre de points saillants utilisés pour représenter l'image, le descripteur fovéal donne de biens meilleurs taux de classification.

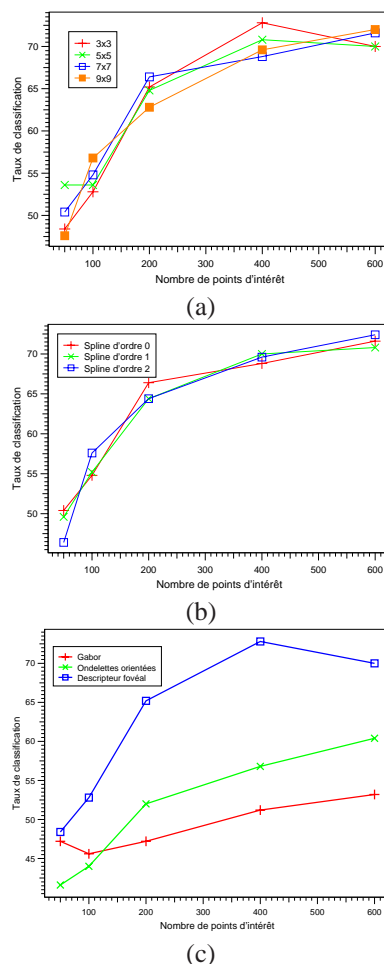


FIG. 5 – Taux de classification en fonction du nombre de points saillants détectés. Variation de la taille des régions support (a), variation de l'ordre des splines (b) et comparaison avec le descripteur de Gabor et les ondelettes orientées (c).

¹<http://wang.ist.psu.edu/jwang/test1.tar>

3.2 Reconnaissance de visages

Dans cette seconde application, nous avons choisi de tester le pouvoir discriminant de notre descripteur fovéal sur une application de reconnaissance de visages. Pour ce faire, nous avons créé un modèle de visage dans lequel nous avons positionné manuellement 20 points fiduciaux localisés sur des singularités faciales. Chaque visage sera donc représenté par 20 signatures fovéales de taille 48 (en utilisant 8 orientations et 3 échelles), chacune d'elles étant calculée dans un voisinage de taille 7×7 autour d'un des 20 points. Pour comparer deux signatures fovéales dans cette application, nous avons utilisé la distance cosinus.

Cette approche fait partie des techniques analytiques de reconnaissance de visages, par opposition aux approches holistiques considérant le visage dans sa globalité.

Pour tester notre approche, nous avons utilisé la base de visage FERET. Pour construire la galerie (i.e. la base de connaissance), nous avons choisi le corpus *fa* de la base FERET, comprenant 1196 visages (un exemple par individu) photographiés sous illumination normale et expression faciale neutre. Les requêtes ont, quant à elle, été choisies dans le corpus *fb* représentant des variations d'expressions faciales. Nous avons utilisé la totalité du corpus *fb*, soit 1196 requêtes.

La figure 6 illustre les résultats obtenus par le descripteur fovéal pour l'application de reconnaissance de visage. Comme on peut le voir, notre descripteur permet d'obtenir un taux de reconnaissance d'environ 70% en ne considérant que le premier visage retourné par le système.

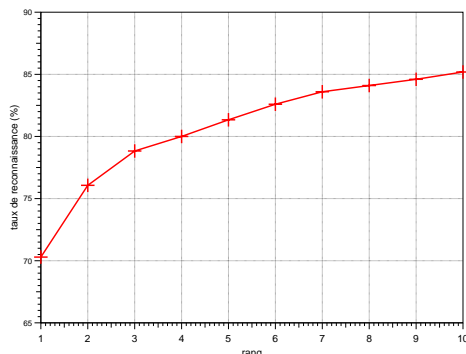


FIG. 6 – Taux de reconnaissance de visage en fonction du rang des visages retournés par le système.

4 Conclusion

Dans cet article, nous avons présenté un descripteur utilisant la théorie des ondelettes fovéales pour caractériser l'orientation et la régularité des singularités d'une image. Nous avons montré son aptitude à discriminer le contenu des images dans le cadre de deux applications différentes et nous travaillons sur d'autres expérimentations, notamment pour améliorer la reconnaissance des visages. Par exemple, l'utilisation d'autres bases d'ondelettes fovéales (comme les ondelettes fovéales régulières à support minimum présentées dans [11]) semble être une piste à suivre.

De plus, nous nous intéressons actuellement à l'implémentation d'un détecteur décrivant uniquement la répartition des orien-

tations et des coefficients de Lipschitz autour de chaque point d'intérêt. Enfin, le calcul de l'échelle caractéristique associée à chaque point d'intérêt semble inévitable pour améliorer les résultats.

Références

- [1] Biederman I. Recognition-by-Component : A Theory of Human Image Understanding. *Psychological Review*, 94(2) :115–147, 1987.
- [2] Bres S. and Jolion J.-M. Detection of interest points for image indexation. In *VISUAL '99 : Proceedings of the Third International Conference on Visual Information and Information Systems*, pages 427–434. Springer-Verlag, 1999.
- [3] Bresenham J. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1) :25–30, 1965.
- [4] Daugman J.G. Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7) :1169–1179, Juillet 1988.
- [5] Gouet V. and Boujeema N. Object-based queries using color points of interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 30–36, Hawaii, Décembre 2001.
- [6] Harris C. and Stephens M. A combined corner and edge detector. In *Proceedings of The 4th Alvey Vision Conference*, pages 147–151, 1988.
- [7] Kim N.D. and Udpa S. Texture Classification Using Rotated Wavelet Filters. *IEEE Transactions on System, Man, and Cybernetics - Part A : Systems and Humans*, 30(6) :847–852, Novembre 2000.
- [8] Laurent C., Laurent N., and Visani M. Color Image Retrieval Based on Wavelet Salient Features Detection. In *Proceedings of The 3rd Int. Workshop on Content-Based Multimedia Indexing*, pages 327–334, Rennes, France, Septembre 2003.
- [9] Loupiaz E., Sebe N., Bres S., and Jolion J.-M. Wavelet-based salient points for image retrieval. In *Proceedings of the International Conference on Image Processing*, volume 2, pages 518–521, Octobre 2000.
- [10] Lowe D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.
- [11] Mallat S. Foveal Approximations for Singularities. *Applied and Computational Harmonic Analysis*, 14(2) :133–180, 2003.
- [12] Mallat S. and Hwang W.L. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 32(2) :617–643, 1992.
- [13] Mallat S. and Zhong S. Characterization of Signals from Multiscale Edges. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(7) :710–732, 1992.
- [14] Manjunath B.S. and Ma W.Y. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8) :837–842, Août 1996.
- [15] Schmid C. and Mohr R. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5) :530–535, 1997.