

Extraction de la forme et de la perspective dans des textures artificielles et naturelles par modèles corticaux

Corentin MASSOT, jeanny HÉRAULT

Laboratoire des Images et des Signaux
46, Avenue Felix Viallet, 38031 Grenoble Cedex, France
corentin.massot@lis.inpg.fr, jeanny.herault@lis.inpg.fr

Résumé – Comment le système visuel extrait-il une information de perspective à partir de surfaces texturées? Afin de répondre à cette question, nous proposons un algorithme biologiquement plausible basé sur un modèle simplifié du système visuel. Tout d’abord de nouveaux filtres log-normaux sont présentés en remplacement des filtres de Gabor classiquement utilisés. Notamment ces filtres sont à variables séparables en fréquence et en orientation ce qui permet d’obtenir une méthode robuste d’estimation de la fréquence locale moyenne dans l’image. Basée sur cette nouvelle approche, une décomposition de l’image en imquettes, après un pré-traitement de type rétinien, permet d’estimer la variation de fréquence sur l’ensemble de l’image. La relation entre la variation de fréquence locale et les paramètres géométriques de la surface vue en projection perspective est calculée, ce qui permet d’obtenir l’orientation et la forme 3D de l’image originale. La robustesse de la méthode est testée et discutée sur différents types de textures, à la fois régulières et irrégulières, de même que sur des scènes naturelles.

Abstract – How does the visual cortex extract perspective information from textured surfaces? To answer this question, we propose a biologically plausible algorithm based on a simplified model of the visual processing. First, new log-normal filters are presented in replacement of the classical Gabor filters. Particularly, these filters are separable in frequency and orientation and this characteristic is used to derive a robust method to estimate the local mean frequency in the image. Based on this new approach, a local decomposition of the image into patches, after a retinal pre-treatment, leads to the estimation of the local frequency variation all over the surface. The analytical relation between the local frequency and the geometrical parameters of the surface, under perspective projection, is derived and finally allows to solve the so-called problem of recovering the shape from the texture. The accuracy of the method is evaluated and discussed on different kind of textures, both regular and irregular, and also on natural scenes.

1 Extraction de la perspective dans les scènes naturelles

Comment le système visuel obtient-il une information de perspective dans les images de scènes naturelles? Basé sur les connaissances actuelles du fonctionnement du système visuel primaire, nous proposons un algorithme permettant d’obtenir une information sur l’orientation et la forme d’une surface sur laquelle est plaquée une texture, c’est-à-dire un motif plus ou moins régulier possédant différentes composantes de fréquence et d’orientation (figure 1).

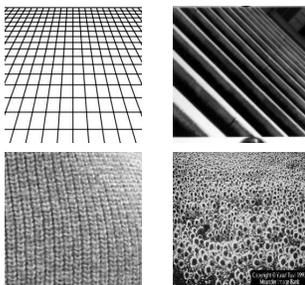


FIG. 1: Exemples de textures étudiées: quadrillage régulier, texture tirée de Super *et al* ([1]), mailles d’un pull tirée de Clerc *et al* ([2]), champ de tournesols.

Depuis le début des années 90, l’utilisation de l’information spectrale a conduit à plusieurs algorithmes efficaces d’extrac-

tion de la forme par la texture. Parmi ceux-ci on pourra notamment citer [3], [4] qui permettent de traiter des textures régulières ou possédant au moins deux composantes d’orientation et [1], [2], [5] qui ne font aucune hypothèse sur la forme du spectre et qui obtiennent également des résultats sur des textures irrégulières. Ce type de textures est plus difficile à traiter compte tenu des fortes non-stationnarités présentes dans l’image (par exemple dans un champ de tournesols, le changement de taille des fleurs ou la présence d’une personne créent des non-stationnarités locales).

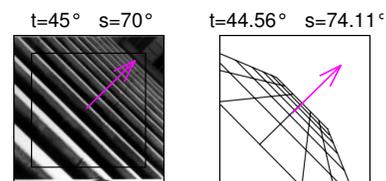


FIG. 2: Exemple de résultat obtenu sur la textures de Brodatz de la figure 1; à gauche sont indiqués les angles réels, sur l’image est reportée la normale estimée au plan; à droite sont indiqués les angles estimés et l’image représente un quadrillage incliné équivalent.

Notre travail s’inscrit dans le cadre d’une étude du fonctionnement du système visuel biologique. Les principaux mécanismes du système visuel primaire sont maintenant bien identifiés [6]. La rétine effectue un ensemble de pré-traitements

permettant notamment de séparer l'information d'illumination de celle de texture [7]. Chaque région de l'espace est analysée à travers un ensemble de champs récepteurs en recouvrement les uns par rapport aux autres. Le signal est ensuite transmis à l'aire visuelle primaire V1, où à chaque champ récepteur est associé un ensemble de cellules simples et complexes qui effectuent un échantillonnage du signal en fréquences et en orientations. Ces cellules sont organisées en macro-colonnes de même orientation et de même fréquence et la réponse des cellules complexes peut se modéliser par un filtrage passe-bande de type filtre de Gabor du spectre d'énergie correspondant à la région de l'espace visual analysée.

La projection d'une image du monde réel sur une surface induit des déformations affines (des gradients) de la texture qui recouvre la surface. Or en vision monoculaire statique, le système visuel est capable de retrouver les caractéristiques tridimensionnelles d'une image. Dans ce papier, une nouvelle technique permettant de retrouver la forme par l'analyse des gradients de texture, basée sur le modèle simplifié, précédemment décrit du système visuel, est présentée. Elle repose sur la décomposition locale d'une texture homogène en imagerie locales (de manière similaire à l'organisation des macro-colonnes de V1). La section 2 présente un nouveau type de filtres log-normaux, mieux adaptés que les filtres de Gabor pour simuler les filtres corticaux. Ils sont utilisés pour réaliser l'échantillonnage du spectre. Ces filtres permettent de développer une technique pour estimer la fréquence locale qui tire avantage de l'ensemble des échelles disponibles après filtrage. La section 3 présente le modèle de projection perspective utilisé et la relation entre la variation de fréquence locale et les paramètres géométriques de la surface. La section 4 décrit l'algorithme final ainsi obtenu et montre différents résultats sur des textures à la fois régulières et irrégulières, de même que sur des scènes naturelles.

2 Filtres log-normaux et estimation de la fréquence locale

Les filtres de Gabor sont habituellement utilisés pour modéliser les cellules corticales. Ils présentent l'avantage d'être localisés à la fois dans l'espace et en fréquence mais ils ne sont pas à variables séparables en fréquence et en orientation. De plus ils ne sont pas symétriques exprimés sur une échelles log-polaire et ont une transmission non nulle en $f = 0$ ce qui donne des réponses bruités en basse fréquence ([8]). Nous présentons ici de nouveaux filtres *log-normaux* qui sont mieux adaptés à notre problème et qui représentent une meilleure approximation de la réponse des cellules corticales.

2.1 Filtres log-normaux

L'expression de ces nouveaux filtres repose sur la loi log-normale:

$$\begin{aligned} |G_{i,j}(f, \theta)|^2 &= A \frac{1}{f^2} \exp\left(-\frac{1}{2} \left(\frac{\ln(f/f_i)}{\sigma_r}\right)^2\right) \cdot \cos^{2n}\left(\frac{\theta - \theta_j}{2}\right) \\ &= |G_r \cdot G_\theta|^2 \end{aligned}$$

où $G_{i,j}$ est la fonction de transfert du filtre, f_i , la fréquence centrale, θ_j , l'orientation centrale, σ_r , la bande passante sui-

vant les fréquences, n contrôle la bande passante suivant les orientations. A est un facteur de normalisation tel que $\|G_{i,j}(f, \theta)\|^2 = 1$ soit en prenant $A = \frac{2^{2n}}{2\pi C_{2n} \sigma_r \sqrt{2\pi}}$. Enfin G_r et G_θ représentent respectivement la composante radiale et la composante angulaire du filtre; ces filtres sont donc à variable séparables en fréquence et en orientation.

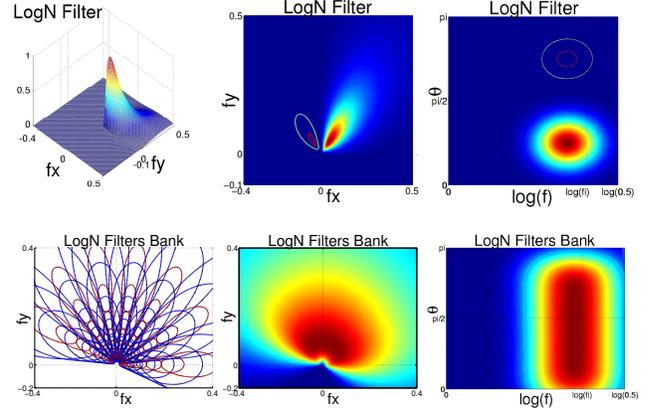


FIG. 3: 1ère ligne: filtres log-normaux, à gauche: représentation 3D; au milieu et à droite: filtre et contours représentant 50% et 90% de l'énergie maximale en coordonnées cartésiennes dans le plan des fréquences et en coordonnées log-polaires. 2ème ligne: banc de filtres log-normaux, à gauche: contours des filtres composant le banc; au milieu et à droite: banc de filtres en coordonnées cartésiennes dans le plan des fréquence et en coordonnées log-polaires.

La première ligne de la figure 3 présente les filtres log-normaux en coordonnées cartésiennes et log-polaires. Contrairement aux filtres de Gabor, les filtres log-normaux sont insensibles à la composante continue quelle que soit la largeur de bande choisie. Ces filtres sont asymétriques en coordonnées cartésiennes mais deviennent symétriques en coordonnées log-polaires, de manière similaire à la forme des réponses des cellules complexes du cortex visuel primaire ([6]). La deuxième ligne de la figure 3 présente un banc de filtres log-normaux. Il est possible d'observer la bonne couverture du spectre et la régularité de l'échantillonnage, notamment en log-polaire.

2.2 Estimation de la fréquence locale moyenne

Dans cette partie nous présentons une méthode permettant d'estimer la fréquence moyenne à une position donnée de l'image. Prenons la composante fréquentielle d'un filtre log-normal à la fréquence centrale f_i :

$$G_i^2(f) = \frac{1}{f^2} \exp\left(-\frac{1}{2} \left(\frac{\ln(f/f_i)}{\sigma_r}\right)^2\right)$$

De manière similaire à Knutsson *et al* [9], le rapport de filtres adjacents en f_i et f_{i+1} s'exprime par:

$$\begin{aligned} \frac{G_{i+1}^2(f)}{G_i^2(f)} &= \exp\left(-\frac{1}{2\sigma_r^2} [(\ln(f/f_{i+1}))^2 - (\ln(f/f_i))^2]\right) \\ &= \left(f/\sqrt{f_i f_{i+1}}\right) \frac{\ln(f_{i+1}/f_i)}{\sigma_r^2} \end{aligned}$$

En posant $\sigma_r^2 = \ln(f_{i+1}/f_i)$, nous obtenons finalement:

$$G_{i+1}^2(f) = \frac{f}{\sqrt{f_i f_{i+1}}} G_i^2(f)$$

Afin d'extraire une information d'échelle indépendante des orientations locales, nous considérons des filtres par bande de fréquence (notés FBF) qui résultent de la somme sur l'ensemble des orientations j des réponses des filtres centrés à la même fréquence f_i sur le spectre d'énergie de l'image $S(f, \theta)$:

$$C_i = \int_f G_i^2(f) \int_\theta S(f, \theta) \sum_j G_j^2(\theta) f df d\theta \quad (2) \quad \text{et}$$

$$\frac{C_{i+1}}{C_i} = \frac{1}{\sqrt{f_i f_{i+1}}} < f > \quad (3)$$

Le rapport de deux FBF adjacents donne l'équation 3 et représente une estimation de la fréquence locale moyenne pour la bande située à la i ème fréquence centrale du banc de filtres utilisé. Finalement en sommant sur l'ensemble des estimations, nous obtenons:

$$< f > = \sum_i \frac{C_i}{\sum_i C_i} < f >_i \quad (4)$$

Il est à noter que cette méthode tire partie des différentes estimations effectuées sur l'ensemble des bandes de fréquence disponibles, ce qui la rend robuste aux variations d'échelle entre les différentes textures étudiées. De plus cette méthode repose sur la séparation entre l'information d'échelle et l'information angulaire, ce qui permet d'utiliser uniquement l'information de fréquence sans hypothèse ou perturbation aucune provenant de l'information d'orientation contenue dans la texture analysée. Cette méthode s'avère simple et efficace car elle ne repose que sur la combinaison au premier ordre de différents filtres et s'adapte bien à un modèle *feedforward* du traitement visuel de bas niveau ce qui n'est pas le cas des méthodes similaires, telles que [1] ou [5].

3 Géométrie et estimation de l'orientation de surfaces planes

3.1 Projection perspective

Nous considérons un modèle de projection perspective pour le passage du monde tridimensionnel au monde bidimensionnel de l'image. La figure 4 présente le système de coordonnées de la projection d'une surface plane. (x_w, y_w, z_w) représente les coordonnées du monde, (x_s, y_s) , les coordonnées de la surface et (x_i, y_i) , les coordonnées de l'image. d (resp. z_w0) représente la coordonnée de l'image (resp. de la surface) sur l'axe z_w . τ (*tilt*) représente l'angle entre x_i et la projection de la normale z_s à la surface sur le plan de l'image. σ (*slant*) est l'angle entre z_w et la normale à la surface en z_w0 . Sa valeur est comprise entre 0 et $\pi/2$.

La relation entre les coordonnées (x_s, y_s) de la surface et les coordonnées (x_i, y_i) de l'image s'exprime par (voir également [1]):

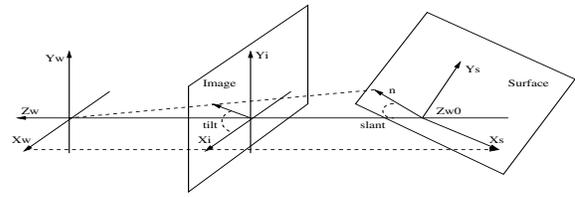


FIG. 4: Modèle de la projection perspective

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{\begin{bmatrix} \cos(\sigma) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\tau) & \sin(\tau) \\ -\sin(\tau) & \cos(\tau) \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix}}{a_i} = \frac{A}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (5)$$

avec $a_i = \frac{-\sin(\sigma)\sin(\tau)x_i + \cos(\tau)\sin(\sigma)y_i + d\cos(\sigma)}{d + z_w0}$ correspondant à un facteur de zoom en fonction de la position spatiale (x_i, y_i) , noté x_i (de même on notera f_i le vecteur (f_{ix}, f_{iy})). Autour de l'axe optique, le spectre d'énergie de l'image $I_i(f_i)$ est relié à celui de la surface $I_s(f_s)$ par:

$$\begin{aligned} I_i(f_i) &= \int_{x_i} i_s(x_s) e^{-j2\pi x_i^t f_i} dx_i \\ &= \int_{f_s} I_s(f_s) \int_{x_i} e^{j2\pi x_i^t (\frac{A^t}{a_i} f_s - f_i)} dx_i df_s \end{aligned}$$

En supposant la luminosité constante avec la projection, on obtient $i_i(x_i) = i_s(x_s)$ avec $x_s = \frac{A}{a_i} x_i$. En prenant a_i constant dans une région limitée centrée autour de x_i , le changement de variable $f_s \rightarrow a_i A^{-t} f_i$ donne:

$$I_i(f_i, x_i) = \int_{f_s} I_s(f_s) \delta\left(\frac{A^t}{a_i} f_s - f_i\right) df_s \quad \text{d'où} \quad \boxed{f_i = \frac{A^t}{a_i} f_s} \quad (6)$$

L'équation 6 donne la relation entre la fréquence de l'image f_i et la fréquence de la surface f_s .

3.2 Estimation de l'orientation de surfaces planes

Afin de relier la variation de fréquence avec l'orientation de la surface couverte par une texture homogène, la méthode ne repose que sur une hypothèse de stationnarité locale. La variation de fréquence sur l'image est alors considérée comme ne provenant que de l'inclinaison de la surface en profondeur. A partir de l'équation 6, la variation locale de fréquence de l'image f_i s'exprime par:

$$df_i = -\frac{\nabla^t a_i dx}{a_i^2} A^t f_s = -\frac{\nabla^t a_i dx}{a_i} f_i \quad (7)$$

En coordonnées polaires, la fréquence de l'image peut s'exprimer par $f_i = v_i [\cos(\varphi_i) \sin(\varphi_i)]^t$. L'équation 7 donne:

$$\begin{aligned} df_i &= dv_i [\cos(\varphi_i) \sin(\varphi_i)]^t + v_i [-\sin(\varphi_i) \cos(\varphi_i)]^t d\varphi_i \\ &= -\frac{\nabla^t a_i dx}{a_i} v_i [\cos(\varphi_i) \sin(\varphi_i)]^t \quad (8) \end{aligned}$$

Si on considère le gradient df_i suivant la direction de φ_i , en multipliant par $[\cos(\varphi_i) \sin(\varphi_i)]^t$ l'équation 8 devient:

$$\frac{dv_i}{v_i} = -\frac{\nabla^t a_i dx}{a_i} \quad \text{d'où}$$

$$d \ln(v_i) = \frac{-\sin(\sigma)}{a_i} [-\sin(\tau) \cos(\tau)] [dx_i dy_i]^t$$

Nous obtenons finalement que pour une surface homogène et localement stationnaire, l'angle de tilt correspond à la direction du gradient de fréquence et le slant est proportionnel à la norme du gradient du logarithme des fréquences avec la relation:

$$\tan(\sigma) = \left(\frac{d}{d + zw_0} \frac{1}{|d \ln(v_i)|} - ([-\sin(\tau) \cos(\tau)] [x_i y_i]^t / d) \right)^{-1} \quad (9)$$

4 Algorithme final et résultats

L'algorithme final peut alors se résumer par:

1. Préfiltrage rétinien de l'image et décomposition en images (par exemple: de taille 96X96 pixels avec un recouvrement de 8 pixels).
2. Extraction des coefficients des filtres log-normaux (par exemple: 7X7 filtres), normalisation par bande d'orientation et sommation sur l'ensemble des orientations afin d'obtenir les FBF.
3. Combinaison des réponses avec l'équation 2 et combinaison finale avec l'équation 4 afin d'obtenir une estimation de la fréquence moyenne locale sur l'ensemble de l'image.
4. Calcul des gradients locaux de fréquence, extraction du tilt et du slant et moyennage final.

La figure 5 présente les résultats obtenus sur différentes textures et sur des scènes naturelles. Nous obtenons une précision de l'ordre de 3° pour le tilt et de 5° pour le slant ce qui est comparable à la précision obtenue par d'autres auteurs tels que [1] et [5]. La méthode donne également des résultats fiables sur des textures irrégulières.

5 Conclusions

Dans ce papier nous avons présenté un nouveau type de filtres log-normaux, à variables séparables, permettant d'obtenir une méthode simple, efficace et robuste d'estimation de la fréquence moyenne locale dans une image. Un algorithme complet d'analyse de la forme par la texture a ainsi pu être présenté et évalué sur différents types de textures à la fois régulières et irrégulières, de même que sur des scènes naturelles. La précision obtenue est comparable à des méthodes proches mais plus complexes et s'accorde bien avec un traitement bas niveau du système visuel. La précision de cette méthode pourra être notamment améliorée par des méthodes de régularisation robustes.

Références

[1] B.J. Super et A.C. Bovik, *Planar surface orientation from texture spatial frequencies*, Pattern Recognition 28, n°5, pp. 729-743, 1995.

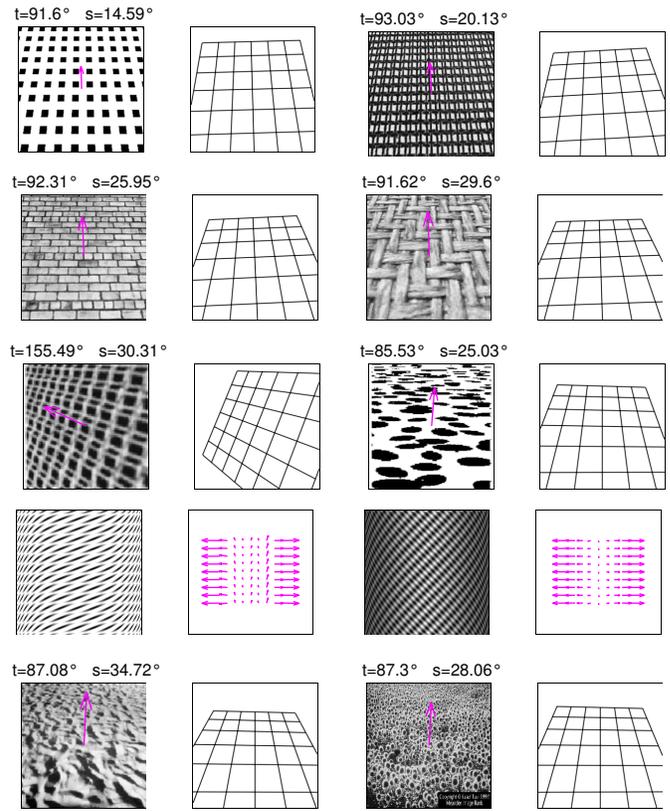


FIG. 5: Résultats obtenus sur des textures régulières, irrégulières de même que sur des scènes naturelles.

[2] M. Clerc et S. Mallat, *The texture gradient equation for recovering shape from texture*, IEEE Trans. PAMI 24, n°4, 2002.

[3] J. Malik et R. Rosenholtz, *Computing local surface orientation and shape from texture for curved surfaces*, International Journal of Computer Vision 23, n°2, pp. 149-168, 1997.

[4] E. Ribeiro et E.R. Hancock, *Shape from periodic texture using the eigen vectors of local affine distortion*, IEEE Trans. PAMI 23, n°12, 2001.

[5] W.L. Hwang, C.S. Lu et P.C. Chung, *Shape from texture estimation of planar surface orientation through the ridge surfaces of continuous wavelets transform*, Trans. on Image Processing 7, n°5, 1998.

[6] L. Spillmann et J.S. Werner, *Visual Perception: The Neurophysiological Foundations*, Academic Press, Inc., 1990.

[7] W.H. Beaudot, *The neural information in the vertebra retina: a melting pot of ideas for artificial vision*, PhD thesis, tirf laboratory, Grenoble, France, 1994.

[8] G. Wallis, *Linear models of simple cells: Correspondence to real cell responses and space spanning properties*, Spatial Vision 14, n°3,4, pp. 237-260, 2001.

[9] H. Knutsson, C.F. Westin et G. Granlund, *Local multiscale frequency and bandwidth estimation*, IEEE International Conference on Image Processing (ICIP'94), Austin, Texas, 1994.