

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

## Shape and Perspective Extraction in Artificial Textures and Natural Scenes by Cortical Models

Corentin Massot, Jeanny Hérault

Laboratoire des Images et des Signaux,  
corentin.massot@mcgill.ca, jeanny.herault@lis.inpg.fr

Manuscrit reçu le 5 octobre 2005

### Résumé et mots clés

Dans ce travail nous présentons un nouvel algorithme d'extraction de la forme par la texture appliqué à l'analyse des scènes naturelles. L'originalité de cette approche est basée sur la structure du cortex visuel primaire (V1) dont elle modélise les fonctions. L'algorithme est capable d'analyser une grande variété de textures présentant différents types d'irrégularités. Tout d'abord pour réaliser l'échantillonnage du spectre d'amplitude, nous proposons de nouveaux filtres, appelés *filtres log-normaux*, inspirés du fonctionnement des cellules complexes de l'aire V1, en remplacement des filtres de Gabor classiques. Ces filtres s'avèrent particulièrement appropriés aux techniques de reconnaissance de forme de part leurs différentes propriétés théoriques, notamment leur profil en fréquence radiale (adapté à la décroissance en  $1/f$  des scènes naturelles) et leur séparabilité en orientation et en fréquence. Nous utilisons ensuite une méthode d'estimation de la fréquence moyenne locale appliquées sur des signaux naturels. Celle-ci ne nécessite pas la recherche d'une échelle adaptée à l'analyse et tire avantage de l'ensemble des fréquences du banc de filtres utilisé. Finalement, à partir de l'estimation locale, l'orientation et la forme sont extraits en utilisant les propriétés géométriques de la projection perspective. La précision de la méthode est évaluée sur différents types de textures, à la fois régulières et irrégulières, et sur des scènes naturelles. La méthode présentée permet d'obtenir des résultats se comparant favorablement aux meilleures techniques existantes tout en conservant un faible coût de calcul. Enfin le modèle peut être adapté à d'autres applications telles que l'analyse de textures, l'extraction de points caractéristiques ou l'indexation d'images par le contenu.

Perception 3D, texture, scènes naturelles, filtres log-normaux

### Abstract and key words

In this work we present a new shape from texture algorithm applied to natural scenes analysis. The originality of this approach is based on the modeling of the structure of the primary visual cortex (V1). The algorithm is able to deal with a large variety of textures presenting different types of irregularities. First to sample the amplitude spectra, we present new filters, called log-normal filters, inspired from the complex cells of V1, in replacement of the classical Gabor filters. These filters appear to be suitable for pattern analysis techniques due to their different theoretical properties, notably their radial frequency profile (adapted to the  $1/f$  frequency profile of natural scenes) and their separability in orientation and frequency. We then use an estimation method of the local mean frequency applied to natural signals. This one does not imply the search for the adapted scale for the analysis and takes advantage of the frequencies of the used bank of filters.

Finally, from a local estimation, the orientation and shape are extracted using the geometrical properties of the perspective projection. The precision of the method is evaluated on different types of textures, both regular and irregular, and on natural scenes. The presented method allows to obtain favorably comparable results to existing best known methods with a low computational cost. Finally the model can be adapted to other applications like texture analysis, characteristic points extraction or content-based image indexation.

3D perception, texture, natural scenes, log-normal filters

## 1. La perspective dans les images naturelles

L'information 3D dans les scènes naturelles (comme la profondeur) peut être extraite sur l'image globale [1]. Cependant son extraction peut être également vu comme un procédé local si la scène est subdivisée en différentes régions. Chaque région peut être considérée comme étant constituée d'une texture homogène c'est-à-dire composée d'une distribution plus ou moins régulière d'éléments possédant des composantes caractéristiques en fréquence et en orientation (figure 1). Notre objectif est d'extraire l'orientation (et la forme) d'une surface locale supportant cette texture (figure 2).

S

Depuis le début des années 90, l'utilisation de l'information spectrale a conduit à de nombreux algorithmes d'extraction de la forme par la texture. Les méthodes développées par Malik *et al.* [4] et Ribeiro *et al.* [5] sont capables de traiter des textures contenant au moins deux composantes discrètes en orientation; Sakai *et al.* [6] et Guerin-Dugue *et al.* [7] ont développé des méthodes basées sur le suivi de pics d'énergie présents dans le spectre; d'autres méthodes [2], [8], [9], [10], [3] ne font aucune hypothèse sur les composantes spectrales. Ils obtiennent des résultats parmi les plus précis, sur des textures à la fois régulières et irrégulières. Ce dernier type de texture est difficile à analyser de par la présence de fortes non-stationnarités dans l'image (par exemple dans le champ de tournesols, le changement de taille des fleurs ou la présence d'une personne).

Exceptés les travaux de Sakai *et al.* [6], aucun de ces travaux n'établit une relation avec le fonctionnement du système visuel. Notamment l'utilisation de techniques spécifiques d'optimisation (pour l'estimation de déformations affines) sont difficilement comparables à des mécanismes biologiques.

Les principaux mécanismes du système visuel primaire sont maintenant globalement bien identifiés [11]. La rétine réalise un ensemble de prétraitements conduisant à la séparation de l'information d'illumination et de texture [12]. Chaque région de l'espace est analysée à travers un grand nombre de champs récepteurs, en superposition les uns par rapport aux autres. Le signal visuel est alors transmis à l'aire primaire V1 où un ensemble de cellules corticales (les cellules simples et les cellules complexes), associées à chaque champ récepteur, réalisent la décomposition du signal en orientation et en fréquence. La réponse des cellules complexes est classiquement modélisée par un filtre passe-bande de type filtre de Gabor appliqué au spectre d'amplitude associé à la région du champ visuel étudié. Le modèle développé par Heeger suit cette architecture en modélisant le comportement non-linéaire des cellules simples par la réponse de filtres orientés (de type filtres de gabor) semi-rectifiée [13]. Une étape d'inhibition divisive permet ensuite de moduler la réponse d'un filtre par les réponses des filtres dont le champ récepteur recouvre en partie celui de la cellule considérée. Cette étape permet de réaliser une « normalisation de contraste » en renforçant les différences entre les réponses des cellules. Ce modèle permet de rendre compte de plusieurs propriétés des cellules corticales.

En parallèle des études en psychophysique Zaidi *et al.* [14] émettent l'hypothèse de l'existence de mécanismes spécifiques dédiés à l'extraction des variations de fréquence et d'orientation pour la perception de la forme à partir des distortions de la texture. Cette approche peut être considérée comme une alternative aux modèles basés sur l'extraction de pics spectraux tel que celui développé par Sakai *et al.* [6].

Dans ce papier nous présentons un algorithme permettant de résoudre le problème de l'estimation de la forme par la texture, notamment sur des textures irrégulières, en se basant sur le fonctionnement du système visuel. Les principales caractéris-

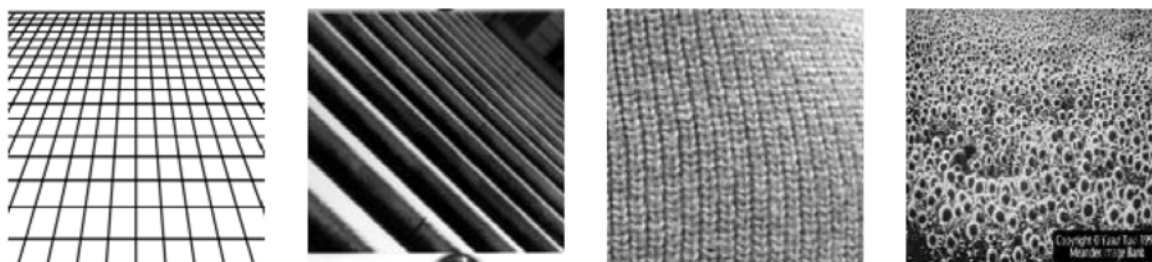


Figure 1. Exemples de textures étudiées : quadrillage régulier, textures de Super et al. ([2]), mailles de pull de Clerc et al. ([3]), un champ de tournesols.

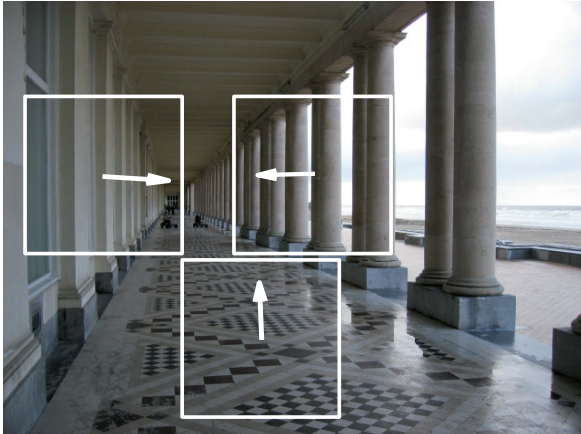


Figure 2. Exemple d'extraction de l'information de perspective dans une scène naturelle (corridor du château royal à Ostende); une estimation de l'orientation locale (les angles de tilt et slant) est obtenue sur des régions de l'image constituées d'une texture homogène.

tiques de l'algorithme sont : une hypothèse d'homogénéité de la texture ; aucune hypothèse sur les composantes du spectre permettant de traiter un nombre varié de textures ; cet algorithme est en accord avec un traitement bas-niveau de l'information 3D dans le système visuel et chaque étape peut être associée à un mécanisme cortical.

La section 2 présente un nouveau type de filtres log-normaux, réalisant une meilleure modélisation des cellules corticales que les filtres de Gabor. La section 3 présente une technique pour estimer la fréquence moyenne locale sur des textures naturelles. La section 5 présente la relation entre une variation locale de fréquence et les paramètres géométriques de la surface. La section 6 présente une comparaison quantitative avec d'autres algorithmes connus, des résultats sur une base importante de textures à la fois régulières et irrégulières, et enfin des résultats sur des scènes naturelles.

## 2. Échantillonnage du spectre avec les filtres log-normaux

Différents types de filtres sont habituellement utilisés dans les techniques de reconnaissance de forme, notamment les filtres de Gabor. Ceux-ci présentent l'avantage d'être localisés dans l'espace et dans les fréquences et permettent une paramétrisation aisée en fréquence et en orientation. Cependant ils ne sont pas à variables séparables et pour des largeurs de bande suffisantes, ils n'assurent pas la nullité de la composante continue en  $f = 0$ . Dans ce papier de nouveaux filtres, appelés filtres *log-normaux*, sont introduits pour éviter ces limitations.

### 2.1. Filtres log-normaux

Dans cette section, nous présentons un nouveau type de filtres obtenus à partir de la fonction log-normale et définis par :

$$|G_{i,j}(f,\theta)|^2 = A \frac{1}{f^2} \exp\left(-\frac{1}{2} \left(\frac{\ln(f/f_i)}{\sigma_r}\right)^2\right) \cdot \cos^{2n}\left(\frac{\theta - \theta_j}{2}\right) = |G_i(f) \cdot G_j(\theta)|^2 \quad (1)$$

Avec  $G_{i,j}$ , la fonction de transfert du filtre,  $G_i(f)$  et  $G_j(\theta)$  représentant respectivement la réponse fréquentielle (radiale) et en orientation (tangentielle) du filtre. Leur enveloppe radiale est gaussienne sur une échelle log-fréquence.  $f_i$  est la fréquence centrale,  $\theta_j$ , l'orientation centrale,  $\sigma_r$ , la largeur de bande en fréquence.  $n$  contrôle la largeur de bande en orientation de manière à se rapprocher d'une enveloppe gaussienne.

$A (= \frac{2^{2n}}{2\pi C_{2n}^n} \frac{1}{\sigma_r \sqrt{2\pi}})$  est un facteur de normalisation tel que  $\|G_{i,j}(f,\theta)\|^2 = 1$ .

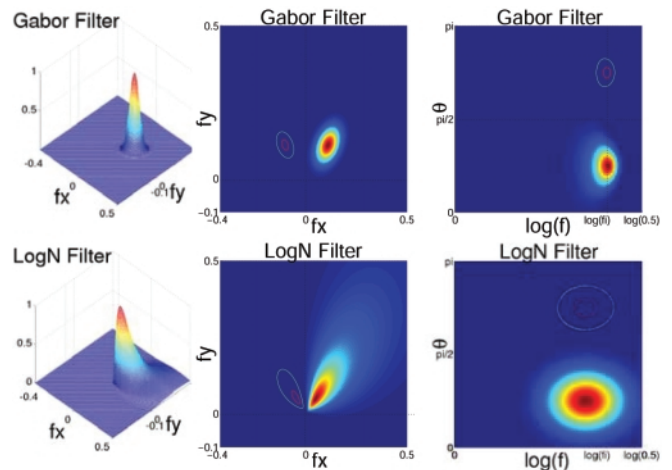


Figure 3. Comparaison entre les filtres de Gabor (première ligne) et les filtres log-normaux (deuxième ligne) ; à gauche : représentation 3D ; au milieu : filtre et ses contours à 50 % et 90 % du maximum d'énergie en coordonnées cartésiennes ; à droite : même filtre mais représenté en coordonnées log-polaires : les filtres log-normaux deviennent symétriques de manière similaire au profil des cellules corticales [11].

Les filtres log-normaux partagent les mêmes propriétés que les filtres de Gabor : localisation spatiale et paramétrables en fréquence et en orientation. Ils sont séparables en fréquence et en orientation (équation (1)). La composante radiale  $G_i(f)$  représente la fonction log-normale appliquée aux fréquences. La composante en orientation  $G_j(\theta)$  est une forme en cosinus assurant une largeur de bande  $2\pi$ -périodique en orientation et une valeur nulle aux limites angulaires du filtre (évitant d'effectuer une troncature). Finalement, le gain du filtre est nul en  $f = 0$  quelque soit la largeur de bande, le filtre log-normal est donc toujours bien défini même en très basses fréquences.



La figure 3 présente une comparaison entre les filtres de Gabor et les filtres log-normaux qui sont représentés en 3D, en coordonnées cartésiennes et log-polaires. La figure 4 présente les différents profils d'un banc de filtres log-normaux. Pour ce dernier, les profils sont représentés sans le coefficient en amplitude  $1/f$  afin de pouvoir observer la dissymétrie en basses fréquences.

Pour échantillonner la moitié supérieure du spectre d'amplitude, nous définissons un banc de filtres log-normaux avec des largeurs de bandes constantes en fréquence relative et en orientation. Sur la figure 5 à droite, il est possible de vérifier que cette définition conduit à une bonne couverture du spectre en coordonnées cartésiennes et à un échantillonnage en coordonnées log-polaires plus régulier qu'en utilisant un banc équivalent composé de filtres de Gabor.

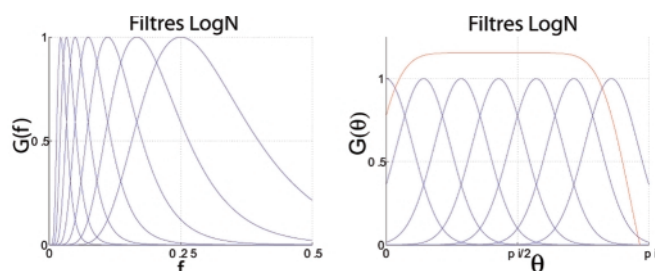


Figure 4. Présentation des profils des filtres ;

à gauche : profils des composantes fréquentielles suivant la loi log-normale sans le coefficient  $1/f$  ; à droite : profils de la composante angulaire (leur somme est constante).

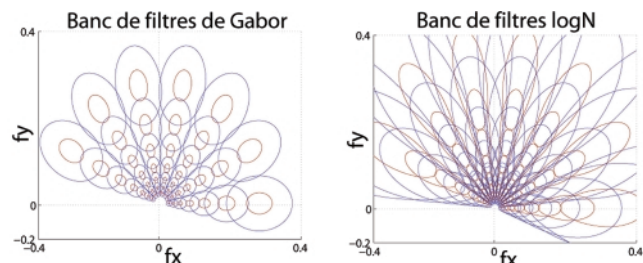


Figure 5. Comparaison entre un banc de filtres de Gabor (à gauche) et un banc de filtres log-normaux (à droite) en coordonnées cartésiennes ; seuls les contours des filtre à 50 % et à 90 % du maximum d'énergie sont représentés.

## 2.2. Analyse spectrale sur des images naturelles

Le spectre d'énergie des images naturelles a une décroissance suivant une pente en  $1/f^\alpha$  [15] avec  $\alpha \approx 2$ . Comme le montre la figure 6 à droite, la réponse des filtres log-normaux est importante en basse fréquence. Pour compenser cet effet, chaque image est prétraitée par un filtre basé sur un modèle du fonctionnement de la rétine [12] [16] [17]. Celle-ci réalise un blanchiment spectral (figure 6 gauche et droite). Après le préfiltrage rétinien (figure 6 à droite), les réponses des filtres log-normaux sont plus importantes en moyenne fréquence et réduites en basse fréquence. Il est possible d'observer que, dans les mêmes conditions, les filtres de

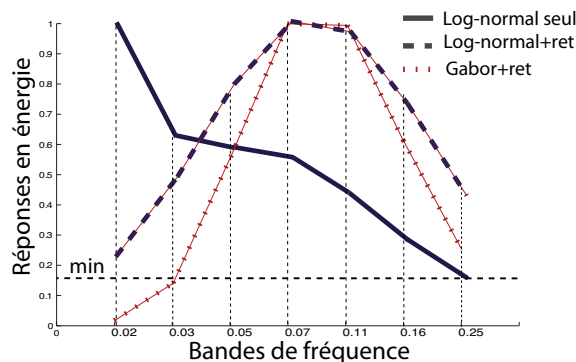
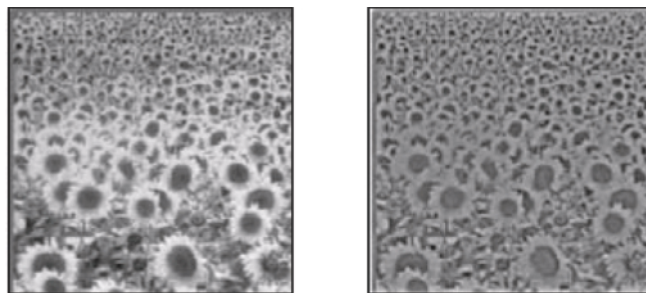


Figure 6. Effet du préfiltrage rétinien ; à gauche : image d'un champ de tournesols ; même image après préfiltrage rétinien ; à droite : réponses des filtres log-normaux et des filtres de Gabor sur le spectre d'amplitude global de l'image avant et après préfiltrage rétinien ; chaque point correspond à la réponse des filtres dans une bande de fréquence particulière (voir la section 3 pour le calcul des réponses par bandes de fréquence).

Gabor ne donnent pas de réponses bien définies en basse fréquence (sur la figure 6 droite) : les deux bandes situées en très basse fréquence (centrées sur  $f_0 = 0.02$  et  $f_0 = 0.03$ ) ont des réponses quasiment nulles contrairement à celles des filtres log-normaux équivalents qui rendent bien compte de la présence des composantes très basses fréquences de l'image.

# 3. Estimation de la fréquence moyenne locale

## 3.1. Extraction large bande

Dans cette section, la propriété de séparabilité des filtres est utilisée pour estimer la fréquence locale de l'image. Prenons la réponse du  $i$ -ème filtre log-normal :

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

De manière similaire à Knutsson *et al.* [18] [19], le rapport des réponses de deux filtres adjacents peut s'exprimer par :

$$\begin{aligned} \frac{G_{i+1}^2(f)}{G_i^2(f)} &= \exp\left(-\frac{1}{2\sigma_r^2}[(\ln(f/f_{i+1}))^2 - (\ln(f/f_i))^2]\right) \\ &= \left(\frac{f/\sqrt{f_i f_{i+1}}}{\sigma_r^2}\right)^{\frac{\ln(f_{i+1}/f_i)}{\sigma_r^2}} \end{aligned} \quad (3)$$

En posant  $\sigma_r^2 = \ln(f_{i+1}/f_i)$ , nous obtenons la relation suivante entre les réponses des filtres :

$$G_{i+1}^2(f) = \frac{f}{\sqrt{f_i f_{i+1}}} G_i^2(f) \quad (4)$$

Afin d'extraire une information d'échelle indépendante des orientations locales, nous considérons des réponses par bandes de fréquence obtenues par sommation sur toutes les orientations  $j$  des réponses des filtres centrés sur la même fréquence  $i$  sur le spectre de l'image  $S(f, \theta)$  :

$$C_i = \int_f G_i^2(f) \int_\theta S(f, \theta) \sum_j G_j^2(\theta) f df d\theta \quad (5)$$

À l'aide de l'équation (4), le rapport des réponses de deux filtres par bandes de fréquence adjacentes  $C_{i+1}$  et  $C_i$  donne :

$$\frac{C_{i+1}}{C_i} = \frac{1}{\sqrt{f_i f_{i+1}}} \langle f \rangle_i \quad (6)$$

où  $\langle f \rangle_i$  représente la fréquence moyenne locale estimée à la  $i$ -ème bande de fréquence. L'équation (6) montre qu'elle peut être facilement extraite par le rapport de deux filtres adjacents en connaissant leur fréquence centrale  $f_i$  et  $f_{i+1}$ .

Finalement en sommant sur toutes les estimations de la fréquence moyenne  $\langle f \rangle_i$  à différentes bandes de fréquence  $i$  et en pondérant chaque estimation par la réponse normalisée de la bande de fréquence correspondante, nous obtenons une estimation large bande de la fréquence moyenne  $\langle f \rangle$  :

$$\langle f \rangle = \sum_i \frac{C_i}{\sum_j C_j} \langle f \rangle_i \quad (7)$$

La réponse  $C_i$  dans l'équation (5) peut être interprétée comme la mesure du poids de la caractéristique fréquentielle locale extraite de l'image parmi l'ensemble des caractéristiques extraites. En effet si la réponse  $C_i$  est forte alors cela signifie que la fréquence moyenne  $\langle f \rangle_i$ , mesurée autour des 2 fréquences centrales  $f_i$  et  $f_{i+1}$ , est une composante importante de l'image observée.

Au contraire l'estimation finale  $\langle f \rangle$  de l'équation (7) est basée sur une estimation globale sommée sur l'ensemble de toutes les bandes de fréquence. Ainsi contrairement à [8] et à [3], dans cette approche il n'y a pas besoin d'estimer au préalable une fréquence diagnostique. Enfin l'estimation large bande permet d'obtenir une estimation moins influencée par un bruit qui pourrait être présent à une fréquence particulière.

Cette méthode présente également la caractéristique de réaliser la séparation entre l'information de fréquence et d'orientation, ainsi que le suggèrent les travaux de Zaidi *et al.* [14]. La fréquence locale est estimée sans aucune hypothèse sur les statis-

tiques des orientations, c'est-à-dire notamment sans hypothèse d'isotropie, comme il l'a été suggéré à de multiples reprises [20] [21]. Seule une hypothèse d'homogénéité locale sur les statistiques des composantes des fréquences est utilisée.

### 3.2. Largeurs de bandes relatives non-constantes

Afin d'évaluer la précision de la méthode d'extraction, celle-ci est appliquée sur un réseau dont la fréquence est parfaitement connue (figure 7 haut). Ce réseau contient une augmentation linéaire de la fréquence partant des très basses fréquences (fréquence nulle) jusqu'aux très hautes fréquences ( $f = 0.33$ ). L'algorithme d'extraction de la fréquence moyenne locale, décrit précédemment, est appliqué à chaque position spatiale suivant l'axe des abscisses à travers un ensemble de fenêtres locales appelées dans la suite *imagettes* (voir section 4).

Les courbes d'estimation finales (figure 7 bas) sont obtenues après approximation (méthode du simplex) de la fonction de l'ensemble des estimations obtenues aux différentes positions spatiales du réseau. La courbe en pointillés représente l'estima-

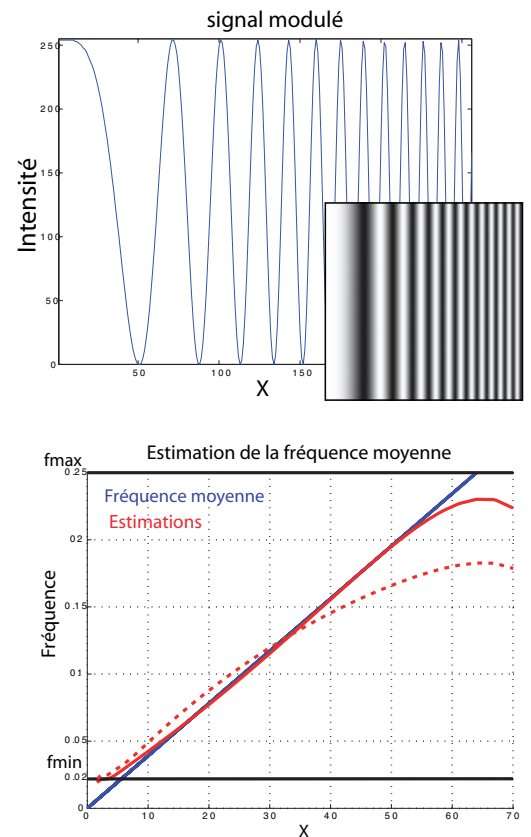


Figure 7. Estimation de la fréquence locale sur un réseau; en haut: projection du réseau avec le signal 1D représentant la modulation correspondante du signal; en bas: fréquence moyenne réelle du réseau (ligne pleine, rectiligne); estimation de la fréquence moyenne avec une largeur de bande relative constante des filtres log-normaux (ligne en pointillée); estimation de la fréquence moyenne avec une largeur de bande relative non-constante (ligne pleine).

tion obtenue avec le banc de filtres log-normaux décrit à la section 2. L'erreur d'estimation apparaît faible dans les basses fréquences mais augmente rapidement vers une valeur asymptotique dans les hautes fréquences.

Une des caractéristiques des filtres log-normaux est d'avoir leur largeur de bande relative en fréquence ne dépendant pas de leur fréquence centrale  $f_0$ . En effet la fréquence moyenne d'un filtre peut s'exprimer par  $u_f = e^{\sigma_r^2/2} f_0$  (moment du premier ordre) et sa largeur de bande, par  $\Delta_f = \sqrt{(e^{\sigma_r^2} - 1)} e^{\sigma_r^2/2} f_0$  (moment du deuxième ordre). Donc la largeur de bande en fréquence est donnée par  $\Delta_f/u_f = \sqrt{(e^{\sigma_r^2} - 1)}$  qui ne dépend que de la largeur de la gaussienne  $\sigma_r$ .

Ici, l'équation (4) impose une valeur de  $\sigma_r$  égale à  $\ln(f_{i+1}/f_i)$  afin de conserver leur rapport constant et égal à 1. Les filtres du premier banc décrits à la section 2 sont donc définis avec une largeur de bande relative constante égale à 1.4 octaves (avec  $f_{i+1}/f_i = 1.5$ ).

Si nous relâchons la condition sur  $\sigma_r$  de manière à ce que le rapport décrit précédemment devienne légèrement supérieur à 1, l'équation (6) reste une approximation valide de la fréquence moyenne locale. Par contre l'analyse de l'équation (3) montre que cela induit une compensation de la diminution de la précision de l'estimation dans les hautes fréquences. Pour obtenir cet effet, un coefficient imposant une décroissance linéaire de  $\sigma_r$  est ajouté et différentes valeurs sont testées. Avec un coefficient égal à 2, la méthode permet d'obtenir une estimation très précise sur l'ensemble des fréquences disponibles (ligne pleine sur la figure 7 en haut) à partir de la fréquence minimale  $f_{min} = 0.02$  à la fréquence maximale  $f_{max} = 0.25$  de notre banc de filtres. La figure 8 présente différents résultats de l'estimation de la méthode pour différentes tailles d'images (avec un coefficient identique égal à 2), depuis des tailles très petites ( $48 \times 48$ ) jusqu'à des tailles très importantes ( $128 \times 128$ ). Il est possible d'observer que la précision de la méthode n'est que faiblement influencée par la taille de l'image, qui pourra donc être choisie librement.

Cette largeur de bande relative non-constante en fréquence a été également observée sur les cellules corticales. Les données collectées sur le cortex chez l'Homme et chez le macaque [11]

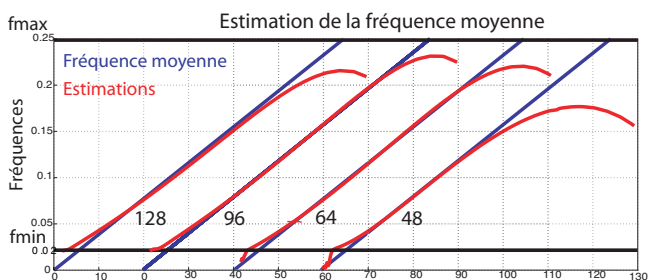


Figure 8. Estimation de la fréquence locale pour différentes tailles d'images (depuis  $48 \times 48$  pixels (images très petites) jusqu'à  $128 \times 128$  pixels (images très grandes)).

montrent clairement que la largeur de bande des cellules corticales de V1 n'est pas constante mais décroît linéairement avec les fréquences centrales. Ainsi le modèle de V1 pour l'analyse et l'extraction de la fréquence locale ici présenté suggère une explication empirique possible de la configuration particulière des cellules corticales.

## 4. Analyse des images naturelles

### 4.1. Décomposition en images

Afin d'estimer la fréquence locale sur toute la surface de l'image, celle-ci est décomposée en images (figure 9). La définition automatique d'une taille optimale des images afin d'étudier les propriétés locales d'une texture est un problème encore non résolu. Dans notre cas, des images de taille  $96 \times 96$  pixels sont jugées appropriées pour analyser des images de taille  $256 \times 256$  pixels. Une fenêtre de Hamming est appliquée sur chaque image afin d'éviter les effets de bord dans la transformée de Fourier, réduisant la zone d'analyse à une ouverture circulaire de 85 pixels de rayon. Ainsi pour une taille de pixels de 0.21mm, une image de 96 pixels de côté perçue à une distance de 1m correspond à un angle visuel de  $1^\circ$  ce qui est la taille moyenne des champs récepteurs de V1. La précision spatiale peut être adaptée en faisant varier le décalage entre les images. Généralement un décalage de 8 ou 4 pixels

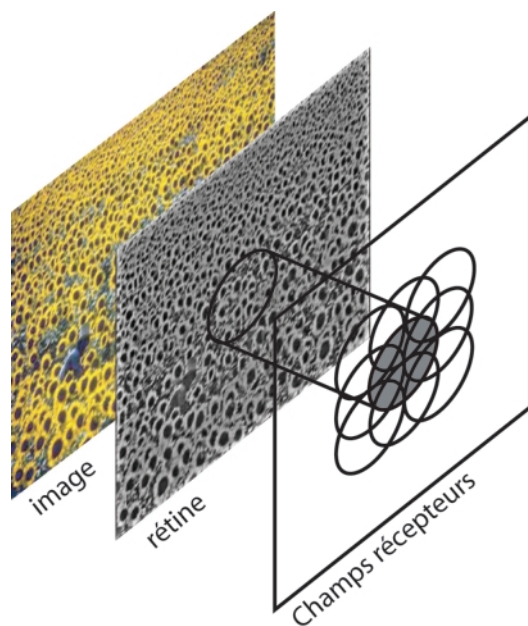


Figure 9. Les différentes étapes du prétraitement appliqué à chaque image : image initiale d'un champ de tournesols ; préfiltrage par la rétine ; décomposition locale en images de l'ensemble de l'image de manière similaire aux champs récepteurs des cellules corticales.

est choisi, ce qui correspond à une décomposition de l'image en  $21 \times 21$  ou  $42 \times 42$  imagettes. La figure 9 présente les différentes étapes de prétraitement appliqué sur chaque imagette avant l'application des filtres corticaux.

Cette décomposition locale en imagettes peut être vue comme un modèle de l'échantillonnage opéré par les cellules corticales. La région spatiale qu'elles recouvrent correspond au concept de champ récepteur spatial, introduit en physiologie.

#### 4.2. Normalisation corticale

Les réponses par bandes de fréquence, correspondant aux coefficients  $C_i$  dans les équations (6) et (7), sont obtenues après sommation sur l'ensemble des filtres à la même fréquence centrale et sur toutes les orientations. Afin de compenser une partie des irrégularités locales dans la texture, une normalisation locale est appliquée. Plus précisément, une variation d'énergie d'une position spatiale à l'autre peut apparaître dans la réponse des filtres. Afin de réduire ces variations, en se basant sur l'hypothèse d'homogénéité de la texture analysée, l'ensemble des réponses des filtres, à la même orientation, est normalisé par la somme de leur réponse sur l'ensemble des fréquences centrales. La réponse normalisée du filtre  $G_{i,j}^2$ , notée  $G_{i,j,norm}^2$ , peut alors se réécrire de la manière suivante :

$$\begin{aligned} G_{i,j,norm}^2(f,\theta) &= \frac{G_{i,j}^2(f,\theta)}{\sum_k G_{k,j}^2(f,\theta) + \epsilon} \\ &= \frac{G_i^2(f)}{\sum_k G_k^2(f) + \frac{\epsilon}{G_j^2(\theta)}} \end{aligned} \quad (8)$$

la constante  $\epsilon$  permet d'éviter des réhaussements de réponses des filtres dans le cas où l'énergie globale est trop faible dans la bande d'orientation considérée ( $\epsilon = 0.1$ ). En introduisant  $G_{i,j,norm}^2$  au lieu de  $G_{i,j}^2$  dans l'équation (5), l'équation (6) reste inchangée. La combinaison finale est obtenue en appliquant l'équation (7).

Ce processus est comparable à la normalisation divisive de Heeger [13] bien qu'ici les réponses soient renforcées uniquement selon l'orientation. Il est également important de remarquer que ce calcul de la réponse par bandes de fréquence représente une manière simple de séparer l'information de fréquence de celle d'orientation. Cela est à mettre en relation avec les travaux en psychophysique menés par Zaidi *et al.* [14] qui émettent l'hypothèse de l'existence dans le système visuel humain de deux mécanismes spécialisés pour l'analyse de la variation de fréquence d'une part et de la variation d'orientation d'autre part.

## 5. Extraction de l'orientation d'une surface plane

Dans cette section, nous établissons la relation entre la fréquence locale de la surface perçue et la fréquence locale de l'image, en projection perspective. Ensuite les paramètres d'orientation (angles de tilt et de slant, définis plus bas) de la surface sont obtenus à partir de la variation locale de fréquence. Finalement un algorithme complet d'extraction de la forme par la texture est présenté.

#### 5.1. Relations géométriques

La figure 10 présente le système de coordonnées d'une projection perspective associé à une surface plane.  $(x_w, y_w, z_w)$  représentent les coordonnées du monde,  $(x_s, y_s)$ , les coordonnées de la surface et  $(x_i, y_i)$ , les coordonnées de l'image. L'axe  $z_w$  correspond à l'intersection entre le centre de projection, l'origine des coordonnées de l'image et l'origine des coordonnées de la surface.  $d$  (resp.  $z_{w0}$ ) est la coordonnée de l'image (resp. de la surface) sur l'axe  $z_w$ . On note  $dz_{w0}$  la distance entre l'image et la surface.  $\tau$  représente l'angle de tilt qui est l'angle entre  $x_i$  et la projection de la normale  $z_s$  sur la plan de l'image.  $\sigma$  est l'angle de slant qui est l'angle entre l'axe  $z_w$  et la normale à la surface en  $z_{w0}$  et sa valeur est comprise entre 0 et  $\pi/2$ .

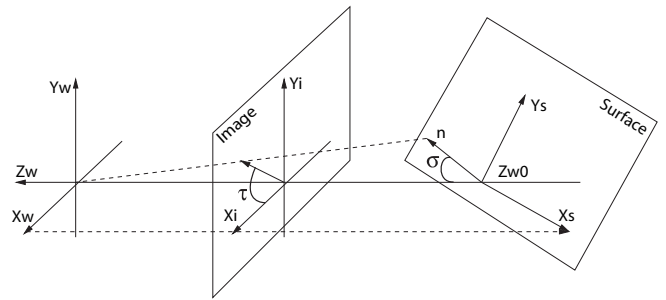


Figure 10. Modèle de projection perspective.

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 [2]) :

$$\begin{aligned} \begin{bmatrix} x_s \\ y_s \end{bmatrix} &= \frac{\begin{bmatrix} \cos(\sigma) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\tau) & \sin(\tau) \\ -\sin(\tau) & \cos(\tau) \end{bmatrix}}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \\ &= \frac{A}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \end{aligned} \quad (9)$$

$$\text{avec } a_i = \frac{-\sin(\sigma)\sin(\tau)x_i + \cos(\tau)\sin(\sigma)y_i + d\cos(\sigma)}{d + dz_{w0}}$$



correspondant à un facteur de zoom en fonction de la position spatiale  $(x_i, y_i)$ .

En supposant l'analyse réalisée dans une région  $L_i$  centrée sur une position  $x_i$ , la transformée de Fourier locale correspondante  $I_{L_i}$  s'exprime par :

$$I_{L_i}(f_i, x_i) = \int_u i_i(u) w_i(u - x_i) e^{-j2\pi(u-x_i)^t f_i} du \quad (10)$$

où  $i_i$  est le signal image et  $w_i$ , une fenêtre spatiale dans cette image. En prenant dans la surface  $i_s$  une fenêtre spatiale équivalente  $w_s$  et en posant  $v = T^{-1}u$  et  $x_s = T^{-1}x_i$ , les régions de l'image et de la surface sont reliées par :  $i_s(v)w_s(v - x_s) = i_i(u)w_i(u - x_i)$ . En appliquant la transformée de Fourier inverse de  $i_s(v)w_s(v - x_s)$  dans l'équation (10), on obtient :

$$\begin{aligned} I_{L_i}(f_i, x_i) &= \int_u \left( \int_{f_s} I_{L_s}(f_s) e^{j2\pi(v-x_s)^t f_s} \right) e^{-j2\pi(u-x_i)^t f_i} du df_s \\ &= \int_{f_s} I_{L_s}(f_s) \int_u e^{j2\pi((v-x_s)^t f_s - (u-x_i)^t f_i)} du df_s \quad (11) \end{aligned}$$

D'après l'équation (9), l'approximation au premier ordre de  $(v - x_s)^t$  donne :

$$(v - x_s)^t = \frac{1}{a_i} \left( I - \frac{\nabla a_i x_i^t}{a_i} \right) A^t (u - x_i)^t = R^t (u - x_i)^t.$$

En remplaçant dans l'équation (11), on obtient la relation entre  $I_{L_i}$  et  $I_{L_s}$  :

$$\begin{aligned} I_{L_i}(f_i, x_i) &= \int_{f_s} I_{L_s}(f_s) \int_u e^{j2\pi(u-x_i)^t (R^t(x_i) f_s - f_i)} du df_s \\ &= \int_{f_s} I_{L_s}(f_s) \delta(R^t(x_i) f_s - f_i) df_s \quad (12) \\ &= \frac{1}{|\det(R)|} I_{L_s}(R^{-t}(x_i) f_i) \end{aligned}$$

Finalement, on obtient la relation entre  $f_i$  et  $f_s$  (pour  $\delta$  non nul) :

$$f_i = R^t(x_i) f_s \approx \frac{1}{a_i} \left( I - \frac{\nabla a_i x_i^t}{a_i} \right) A^t f_s \quad (13)$$

Afin de relier la variation de fréquence avec la forme de la surface ou son orientation, une hypothèse d'homogénéité de la texture est nécessaire, comme décrit précédemment. Ainsi la variation de fréquence sur l'image permet de retrouver l'inclinaison de la surface avant projection.

En utilisant l'équation (13), l'expression de la variation locale de fréquence de l'image est :

$$df_i = -\frac{1}{a_i} [\nabla^t a_i dx_i + \nabla a_i dx_i^t] f_i \quad (14)$$

La fréquence de l'image  $f_i$  peut s'exprimer en coordonnées polaires par  $f_i = v_i [\cos(\varphi_i) \sin(\varphi_i)]^t$ . L'équation (14) devient :

$$\begin{aligned} df_i &= dv_i [\cos(\varphi_i) \sin(\varphi_i)]^t + v_i [-\sin(\varphi_i) \cos(\varphi_i)]^t d\varphi_i \\ &= -\frac{\nabla^t a_i dx}{a_i} v_i [\cos(\varphi_i) \sin(\varphi_i)]^t \\ &\quad - \frac{1}{a_i} \nabla a_i dx_i^t v_i [\cos(\varphi_i) \sin(\varphi_i)]^t \quad (15) \end{aligned}$$

Si l'on considère le gradient  $df_i$  dans la direction  $\varphi_i$ , en multipliant par  $[\cos(\varphi_i) \sin(\varphi_i)]^t$  l'équation (15) devient :

$$\begin{aligned} dv_i &= -\frac{\nabla^t a_i dx}{a_i} v_i \\ &\quad - \frac{1}{a_i} [\cos(\varphi_i) \sin(\varphi_i)] \nabla a_i dx_i^t [\cos(\varphi_i) \sin(\varphi_i)]^t v_i \quad (16) \end{aligned}$$

En intégrant sur  $\varphi_i$  (sur  $[0, 2\pi]$ ) on obtient :

$$\frac{dv_i}{v_i} = -\frac{3}{2} \frac{\nabla a_i^t dx_i}{a_i} \quad (17)$$

soit

$$d \ln(v_i) = -\frac{3}{2} \frac{\sin(\sigma) [-\sin(\tau) \cos(\tau)]^t [dx_i dy_i]}{a_i} \quad (18)$$

Finalement pour une région homogène, l'angle de tilt correspond à la direction du gradient de fréquence (rapport entre les composantes en  $x_i$  et  $y_i$ ) et l'angle de slant est proportionnel à la norme du gradient du logarithme des fréquences dans la direction du tilt, soit :

$$\tan(\sigma) = \frac{|d \ln(v_i)| f}{\frac{3}{2} |dx| - |d \ln(v_i)| (-\sin(\tau) x_i + \cos(\tau) y_i)} \quad (19)$$

avec  $|dx|$  correspondant au pas unitaire de déplacement spatial.

## 5.2. Extraction de l'orientation de l'image

La méthode décrite ne repose pas sur une procédure d'optimisation comme [4] ou [3]. Au contraire, ici la méthode d'estimation est basée sur l'extraction de gradients locaux de fréquence. Pour ce faire, la surface obtenue après estimation locale de la fréquence est analysée en sous-régions (*i.e* moyennée sur un ensemble d'imagettes). Chaque sous-région est supposée couvrir une surface plane et, ainsi, le tilt et le slant peuvent y être considérés comme constants. Le gradient moyen de fréquence sur la sous-région est calculé : son orientation donne une estimation du tilt et l'équation (19) est utilisée pour obtenir une estimation du slant. Ce calcul est réalisé sur l'ensemble des sous-régions couvrant l'image. La taille des sous-régions est un paramètre modifiant le lissage de la surface obtenue lors de l'intégration sur l'ensemble des estimations locales. Sur l'ensemble des tests réalisés cette taille est fixée à  $10 \times 10$  imagettes. Dans le cas d'une surface courbe, les estimations locales sont utilisées pour obtenir la normale à la surface en chaque position (*i.e*



centre des imagettes). Dans le cas d'une surface plane, la moyenne des estimations locales est utilisée pour obtenir l'orientation de la surface.

La complexité est linéaire pour le prétraitement rétinien, l'étape de combinaison des filtres et l'estimation de la forme finale. La partie la plus coûteuse en calcul est l'étape d'obtention de la réponse des filtres qui utilise une transformée de Fourier sur chaque imagette. Celle-ci dépend également du nombre total d'imagettes (en général  $21 * 21 = 441$  imagettes) et du nombre de filtres (maintenu à  $7 * 7 = 49$  filtres). Avec une fréquence d'horloge du processeur cadencée à 2 GHz, l'estimation de la forme sur une image de  $256 * 256$  pixels prend approximativement 1 minute avec une implantation basique en Matlab.

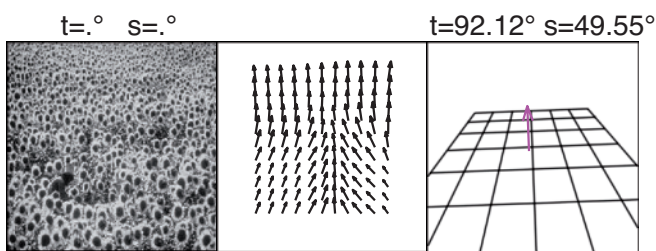


Figure 11. Extraction de l'orientation d'une image ; à gauche : image d'un champ de tournesols ; au milieu : estimations locales du tilt (orientation des flèches) et du slant (longueur des flèches) localisées au centre des imagettes ; à droite : estimation finale de la surface supposée plane et représentée par un quadrillage projeté de manière équivalente (en utilisant les angles d'orientation estimés).

### 5.3. Un algorithme biologiquement plausible

L'algorithme proposé pour extraire la forme à partir de la texture est basé sur une estimation de la fréquence locale et peut être vu comme le résultat d'une combinaison successive de réponses de filtres et d'un moyennage local. Le processus global peut être considéré comme une combinaison relativement simple de réponses de filtres en suivant un modèle uniquement *ascendant*. Ceci est en accord avec un traitement de l'information de texture et des variations dès les premières aires visuelles primaires. La figure 12 présente le schéma décrivant les différentes étapes de l'analyse de la fréquence locale sur l'ensemble du banc de filtres log-normaux et permet de mettre en évidence la simplicité du modèle du mécanisme cortical proposé. Dans chaque imagette, seule l'information du spectre d'amplitude est utilisée, l'information de phase n'est pas prise en compte. Ce modèle correspond à une approche générale de l'analyse corticale et peut être adapté à d'autres applications telles que l'analyse de textures, l'extraction de points caractéristiques ou l'indexation d'images par le contenu.

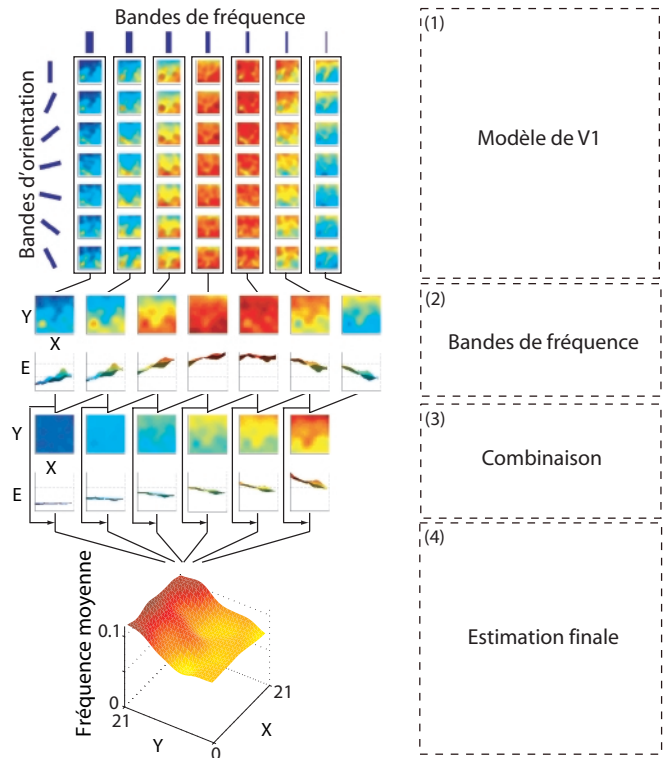


Figure 12. Modèle cortical de l'analyse de la fréquence : (1) décomposition en fréquence et en orientation : chaque vignette représente la réponse d'un filtre log-normal en particulier sur l'ensemble de l'image (i.e. sur chaque imagette pavant entièrement l'image) ; (2) calcul des réponses par bandes de fréquence après la normalisation corticale par orientation ; (3) combinaison des filtres en utilisant l'équation (6) ; (4) estimation finale de la fréquence sur l'ensemble de l'image en utilisant l'équation (7), chaque estimation autour d'une fréquence centrale donnée étant pondérée directement par la réponse par bandes de fréquence associée.

## 6. Résultats

Nous présentons d'abord une évaluation de la précision de la méthode. Celle-ci est comparée avec deux autres techniques développées par Super et Bovik [2] et Hwang *et al.* [9]. La figure 13 présente les différents résultats obtenus sur la base de textures de Super et Bovik. Dans [9], une comparaison de la précision des deux méthodes est présentée sur cette base : les erreurs moyennes obtenues par Super et Bovik sur l'estimation du tilt et du slant sont respectivement de  $3.70^\circ$  et de  $2.84^\circ$  ; Hwang *et al.* obtiennent des erreurs moyennes respectivement de  $1.75^\circ$  et de  $2.18^\circ$ . L'algorithme proposé ici atteint une erreur moyenne de  $2.41^\circ$  pour l'estimation du tilt (sans prendre en compte les résultats pour une inclinaison faible), ce qui est comparable aux autres techniques, et de  $4.95^\circ$  pour l'estimation du slant, ce qui correspond à une précision inférieure. Il est à noter qu'une faible erreur sur l'estimation du slant n'est pas critique

dans le cas de l'analyse de scènes naturelles. En effet, en pratique, il est rare d'avoir à disposition les paramètres de la projection perspective (les valeurs de  $d$  et  $dzw_0$  de la figure 10).

Pour évaluer la robustesse de notre méthode sur différentes caractéristiques de texture, nous avons créé une base composée de 208 exemples de textures (la plupart sont tirées de la collection de textures de Brodatz, certaines sont des textures artificiellement créées, d'autres enfin sont tirées d'images naturelles)<sup>1</sup>. Chaque exemple est d'abord projeté avec un tilt à  $0^\circ$  suivant 3 valeurs de slant:  $30^\circ$ ,  $45^\circ$  et  $60^\circ$ . De même, chaque exemple est également projeté pour un slant fixé à  $45^\circ$  suivant 3 valeurs de tilt:  $0^\circ$ ,  $45^\circ$  et  $90^\circ$ . L'erreur d'estimation est moyennée sur l'ensemble des exemples de texture et sur toutes les projections. Nous obtenons une précision moyenne de  $18.45^\circ$  sur l'estimation du tilt et de  $12.35^\circ$  (variance  $8.1^\circ$ ) sur l'estimation du slant. Il est à noter, qu'à notre connaissance, aucune méthode n'a été évaluée sur une grande base de textures naturelles présentant une très grande variété d'irrégularités comme celle que nous présentons. La figure 14 montre des résultats d'estimation de l'orientation des surfaces texturées provenant de la base utilisée. Des exemples présentant différents types d'irrégularité ont été choisis.

Cependant pour évaluer la précision obtenue, il est possible de la mettre en relation avec les taux obtenus sur une centaine de

textures par la technique développée par Hwang *et al.* [9] et par celle développée par Lelandais *et al.* [10]. La méthode développée par Hwang *et al.* repose sur l'extraction de la fréquence spatiale locale aux points où celle-ci est significative. Les auteurs ont ensuite développé deux méthodes pour estimer l'orientation du plan: rechercher la parabole interpolant au mieux la variation d'échelle locale (inverse de la fréquence); effectuer un vote majoritaire sur les différentes estimations du slant en chaque paire de points aux différentes fréquences centrales du banc de filtre utilisé. La méthode développée par Lelandais *et al.* repose également sur l'extraction des fréquences locales par interpolation des réponses maximales du banc de filtres utilisé. De même que précédemment, l'estimation de l'orientation du plan est obtenue par interpolation d'une parabole pour chaque valeur du tilt. Un critère de minimisation de l'erreur quadratique de l'interpolation est utilisé pour trouver le tilt et en déduire ainsi le slant. Cette méthode apparaît plus robuste aux faibles inclinaisons que la méthode développée par Hwang *et al.* mais le parcours de toutes les valeurs du tilt augmente sensiblement la complexité calculatoire.

Ces deux méthodes obtiennent une précision de l'ordre de  $1^\circ$  sur le tilt. Il est cependant à noter que les textures constituant cette base sont, pour une majorité, relativement régulières. Sur l'estimation du slant les deux méthodes obtiennent respective-

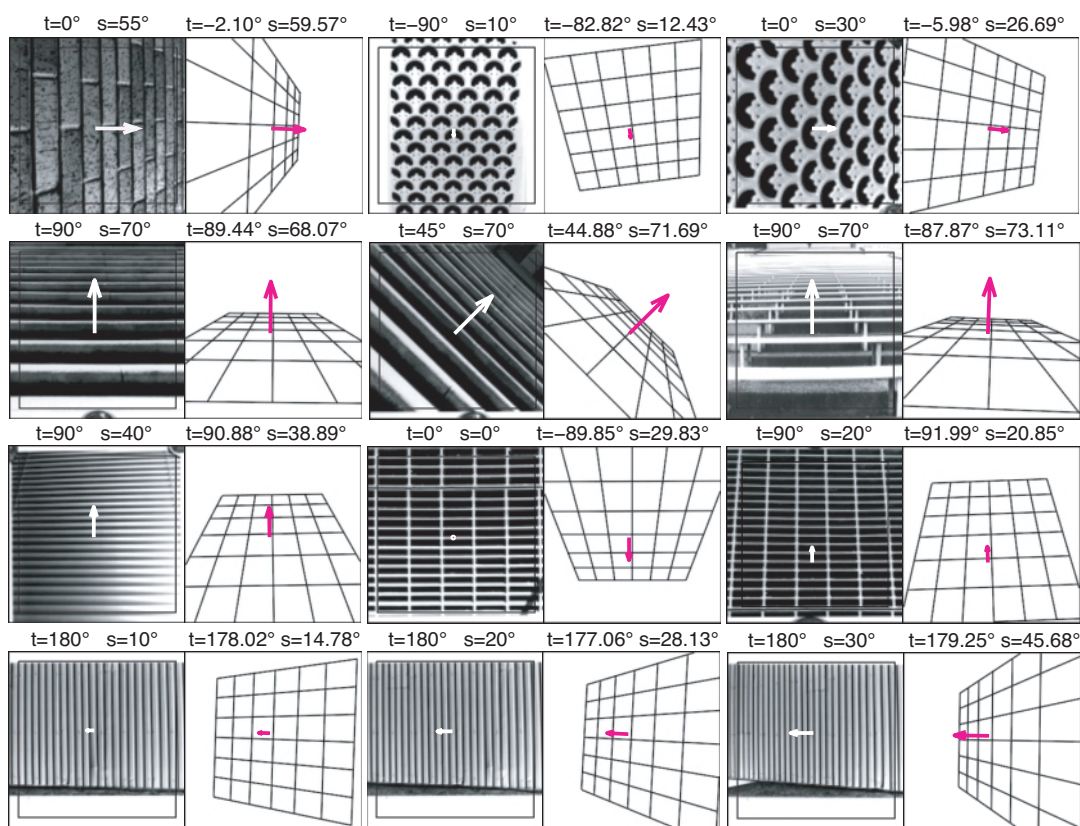


Figure 13. Résultats obtenus sur la base de texture de Super et Bovik [2].

<sup>1</sup> Cette base est accessible à l'adresse suivante:

<http://www.lis.inpg.fr/massot>

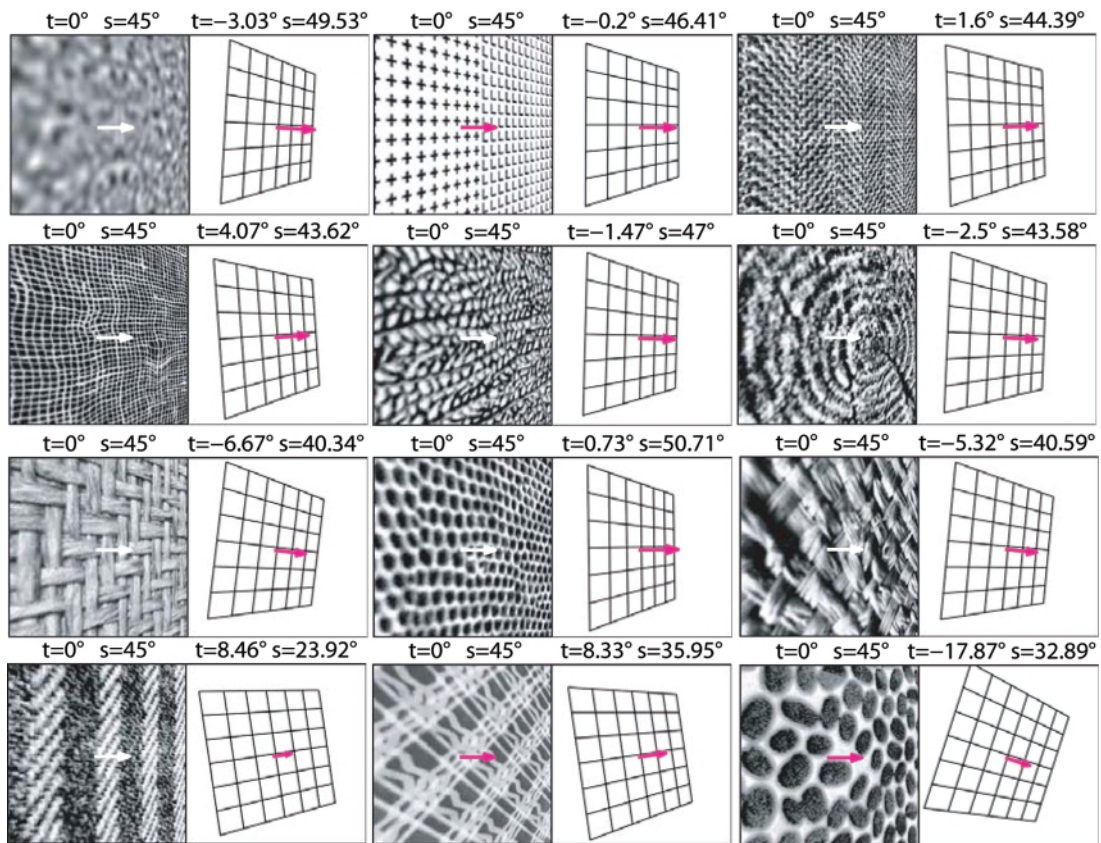


Figure 14. Exemples de résultats obtenus sur notre base de textures (208 exemples, chacun projeté suivant 5 configurations de tilt et slant soit un total de 1040 textures); ces exemples ont été choisis pour refléter les différentes irrégularités présentes dans la base.

ment une précision de  $31.4^\circ$  (variance  $23.3^\circ$ ) et de  $15.9^\circ$  (variance  $14.9^\circ$ ). La précision obtenue par notre méthode est donc comparable tout en présentant, en outre, moins de dispersion (variance plus faible). Ces résultats permettent d'affirmer que les performances de la méthode proposée ici sont comparables à celles obtenues par des techniques spécialement développées pour l'extraction de la forme à partir de la texture lorsque celles-ci sont évaluées sur un grand nombre de textures naturelles. De plus la méthode présentée n'incorporant pas de techniques d'optimisation, son coût calculatoire est inférieur aux deux autres techniques citées.

Finalement, sur la figure 16, nous présentons des résultats sur des scènes naturelles. La première image est tirée de [21], les quatre images suivantes sont tirées de [8] sur lesquelles la technique proposée obtient des résultats tout-à-fait comparables. Les figures 2 et 16 présentent également des estimations sur différentes régions de scènes multi-texturées. Il est à noter que ces résultats sont obtenus avec exactement les mêmes paramètres que pour les précédents tests, exceptée la taille de la zone analysée, choisie manuellement suivant les cas, pour respecter l'hypothèse d'homogénéité (le processus pourrait être rendu automatique à l'aide d'une technique de segmentation région des scènes naturelles, par exemple *normalized cut* [22]). L'ensemble de ces résultats montre la capacité de la méthode présentée à traiter des textures tirées directement de scènes naturelles pour en extraire

les informations de perspective. Il est également à noter que sur les scènes naturelles, la précision de la méthode est suffisante pour extraire une information utilisable, notamment pour comprendre l'organisation spatiale de la scène.

## 7. Conclusion

Dans cet article nous avons présenté un algorithme d'extraction de l'orientation à partir de la texture basé sur l'utilisation de nouveaux filtres (les filtres log-normaux) et l'estimation de la fréquence moyenne locale. Chaque étape s'appuie sur un modèle du système visuel humain, basé sur des données physiologiques et psychophysiques, depuis la rétine jusqu'au cortex primaire et il peut être considéré comme un modèle plausible d'extraction bas-niveau et complètement *ascendant* de l'orientation et de la forme. Nous avons également présenté de nombreux résultats sur différentes bases de données comportant des textures artificielles et naturelles et nous avons montré la capacité de l'algorithme à prendre en compte différents types d'irrégularités. Nous avons montré qu'il peut être appliqué à l'extraction d'une information de perspective dans des scènes naturelles après une étape de segmentation en régions mono-texturées afin de respecter une hypothèse d'homogénéité. Nous



avons montré que les circuits du système visuel primaire permettent vraisemblablement d'extraire une part importante de la perspective monoculaire. Enfin, le modèle développé permet d'aborder le problème de l'extraction de la forme par la texture avec des performances équivalentes aux algorithmes spécialement développés à cet effet avec, en plus, un moindre coût cal-

culatoire. Le développement de ce modèle peut être poursuivi par l'introduction d'un processus de régularisation sur les estimations des fréquences moyennes locales afin d'améliorer la robustesse à certaines irrégularités locales de la texture et également afin d'étendre ce modèle à l'étude des surfaces courbes.

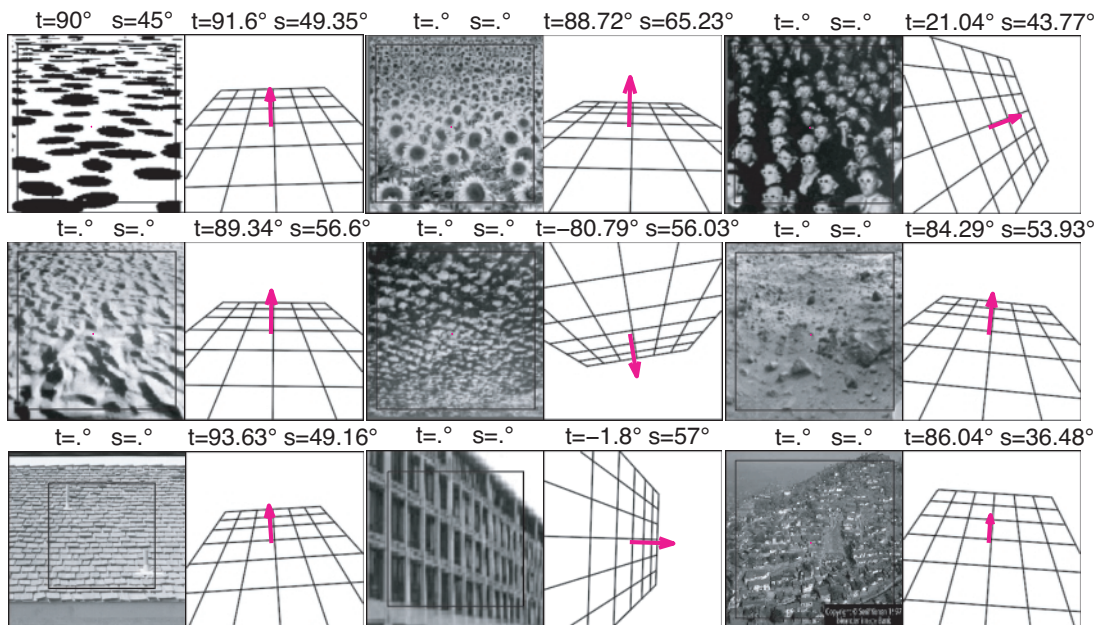


Figure 15. Résultats obtenus sur des scènes naturelles tirées de [21], [8] et sur d'autres images non référencées; il n'y a pas d'estimation théorique accompagnant ces images, excepté pour la première image (le tilt est égal à 90° et le slant est égal à 45°); de la deuxième à la cinquième image les résultats se comparent favorablement à ceux de Lindeberg et Garding ([8]).

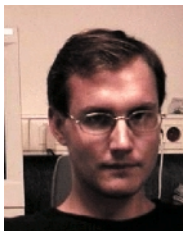


Figure 16. Résultats obtenus sur des scènes naturelles: vue aérienne de Budapest; immeuble de Glasgow; intérieur du musée d'Orsay.



# Références

- [1] A. TORRALBA and A. OLIVA, Depth estimation from image structure. *IEEE Pattern Analysis and Machine Intelligence*, 24, No.9, 2002.
- [2] B.J. SUPER and A.C. BOVIK, Planar surface orientation from texture spatial frequencies. *Pattern Recognition*, 28, No.5, 1995.
- [3] M. CLERC and S. MALLAT, The texture gradient equation for recovering shape from texture. *IEEE Trans. PAMI*, 24, No.4, 2002.
- [4] J. MALIK and R. ROSENHOLTZ, Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 24, No.2, 1997.
- [5] E. RIBEIRO and E.R. HANCOCK, Shape from periodic texture using the eigen vectors of local affine distortion. *IEEE Trans. PAMI*, 23, No.12, 2001.
- [6] K. SAKAI and H. FINKEL, Spatial-frequency analysis in the perception of perspective depth. *Network: Computation in Neural Systems*, 8(3):335-352, 1997.
- [7] A. GUERIN-DUGUE and M. ELGHADI, Shape from texture by local frequencies estimation. *SCIA, Kangerlussuaq (Greenland)*, pages 533-544, 1999.
- [8] J. GARDING and T. LINDBERG, Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2), pp. 163-191, 1996.
- [9] W.S. HWANG, C.S. LU, and P.C. CHUNG, Shape from texture estimation of planar surface orientation through the ridge surfaces of continuous wavelets transform. *Trans. in Image Processing*, 7, No.5, 1998.
- [10] S. LELANDAIS, L. BOUTTÉ, and J. PLANTIER, Shape from texture: Local scales and vanishing line computation to improve results for macrotexels. *Int. J. Image Graphics*, 5(2):329-350, 2005.
- [11] L. SPILLMANN and J.S. WERNER, *Visual Perception: The Neurophysiological Foundations*. Academic Press, Inc., 1990.
- [12] W.H. BEAUDOT, *The neural information in the vertebrate retina: a melting pot of ideas for artificial vision*. PhD thesis, tifr laboratory, Grenoble, France, 1994.
- [13] D.J. HEEGER, Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70:1885-1898, 1993.
- [14] A. LI and Q. ZAIDI, Three-dimensional shape from non-homogeneous textures: carved and stretched surfaces. *Journal of Vision*, 4, No.10(3):860-878, 2004.
- [15] J. ATICK and A. REDLICK, What does the retina know about natural scenes? *Neural Computation*, 4, 1992.
- [16] J. HÉRAULT, A model of colour processing in the retina of vertebrates: from photoreceptors to colour opposition and colour constancy. *Neurocomputing*, 12:113-129, 1996.
- [17] A. TORRALBA and J. HÉRAULT, An efficient neuromorphic analog network form motion estimation. *IEEE Trans. on Circuits and Systems-I: Special Issue on Bio-Inspired Processors and CNNs for Vision*, 46, No.2, 1999.
- [18] H. KNUTSSON, C.F. WESTIN, and G. GRANLUND, Local multiscale frequency and bandwidth estimation. *IEEE International Conference on Image Processing (ICIP'94)*, Austin, Texas, 1994.
- [19] G.H. GRANLUND and H. KNUTSSON, *Signal processing for computer vision*. Kluwer Academic Publishers, ISBN 0-7923-9530-1., 1995.
- [20] R. ROSENHOLTZ and J. MALIK, Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision Research*, 37(16):2283-2293, 1997.
- [21] D.C. KNILL, Discriminating surface slant from texture: Comparing human and ideal observers. *Vision Research*, 38, 1998.
- [22] J. SHI and J. MALIK, Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888-905, 2000.



Corentin Massot

Corentin Massot est actuellement postdoctorant dans le département de physiologie de l'Université McGill à Montréal. Sa recherche se caractérise par une approche multidisciplinaire de la perception de l'espace en mêlant des études théoriques et des expérimentations en psychophysique et en neurophysiologie appliquées à la perception visuelle 3D et au système vestibulaire. Ses travaux concernent également l'interaction Homme-Machine.



Jeanny Hérault

Il est Professeur Emérite au Laboratoire des Images et des Signaux où il a fondé une équipe de recherche sur la Perception Visuelle. Ses thèmes de recherche gravitent autour des modèles de réseaux de neurones artificiels ou naturels: de la séparation de sources à l'analyse de données en grandes dimensions par des réseaux auto-organisés, jusqu'aux modèles de la vision. Dans la thématique de la vision, il a étudié le traitement des signaux dans la rétine (filtrage spatio-temporel, couleur et non-linéarités), et leur exploitation par des modèles du cortex visuel. Les applications visent le traitement adaptatif des images, l'estimation dense du mouvement, la catégorisation de scènes et l'indexation d'images.

