

Extraction d'indices de perspective dans les scènes naturelles par modèles corticaux.

Corentin MASSOT, Jeanny HÉRAULT

Laboratoire des Images et des Signaux (LIS)

46 avenue Félix Viallet, 38031 Grenoble Cedex, France

corentin.massot@lis.inpg.fr, jeanny.herault@lis.inpg.fr

Résumé – Nous nous intéressons à l'extraction d'information de perspective dans les scènes naturelles. Il s'agit d'évaluer les caractéristiques géométriques des surfaces sur lesquelles sont projetées des textures artificielles et naturelles. Nous nous plaçons dans le cadre d'un modèle plausible du fonctionnement du système visuel biologique. Nous divisons l'image en un ensemble d'images. La méthode effectue une analyse par bandes de fréquences dont la variation énergétique est estimée sur toutes les images. Nous montrons expérimentalement que le choix des bandes dont la variation est la plus régulière permet d'obtenir une information de perspective sur un grand nombre de textures.

Abstract – This work aims at extracting the perspective information in natural scenes. Our goal is to evaluate the 3D geometric characteristics of surfaces in which artificial and natural textures are projected. The proposed model can be considered as a realistic assumption of the biological visual system behavior. We divide the image into a set of patches. The method is based on the analysis of bands of frequencies, and the estimation of their energetic variation over every patch. Experimental results show that the choice of the bands whose variation is the most regular, lead us to obtain perspective information of many kind of textures.

1 Information de perspective dans les scènes naturelles

Notre objectif est d'obtenir une information de perspective sur des images de scènes naturelles en s'appuyant sur un modèle biologique du cortex visuel. Nous recherchons une information globale de perspective permettant d'obtenir une indication sur la profondeur relative des zones de l'image et sur la forme de la surface à partir d'une analyse de "textures" au sens large (par exemple: sur laquelle est plaquée la texture).

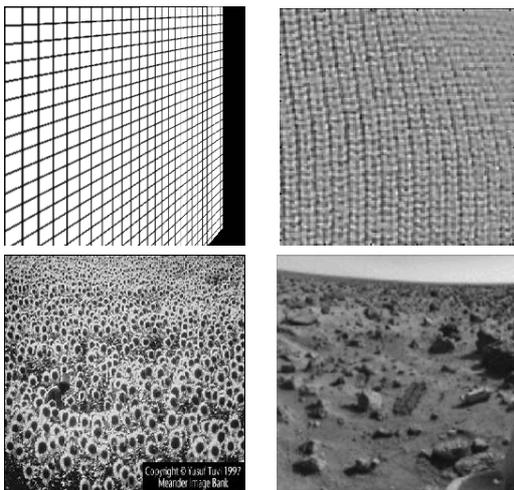


FIG. 1: Exemples de textures étudiées: quadrillage régulier, mailles d'un pull (tiré de [6]), champ de tournesols, surface de mars.

Il existe des méthodes (*shape from texture*) qui donnent de bons résultats sur des textures régulières [1],[2],[3],[4]. Peu

de méthodes traitent également le cas des textures naturelles [5],[6]. Ce problème étant plus difficile compte tenu des irrégularités 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).

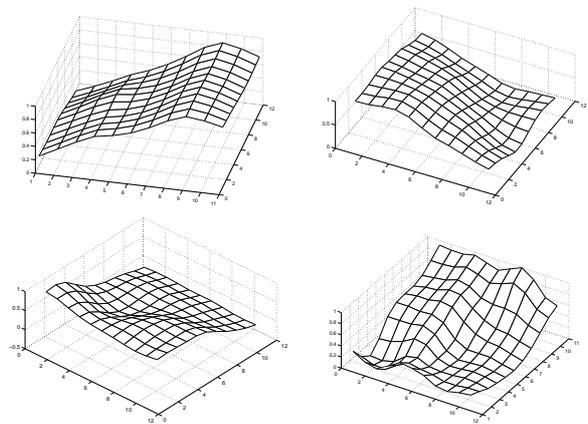


FIG. 2: Résultats obtenus sur les textures de la figure 1; au sol se trouve le plan (x,y) et le troisième axe représente la valeur de l'énergie correspondant à une bande de fréquences; l'angle d'orientation est obtenu au signe près comme il le sera expliqué ultérieurement.

Notre travail s'inscrit dans le cadre d'une étude du fonctionnement du système visuel biologique. Les connaissances en neurophysiologie font apparaître certaines grandes étapes dans le traitement visuel chez les primates. La rétine effectue un ensemble de pré-traitements permettant notamment d'éliminer les variations locales de luminosité [7]. Le signal est ensuite transmis aux premières aires visuelles V1 et V2, dont les cel-

lules simples et complexes effectuent une décomposition en orientations et en fréquences. La réponse des cellules simples peut être modélisée par un filtre passe-bande de type Gabor. Les champs récepteurs de ces cellules représentent chacun une région de l'espace visuel, en recouvrement avec les champs récepteurs voisins. Enfin le signal est modifié par des interactions par bandes de fréquence entre les champs récepteurs des cellules voisines [8].

Une vision monoculaire, statique est suffisante pour permettre au système visuel de retrouver les caractéristiques tridimensionnelles d'une image. La méthode développée ne s'appuie que sur les informations contenues dans la texture. Celles-ci comportent à la fois des indices d'ordre géométrique (par exemple les lignes de fuite) et des indices indépendants du motif (par exemple le zoom produit par la profondeur). C'est ce dernier type d'indice que nous cherchons à exploiter en nous basant sur le modèle simplifié du fonctionnement du système visuel précédemment décrit. Le choix de cet indice impose que la texture soit suffisamment dense pour pouvoir être traitée.

2 Méthode

La méthode est basée sur un découpage de l'image en un ensemble d'images (figure 3), qui permettent une analyse des variations locales de la texture. Leur taille est choisie afin d'avoir suffisamment de signal à analyser (par exemple 96X96 pour des images 256X256). La précision de l'estimation est obtenue en jouant sur le taux de recouvrement des images (par exemple chaque image est décalée de 5 pixels horizontalement et verticalement par rapport à ses voisines).

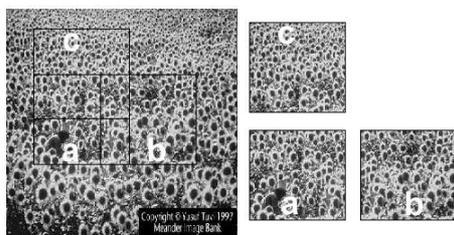


FIG. 3: Découpage d'une image en images en recouvrement local.

Afin de ne prendre en compte que l'information due à la texture, un pré-traitement est nécessaire pour éliminer les variations locales d'intensité. Pour cela nous utilisons un filtre basé sur un modèle de rétine: par un processus de régularisation locale, ce filtre permet d'obtenir une uniformisation des différences d'intensité et aussi un blanchiment du spectre.

Pour procéder à une analyse fréquentielle, une FFT est appliquée sur l'image après l'avoir préalablement multipliée par une fenêtre de Hanning. La décomposition en orientation et en fréquence du module du spectre est effectuée en utilisant une rosace de filtres de Gabor (figure 5). La largeur des filtres augmente exponentiellement avec les fréquences ce qui compense la décroissance du spectre en $1/f$. La largeur des filtres suivant les orientations est calculée pour couvrir deux quadrants successifs du plan des fréquences. Suivant à la fois les fréquences

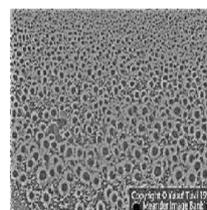


FIG. 4: Image obtenue par le filtre de rétine.

et les orientations, les filtres se recouvrent en partie, permettant un échantillonnage complet de l'espace des fréquences.

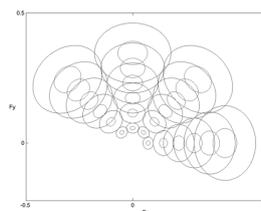


FIG. 5: Rosace de Gabors suivant 4 orientations et 6 fréquences centrales.

En sortie de la rosace on considère l'énergie totale à travers toutes les orientations à un intervalle de fréquences donné (figure 6). On obtient ainsi une réponse par bande de fréquences. Chaque bande est ensuite normalisée par rapport à l'ensemble des réponses de la rosace ce qui permet d'éliminer les légères variations d'intensité d'une image à l'autre.

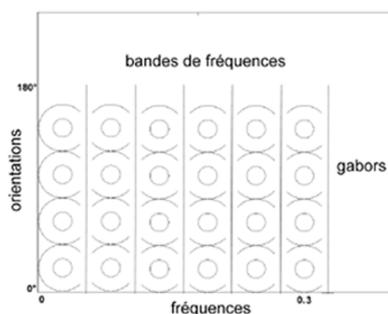


FIG. 6: Représentation de chaque bande de fréquences obtenue en transposant la rosace sur une échelle logpolaire.

La décomposition du spectre en bandes de fréquences présente l'avantage de libérer l'analyse des variations possible de la géométrie de la texture. Par exemple dans le champ de tournesols, les fleurs n'ont pas toutes la même orientation. L'utilisation de filtres de Gabor permet également de réaliser un lissage autour de la fréquence centrale ce qui permet de se libérer des faibles variations d'énergie dans chaque fréquence.

Dans cette étude, nous travaillons sur des textures homogènes et nous faisons l'hypothèse que certaines de ses composantes varient avec l'effet de perspective. Pour cela nous analysons l'image par bandes de fréquences et nous essayons de montrer expérimentalement que la variation d'énergie d'au moins une bande correspond à la géométrie de la surface. Si celle-

ci est douce, développable, alors cette variation doit être relativement régulière. Dans le cas d'une surface plane, parallèle au plan de projection, l'énergie de la bande doit être pratiquement constante. Dans le cas d'une surface orientée ou ayant une forme particulière, la variation doit être en rapport avec le gradient de déformation induit par la projection en deux dimensions de la surface.

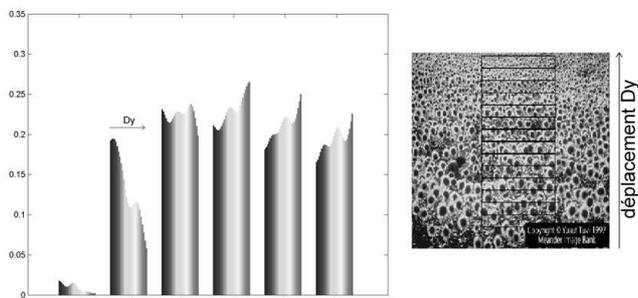


FIG. 7: Graphique représentant les différentes bandes de fréquences à partir du découpage monodimensionnel présenté sur l'image; le déplacement Dy suit la direction de la profondeur.

Pour pouvoir effectuer ce choix, on effectue un découpage en imagerie de l'image, uniquement suivant la direction de la profondeur. La variation d'énergie dans chacune des bandes de fréquence constitue un signal (figure 7). Nous choisissons le signal dont la variation est la plus régulière en regardant le profil de l'évolution de l'énergie dans chacune des bandes. Un graphique du type de la figure 7 permet d'effectuer ce choix avant d'appliquer la méthode sur l'ensemble de la texture. Sur celui-ci on remarque que la bande numéro 2 décroît presque constamment, conformément à l'impression visuelle d'un plan incliné en profondeur donné par le champ de tournesols. Les bandes situées en hautes fréquences ont une variation moins régulière mais les résultats de la figure 8 montrent qu'elles permettent également d'obtenir une information. Les résultats indiquent la variation énergétique dans une bande de fréquences particulière sur l'ensemble de l'image. Au sol se trouve le plan spatial (x,y) et le troisième axe représente la réponse en énergie correspondante.

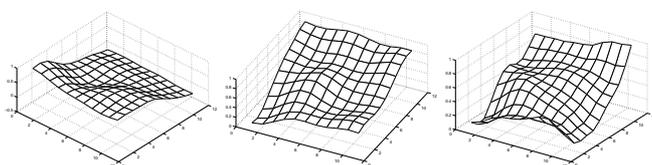


FIG. 8: Résultats obtenus après lissage de la réponse pour les bandes n°2, n°4 et n°6 de l'image du champ de tournesols; on remarquera le changement de signe de l'angle d'inclinaison.

Sur le graphique de la figure 7, il est également possible d'observer un *basculement* entre la décroissance des réponses des bandes situées en basses fréquences avec la profondeur et l'évolution inverse pour les bandes situées en hautes fréquences. Ceci conduit à une estimation de l'angle d'inclinaison au signe près.

Les textures étudiées peuvent contenir des non-stationnarités correspondant à la présence d'autres éléments dans la texture (par exemple le personnage dans le champ de tournesols) ou à l'absence d'éléments. Ces inhomogénéités viennent perturber le signal stationnaire recherché. Une première approche consiste à lisser le signal bidimensionnel obtenu. Pour cela nous utilisons une procédure de régularisation robuste. Ce processus peut être mis en oeuvre par une combinaison des sorties des filtres de Gabor.

3 Expérimentations

Nous avons appliqué la méthode décrite précédemment sur différents types de texture. La figure 9 montre des résultats sur des textures régulières. Les surfaces de ces textures étant planes, nous présentons les résultats dans le plan (x,I) (des abscisses et de la réponse en énergie). Cela permet d'observer le changement progressif d'inclinaison bien visible entre la première et la troisième image.

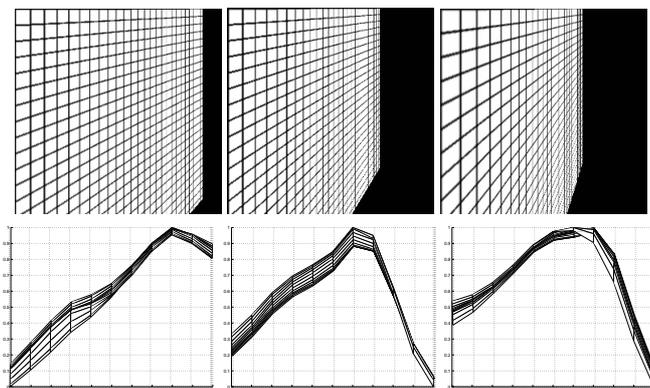


FIG. 9: Textures régulières de quadrillages; les surfaces sont orientées progressivement avec un angle d'inclinaison respectivement de 13° , 18° et 23° .

La figure 10 montre des résultats sur des textures naturelles, avec des irrégularités croissantes. Les surfaces étant toutes inclinées de 13° , cela permet de comparer les résultats pour des textures différentes et d'observer que la réponse obtenue est globalement la même.

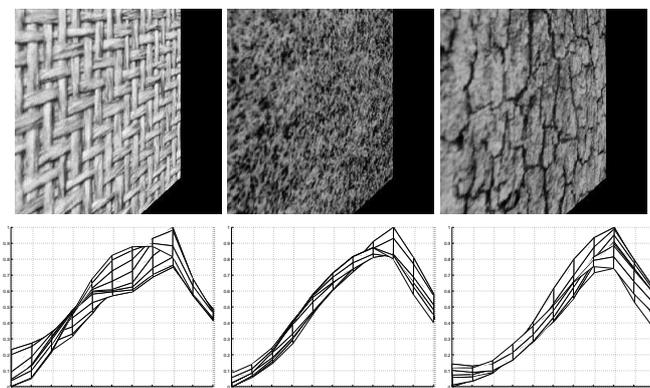


FIG. 10: Textures naturelles avec le même angle d'inclinaison de 13° .

La figure 11 montre des résultats sur des textures artificielles de difficulté croissante projetée sur des surfaces cylindriques. On remarquera la correspondance entre les résultats et l'impression visuelle donnée par ces images.

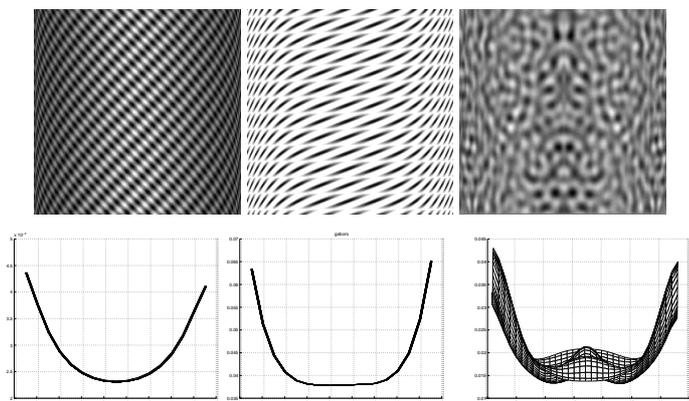


FIG. 11: Textures artificielles sur une surface cylindrique; la troisième image est tirée de [3] sur laquelle peu de résultats ont été montrés.

La figure 12 montre des résultats sur d'autres textures naturelles et artificielles qui présentent des irrégularités.

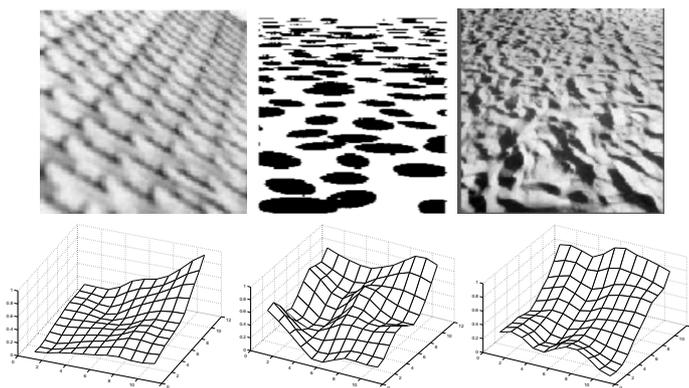


FIG. 12: Textures naturelles et artificielles; la deuxième est tirée d'expériences psychologiques menées par Knill [9]; la troisième est tirée de [5].

4 Discussion

Ce papier présente une étude expérimentale montrant la possibilité de récupérer une information globale de perspective pour des textures variées par une modélisation simplifiée du système visuel. Un des aspects intéressant de la méthode est la ressemblance entre les réponses obtenues et l'impression visuelle donnée par les surfaces. L'utilisation de bandes de fréquences est très similaire à l'analyse par *warplets* uniquement en hautes fréquences introduite par Clerc et Mallat [6]. Cependant cette étude montre qu'il est possible d'obtenir une information sur la géométrie de la texture en utilisant également des bandes situées en basses fréquences. Il est également possible d'envisager des combinaisons entre bandes à la manière des interactions corticales. Le choix de la bande à analyser peut se ramener à un problème d'identification d'un signal bidimen-

sionnel. L'association de la variation de ce signal avec les paramètres géométriques de la surface supportant la texture reste également à faire.

Références

- [1] J. Super, A. Bovik. *Shape from Texture Using Local Spectral Moments*. IEEE Trans. PAMI, 17, mai 1995.
- [2] J. Malik, R. Rosenholtz. *Computing Local Surface Orientation and Shape From Texture for Curved Surfaces*. Int. J. of Computer Vision, 23(2), 1994 (republié en 1997).
- [3] K. Sakai, L. Finkel. *Characterisation of Spatial Frequency in the Perception of Shape From Texture*. J. Opt. Soc. Am., A12, 1995.
- [4] W. Hwang, C. Lu et P. Chung. *Shape from Texture: Direct Estimation of Planar Surface Orientation Through the Ridge Surfaces of Continuous Wavelet Transform*. IEEE Trans. Image Processing, 7, 1998.
- [5] T. Lindeberg, J. Garding. *Shape from Texture from a Multi-Scale Perspective*. Int. Conf. on Computer Vision, 1993.
- [6] M. Clerc, S. Mallat. *The Texture Gradient Equation for Recovering Shape from Texture*. IEEE Trans. PAMI, 24(4), avril 2002.
- [7] J. Héroult, W. Beaudot. *The Vertebrate Retina: a Model of Spatio-Temporal Image Filtering*. 14ème colloque GRETSI, 1993.
- [8] L. Spillmann, J. Werner. *Visual Perception, the Neurophysiological Foundations*. Academic Press, 1990.
- [9] D. Knill *Discriminating Planar Surface Slant from Texture: Human and Ideal Observers Compared*. Vision Research, 38, 1998.