Texture et Perception 3D dans les Scènes Naturelles : Modèles d'Inspiration Biologique et Expérimentations Psychophysiques.

Corentin Massot

UNIVERSITE JOSEPH FOURIER DE GRENOBLE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UJF

Spécialité : "Science Cognitive"

préparée au Laboratoire des Images et des Signaux
dans le cadre de l'École Doctorale
"Ingénierie pour la Santé, la Cognition et l'environnement"

présentée et soutenue publiquement

par

Corentin Massot

le 07/09/2006

Titre :

Texture et Perception 3D dans les Scènes Naturelles : Modèles d'Inspiration Biologique et Expérimentations Psychophysiques.

Directeurs de thèse : Jeanny HÉRAULT et Pascal MAMASSIAN

JURY

M	CHRISTIAN MARENDAZ,	Président
M	KENNETH KNOBLAUCH,	Rapporteur
M	PHILIPPE GAUSSIER,	Rapporteur
MME	MAUREEN CLERC,	Examinatrice
M	JEANNY HÉRAULT,	Directeur de thèse
M	PASCAL MAMASSIAN,	Directeur de thèse

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury :

- Monsieur Christian Marendaz qui m'a fait l'honneur d'être le président de mon jury. J'ai eu l'occasion de le rencontrer à maintes reprises durant ma thèse au LIS et au LPNC et je le remercie de m'avoir toujours ouvert les portes de son laboratoire pour pouvoir y effectuer des expériences.
- Monsieur Philippe Gaussier ainsi que Monsieur Kenneth Knoblauch pour avoir bien voulu être les rapporteurs de ma thèse et pour m'avoir posé des questions pertinentes qui montrent qu'il est nécessaire d'approfondir ce travail. Je les remercie également pour leur rapport pour lequel ils n'ont eu que peu de temps.
- Madame Maureen Clerc pour avoir accepté d'être examinatrice de ce travail en tant que spécialiste du problème de l'extraction de la forme par la texture et pour ses questions précises et pertinentes.

Je tiens à remercier mes directeurs de thèse :

- Jeanny pour m'avoir suivi depuis le DEA, pour m'avoir laissé libre de choisir la voie dans laquelle je me suis lancé, pour avoir toujours répondu présent lors de mes (pas assez) nombreuses sollicitations et pour m'avoir mis en contact très facilement avec Pascal Mamassian.
- Pascal pour avoir bien voulu co-diriger ma thèse, pour m'avoir invité dans son laboratoire à Glasgow pendant 5 mois où j'ai pu découvrir le monde de la psychophysique, enfin pour être toujours enthousiaste et encourageant malgré la distance.

Je remercie également l'ensemble des personnes que j'ai pu côtoyer durant ces 5 années passées à Grenoble. Tout d'abord mes collègues de travail Zakia, Mickaël et Pierre. Les étudiants que j'ai pu encadrer durant plusieurs mois Christian, Javier et Christophe. Mes collègues de laboratoire dont Guillermo, Brice, Nicolas, Reza, Massoud, Stéphane, Michele, Denis, Hervé.

Je remercie enfin mes amis que je connais depuis le DEA et avec qui je suis resté proche : Ronan, Jérôme, Chloé, Magalie et Razika.

Je remercie enfin mes parents pour m'avoir soutenus durant toutes ces années en grande partie par téléphone interposé.

Table des matières

FIG. 1 – Musée d'Orsay, Paris, 2003.

Chapitre 1

Introduction : étude de la perception 3D dans les scènes naturelles

La figure 1, présentée à la page précédente, représente l'intérieur du musée d'Orsay à Paris. Lorsque nous observons cette image, nous pouvons dire très rapidement qu'il s'agit d'une image prise à l'intérieur d'un bâtiment ; qu'il est profond et que le point le plus éloigné de nous dans la scène se situe à peu près en face au niveau de l'horloge ; la scène présente un axe central, une allée, dont les bords convergent en s'éloignant de nous ; que le toit est en demi-cercle ; que le mur du fond est vertical ; que les murs sur les côtés de l'allée sont rectilignes ; qu'au premier plan se trouve un objet, une statue, représentant un globe sphérique ; que dans l'allée se trouvent une multitude de personnes irrégulièrement réparties ; que les lustres sont situés en ligne suivant la direction de l'allée centrale. L'ensemble de ces informations mêlées à certaines connaissances préalables sont suffisantes pour pouvoir décrire la scène de la manière suivante : cette scène se situe dans un endroit publique, fermé, contenant des objets d'une certaine dimension et avec la caractéristique d'avoir une partie supérieure en arc de cercle. Cela donne une description se rapprochant assez d'un musée ou d'une gare, ce qui est le cas du musée d'Orsay construit dans une ancienne gare de chemin de fer ouvert en 1900 et transformée en un musée des oeuvres de la deuxième moitié du XIXème siècle inauguré en 1986.

L'étude cette image nous amène ainsi à nous poser deux types de questions : comment le système visuel fait-il pour extraire et interpréter les informations de la scène de manière aussi précise et détaillée au point de pouvoir reconnaître le musée d'Orsay ? Si nous avions à disposition une grande quantité d'images, la description que nous venons de donner permettrait-elle de retrouver cette image dans la base d'images ? La première question est du domaine des sciences cognitives et de l'étude du fonctionnement du système visuel humain. La seconde question concerne le domaine du traitement d'image et de la reconnaissance de forme. Dans le travail présenté nous allons essayer d'apporter des éléments de réponse à ces deux questions.

Les scènes naturelles sont des images de l'environnement tel que des paysages de campagne, de plages, de montagnes, des prises de vue de villes ou des intérieurs d'habitation. Les informations contenues dans ces images sont nombreuses et correspondent à des niveaux d'interprétation différents. Par exemple la couleur, les formes, la détection de personnes sont des informations générales, dites de bas-niveau, tandis que la reconnaissance explicite qu'il s'agit du musée d'Orsay est une information précise, dite de haut-niveau. Afin de pouvoir extraire ce

type d'information, tout système (biologique ou informatique) s'appuie sur l'extraction d'un minimum d'informations de bas-niveau aussi précises que possible, que nous appellerons les *indices* de l'image. Dans ce travail nous nous intéressons au système visuel de l'être humain et à comment celui-ci réalise l'extraction et l'analyse d'indices présents dans les images de scènes naturelles.

L'étude du fonctionnement du système visuel est un domaine extrêmement actif. Il permet tout d'abord de contribuer au développement des connaissances sur le fonctionnement du cerveau humain. Du point de vue des applications médicales, certaines pathologies sont induites par des déficit visuels (par exemple certaines formes de dyslexie seraient dues à des déficits de la voie magnocellulaire reliant la rétine aux cortex visuel). Au delà de ces recherches en sciences cognitives, cette étude s'applique également à de nombreux domaines en vision par ordinateur. Notamment de plus en plus d'applications requièrent une administration simple et efficace des bases de données contenant de grandes quantité d'images (par exemple la recherche d'images spécifiques sur le Web ou dans une grande base telle que celle de l'Institut National de l'Audiovisuel (INA)). Actuellement peu de systèmes sont capables de donner une description des images à partir de leur contenu et requièrent en général une annotation manuelle ou, lorsque cela est possible, ils se basent sur le nom et l'environnement textuel dans lequel se trouve l'image (par exemple le moteur de recherche *Google image*). Le développement de systèmes d'analyse d'images basés sur le contenu est ainsi un problème encore ouvert et représente un domaine de recherche très actif.

Dans ce travail nous nous intéressons en particulier à l'information tridimensionnelle (3D) contenue dans les images. Plus spécifiquement nous nous intéressons à l'extraction de l'information d'orientation et de forme des différentes régions de l'espace (figure 1.1). Notre étude s'attache ainsi à analyser et modéliser la capacité du système visuel à extraire l'information 3D dans les images. Un des objectifs applicatifs est l'incorporation de cette information dans l'analyse effectuée par un système d'indexation automatique de bases d'images.

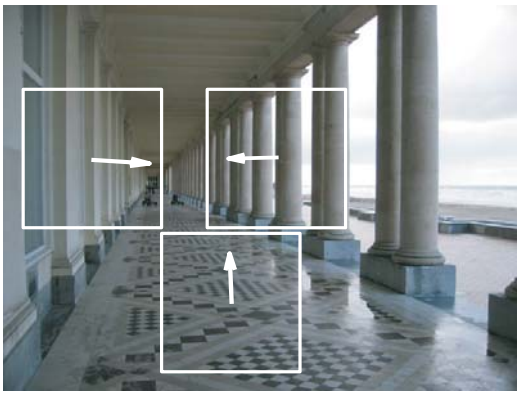


FIG. 1.1 – Exemple d'extraction de l'information 3D dans une scène naturelle (corridor du château royal à Ostende) ; une estimation de l'orientation locale est obtenue sur des régions de l'image constituées d'une texture homogène plus ou moins régulière.

La perception 3D dans les images

La projection du monde réel en 3 dimensions sur le fond de l'oeil (la rétine) ou sur le plan de l'image (photographie) induit des déformations affines des éléments composants la scène ou des

éléments structurels recouvrant une surface (et formant ainsi une texture). Ces déformations créent des indices permettant d'interpréter la scène ou la surface observée en 3D (figure 1.2).

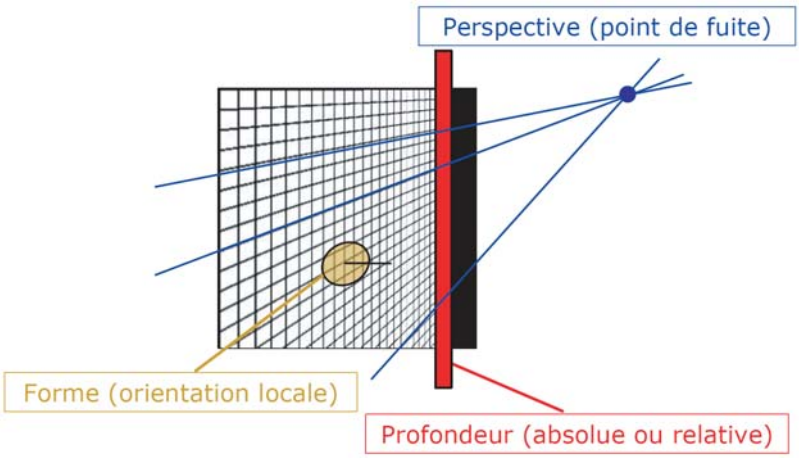


FIG. 1.2 – Attributs 3D sur un plan texturé incliné dans l'espace : la perspective (lignes de fuite convergeant vers le point de fuite) ; la profondeur (absolue ou relative) ; la forme (par intégration de orientations locales sur l'ensemble de la surface).

A partir de l'extraction et l'analyse de ces indices, l'information 3D peut être représentée de différentes manières. Il est ainsi possible de distinguer 3 types d'attributs associés à une surface (figure 1.2) : la **perspective**, la **profondeur** et la **forme**. La perspective correspond au type de point de vue sous lequel est perçue la scène. Cette information est portée par des lignes (lignes de fuite) convergeant vers un unique point appelé le *point de fuite* (indices d'orientation). La profondeur correspond à la mesure de la distance moyenne à l'observateur de la scène entière (profondeur absolue) ou de la distance relative entre différents points de la scène (profondeur relative). La forme correspond à l'intégration d'un ensemble d'estimations locales de l'orientation sur une surface plane ou courbe, permettant ainsi d'effectuer une reconstruction en 3 dimensions de la surface. Cette nomenclature nous amène à nous poser des questions sur le fonctionnement du système visuel : est-ce que celui-ci associe une représentation spécifique de chaque attribut ? Conjointement, existe-il un mécanisme spécifique d'extraction pour chaque attribut ?

Le système visuel est un outil particulièrement puissant et efficace pour analyser l'environnement visuel. Cependant savoir comment celui-ci arrive à extraire et interpréter les informations contenues dans le champ visuel reste un problème encore largement irrésolu. Les principaux mécanismes participant au processus d'analyse et de reconnaissance (figure 1.3) : projection sur la rétine, transmission par le nerf optique à travers le corps genouillé latéral (CGL) et projection sur le cortex visuel primaire (aire V1) situé dans la partie arrière de la structure cérébrale (lobe occipital).

Le système visuel utilise différents indices pour interpréter l'environnement visuel en 3D. Il est possible de distinguer les indices dus à la physiologie du système visuel (convergence et accommodation), les indices binoculaires (stéréoscopie), la parallaxe de mouvement (lorsque l'observateur est en mouvement, les objets proches ont un déplacement spatiale plus important

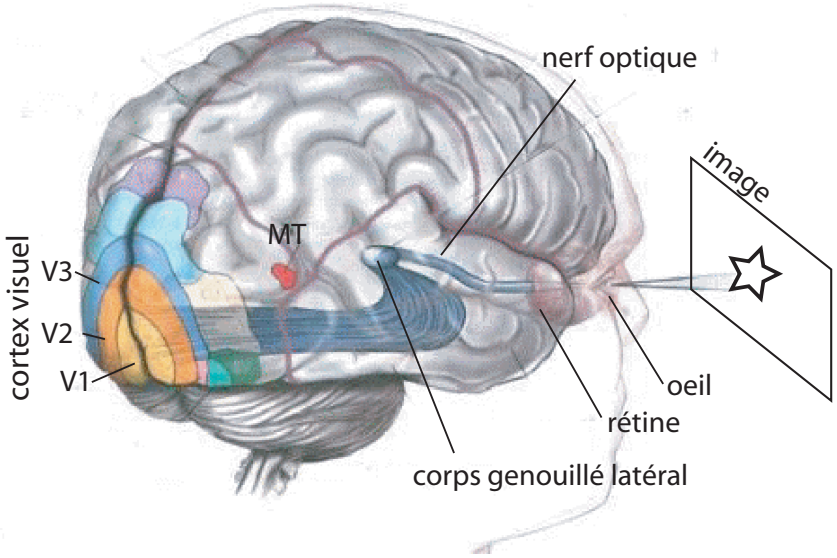


FIG. 1.3 – Premières étapes de traitement de l'information visuelle par le système visuel humain.

que les objets éloignés) et tout un ensemble d'indices monoculaires basés sur une information statique, correspondant notamment à l'information contenue dans les images. Il est possible de distinguer les gradients de texture (la déformation due à la projection induit des variations de la structure des éléments recouvrant la surface), la perspective linéaire (des lignes parallèles dans le monde 3D convergent en un point lors de la projection sur une surface 2D), la perspective atmosphérique (le contraste et la focalisation diminue avec la profondeur), la variation de taille (deux objets de taille identique apparaissent de taille différente lorsqu'ils sont disposés à des profondeurs différentes), la variation de la réflection de la lumière (par exemple cet indice permet d'interpréter une boule de billard comme une sphère circulaire) et l'occlusion (deux objets disposés l'un derrière l'autre permet d'interpréter la scène en 3D).

Les indices physiologiques, les indices binoculaires et les indices de mouvement permettent d'obtenir une mesure précise de la distance d'un objet à l'observateur. Ces indices sont particulièrement utiles dans des tâches telles que la prise d'objets ou la mesure de la distance à un obstacle. Ils ne sont cependant efficaces qu'à une portée limitée par rapport à l'observateur (quelques mètres). Au contraire les indices monoculaires statiques apparaissent particulièrement efficaces pour des distances plus importantes et notamment pour l'interprétation des scènes visuelles en 3D. En effet comme décrit précédemment un observateur est capable d'extraire l'organisation spatiale dans une image de scène naturelle en niveau de gris. Dans ce contexte deux indices semblent jouer un rôle particulièrement important car ils apparaissent de manière systématique : les gradients de texture et la perspective linéaire. Ce sont les indices que nous étudions plus précisément dans ce travail.

Une approche pluridisciplinaire

Comment le système visuel interprète-il l'information 3D présente dans les scènes naturelles ?

Cette question est à la base du travail présenté. Les éléments qui viennent d'être introduits sur le fonctionnement du système visuel et sur les indices participant à la perception 3D per-

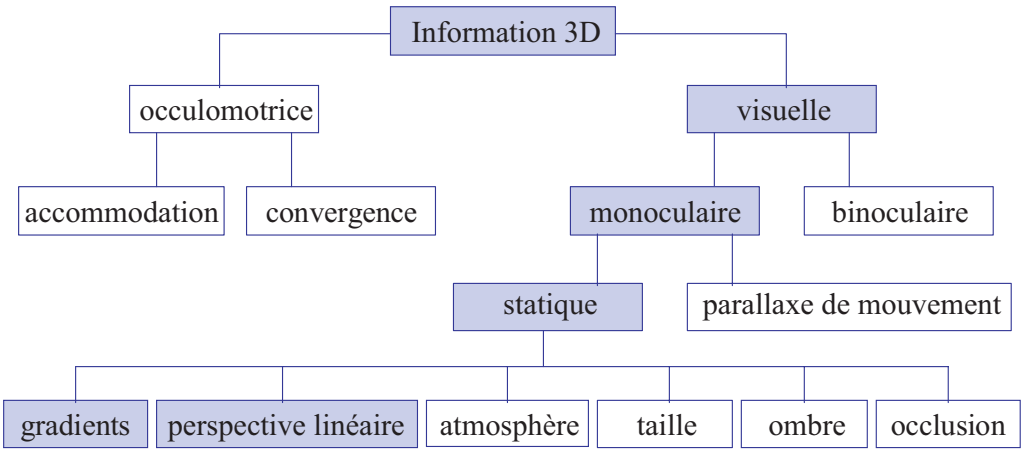


FIG. 1.4 – Indices de perception de l'information 3D ; les indices grisés correspondent à ceux disponibles dans une image et étudiés dans le travail présenté.

mettent d'apporter de premiers éléments de réponse : il est possible que le système visuel code différents types de représentations de la 3D (les attributs) en se basant sur l'extraction et l'analyse d'informations présentes dans les images (les indices).

Il s'agit maintenant de savoir quels sont les processus réellement mis en jeu. Comment sont extraits les indices de l'image, notamment les gradients de texture et la perspective linéaire ? Y-a-t-il un mécanisme spécifique au traitement de chaque indice ? A partir d'un modèle des premières étapes du système visuel est-il possible d'extraire des informations 3D à partir d'une image ? Les réponses à ces questions peut servir alors de base au développement d'un modèle du système visuel de manière biologiquement plausible.

Pour tenter d'apporter des éléments de réponse à ces questions nous adoptons dans ce travail une approche pluridisciplinaire en s'appuyant sur les travaux en neurophysiologie, sur des expérimentations psychophysiques et sur des modèles computationnels inspirés de la biologie.

Les études en neurophysiologie ont permis de mettre en évidence une stratégie du système visuel consistant à décomposer l'information visuelle en un ensemble de composants élémentaires (couleur, intensité lumineuse, disparité binoculaire, fréquences spatiales, orientations) qui sont ensuite combinées de manière à obtenir des informations de plus en plus complexes (les formes, le mouvement puis les objets jusqu'à la reconnaissance de la scène entière). L'information transite et remonte dans les aires supérieures (V1 V2 V3 MT, jusqu'aux aires frontales liées au raisonnement). Cependant il est important de noter que ce processus est beaucoup plus complexe et non-linéaire, notamment du fait de l'existence de nombreuses boucles de retour des informations traitées dans les aires supérieures vers les aires inférieures (ainsi 80% des fibres afférentes du CGL proviennent des aires supérieures et uniquement 20% directement de la rétine).

Les travaux en psychophysique s'attachent à étudier les réponses et les performances d'observateurs humains (les sujets) lorsque ceux-ci doivent réaliser une tâche visuelle particulière en observant des stimuli (par exemple une image crée artificiellement et dont certaines propriétés sont contrôlées par l'expérimentateur). Ces travaux permettent de mettre en évidence

le type d'information utilisée dans les stimuli et leur importance relative dans la réalisation de la tâche étudiée.

La modélisation des mécanismes biologiques s'attache à reproduire le fonctionnement du système visuel depuis le comportement des cellules corticales jusqu'aux performances obtenues par des expérimentations psychophysiques sur des tâches de perception complexes. La modélisaiton peut ainsi s'effectuer à différents niveaux : certaines modélisation étudient les processus chimiques liés au fonctionnement des cellules (approche biologique) ; d'autres étudient les réponse des neurones, leur comportement dynamique et spatio-temporel (approche neurocomputationnelle) ; enfin d'autres modélisent une fonction particulière réalisée par le système visuel (la perception du mouvement, de la couleur, de la 3D) (approche fonctionnelle).

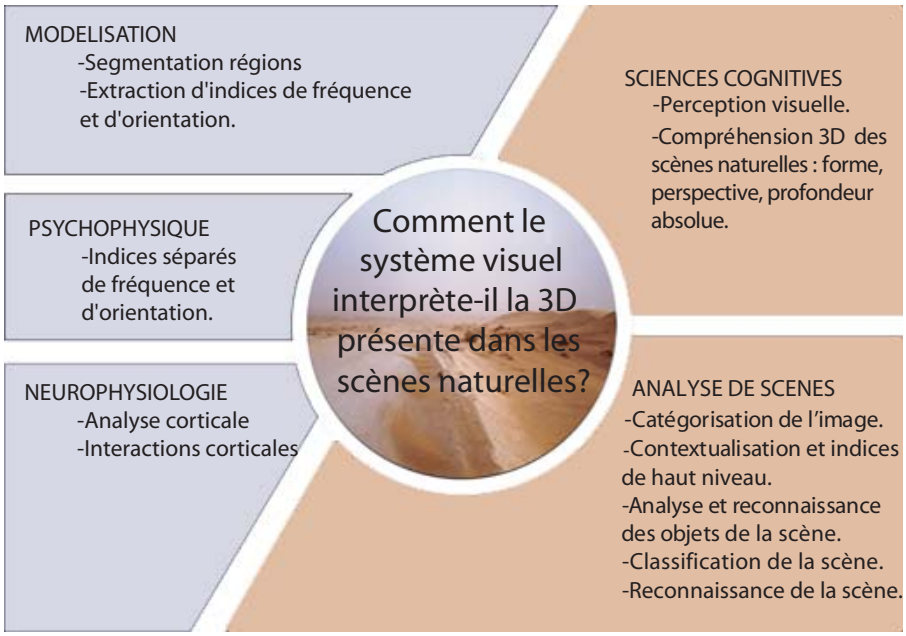


FIG. 1.5 – Approche pluridisciplinaire de la perception 3D dans les scènes naturelles ; au centre : la question posée dans le cadre de ce travail ; à gauche : les différentes disciplines étudiées ; à droite : deux des domaines d'application de ce travail.

L'association de ces trois disciplines n'est pas un travail immédiat. En effet chaque domaine étudie les problèmes qu'il rencontre en fonction des outils d'investigation qui lui sont propres : les expérimentations perceptives en psychophysique ; les outils de formalisation mathématique et une mise en oeuvre informatique pour la modélisation ; les réponses *in vivo* des cellules et l'image de leur activation (neuroimagerie médicale) en neurophysiologie. La deuxième difficulté résulte du fait que chaque domaine ne travaille pas à la même échelle d'analyse, par exemple : la psychophysique travaille au niveau de l'apparence visuelle du stimulus et étudie les performances obtenues (étude comportementale) ; la modélisation simplifie les mécanismes biologiques, souvent trop complexes, afin de pouvoir formaliser le processus et d'en réaliser une simulation ; la neurophysiologie cherche à cartographier la structure des différents groupes de cellules (repérés par zones ou aires) et à identifier quels sont les types d'information qui activent ces groupes de cellules.

Cependant suivant certaines restrictions, une véritable interaction peut être crée entre ces disciplines. Par exemple cela peut être réalisé si en psychophysique l'information étudiée peut être modélisée (par exemple le mouvement, les formes, les fréquences spatiales, les orientations) ; si en vision par ordinateur le modèle développé reprend les schémas principaux du système visuel (par exemple les premières étapes de traitement de l'information visuelle en conservant l'organisation corticale) ; enfin si en neurophysiologie la réponses du groupe de cellules étudiées peut être associée à un type bien défini d'information (par exemple le mouvement, la couleur, les fréquences spatiales, les orientations). Le modèle développé peut alors s'appuyer sur les hypothèses émises (notamment le type d'information extraite) et s'attacher à reproduire les résultats obtenus dans chaque domaine. Ce modèle peut alors servir de base à de nouvelles recherches dans chacun des domaines en suscitant de nouvelles questions sur le fonctionnement réel du système visuel.

Objectif

L'objectif de ce travail de thèse est d'analyser et d'extraire des informations à partir du contenu des images en se basant sur l'étude et la modélisation du système visuel. Dans ce travail nous proposons une étude fonctionnelle de la perception 3D à partir de la texture en s'appuyant sur une approche pluridisciplinaire en mêlant les connaissances acquises en neurophysiologie, des expérimentations psychophysiques et le développement d'algorithmes de traitement basés sur la modélisation des premières étapes du système visuel.

Plan du document

L'organisation générale du travail proposé est la suivante :

Le chapitre 2 intitulé **Des scènes naturelles aux textures** présente le problème de l'analyse des scènes naturelles. Il décrit l'approche retenue consistant à considérer une scène comme étant constituée d'un ensemble de régions composées d'une texture homogène. Une description des principales caractéristiques des textures est présentée ainsi que la projection perspective utilisée. Enfin une revue des principaux modèles d'analyse de la forme à partir de la texture est proposée.

Le chapitre 3 intitulé **Neurophysiologie du système visuel** présente une description de l'architecture du système visuel depuis la rétine jusqu'à l'aire visuelle primaire V1. L'accent est particulièrement mis sur la description des cellules corticales et la notion de champ récepteur. Enfin des travaux appuyant la thèse de l'existence de corrélas neuronaux de la perception 3D à partir de la texture sont présentés.

Le chapitre 4 intitulé **Perception 3D : les indices de texture** présente une revue des principaux travaux et résultats obtenus en psychophysique. L'accent est particulièrement mis sur les travaux récents de Li et Zaidi appuyant la thèse d'une analyse séparée des informations de fréquence et d'orientation pour la perception 3D.

Le chapitre 5 intitulé **Perception 3D : gradient de fréquence et perspective linéaire** présente nos expérimentations psychophysiques sur l'étude des indices de variation de fréquence et de variation d'orientation (perspective linéaire) et leur combinaison. Il s'attache tout d'abord à décrire les stimuli crée spécifiquement pour cette étude puis à analyser les résultats obtenus sur deux tâches de discrimination de l'inclinaison et de l'orientation.

Le chapitre 6 intitulé **Modèle de V1 pour la perception 3D** présente un nouveau type de modèle de cellules corticales : les filtres *log-normaux*. Nous décrivons ensuite un modèle biologiquement plausible d'analyse de la variation de fréquence au niveau de V1 pour extraire l'orientation de surfaces planes et la forme de surfaces courbes. La méthode est évaluée sur

différentes bases de textures et d'images de scènes naturelles et elle est comparée à d'autres techniques existantes.

En **Conclusions** nous présentons une synthèse des travaux réalisés et des perspectives à court et à long terme ouvertes par ce travail.

Chapitre 2

Des scènes naturelles aux textures

Ce chapitre introduit le problème de l'analyse de scènes ainsi que notre approche par analyse locale en régions de texture homogène. Une taxonomie des caractéristiques de la texture et des problèmes rencontrés pour son analyse est présentée. Afin de reproduire le passage du monde 3D au plan de l'image (ou à la surface de la rétine), nous présentons la projection perspective permettant de modéliser les déformations géométriques subies par les éléments de la texture. Nous présentons enfin une revue des modèles existants en extraction de la forme à partir de la texture.

2.1 L'analyse des scènes naturelles

L'approche classiquement retenue pour créer un système d'analyse d'image par le contenu est de conserver avec l'image un ensemble de descripteurs qui décrivent les principales caractéristiques de l'image (par exemple : histogramme de couleur, histogramme des contours, réponses de filtres décrivant la scène suivant les orientations et les fréquences spatiales présentes (filtres de Gabor)). Différents systèmes ont ainsi été développés (par exemple Blobworld [CTB+99], voir [KZB04] pour un état-de-l'art complet). Cependant ces systèmes ne réalisent pas automatiquement de regroupement des images en fonction de leur proximité sémantique, ils ne font que chercher les résultats les plus proches d'une image d'entrée (au sens des descripteurs utilisés). L'étude du système visuel permet d'adopter une approche différente en servant de guide au développement de modèles dédiés à l'extraction d'informations de plus haut niveau dans les images (par exemple la catégorie de l'image et l'organisation spatiale).

Les scènes naturelles sont des images de l'environnement auquel le système visuel est soumis au quotidien. La figure 2.1 présente quelques exemples de scènes naturelles.

Le système visuel est capable de récolter un très grand nombre d'informations sur une image incroyablement rapidement. Différentes expériences ont montré qu'en une seule fixation oculaire et avec un temps de présentation extrêmement court, les sujets sont capables de donner une description de la scène (temps de présentation TP<300ms), de reconnaître une scène cible (TP<120ms) [Pot75], de décider de la présence ou de l'absence d'un animal dans la scène (TP<150ms) [STM96] ou encore d'identifier la catégorie de la scène (par exemple les catégories *plage*, *ville*, *forêt*, *intérieur*) (TP<135ms)[SO94]. Pour résumé, en une seule fixation oculaire sur la scène et en à peu près moins de 150ms de temps de présentation, les sujets sont capables de déterminer avec une bonne précision *l'essentiel* de la scène c'est-à-dire

FIG. 2.1 – Exemples de scènes naturelles ; de gauche à droite : scène de ville, de plage, de montagne et d'intérieur.

sa catégorie conceptuelle et son organisation spatiale générale. Cette identification rapide de la scène peut être ainsi particulièrement utile au système visuel afin de créer un contexte de perception dans lequel les objets présents peuvent être à la fois localisés et identifiés lors de l'exploration de la scène [HH99] [Tor03].

Cela nous amène à nous poser la question suivante : quelles sont les types de représentation et d'informations utilisées dans l'identification des scènes permettant d'obtenir une telle efficacité ?

2.1.1 Représentation par le spectre d'amplitude global

Un modèle particulièrement étudié est le spectre d'amplitude de la scène. Il correspond au module de la transformée de Fourier de l'image entière et décrit la répartition de l'énergie selon les fréquences spatiales et les orientations présentes dans l'image (statistiques du second ordre) (figure 2.2). Il possède en outre la caractéristique d'être invariant à de faibles translations de l'image. En moyenne le spectre d'amplitude des scènes naturelles présente de fortes énergies en basses fréquences et de plus faible énergie vers les hautes fréquences avec une décroissance suivant approximativement une loi en $1/f^{\alpha}$ (avec $\alpha \approx 2$) [AR92]. Cependant cette propriété n'est vérifiée qu'en moyenne sur une base importante de scènes naturelles. La décroissance d'énergie des basses vers les hautes fréquences apparaît systématiquement pour toutes les scènes, mais chacune présente un coefficient de décroissance particulier, des pics d'énergie peuvent n'apparaître qu'à certaines fréquences et n'être distribués qu'à certaines orientations.

Un certain nombre de travaux ce sont intéressés aux propriétés du spectre d'amplitude pour identifier des catégories possibles des scènes naturelles ([GDO00], [OT01], [OTGDH99]). Ces modèles se basent sur une analyse du spectre d'amplitude en réalisant échantillonnage par filtres de Gabor pour extraire les énergies à différentes fréquences et différentes orientations. Une une analyse statistique est ensuite effectuée consistant à projeter la réponse de chaque filtre dans un espace multidimensionnel où un outil de classification non supervisée (carte de Kohonen, Analyse en Composantes Curvilignes) permet de regrouper les scènes en paquets (i.e les scènes partagent des caractéristiques de fréquence et d'orientation) et ainsi de faire émerger des *catégories* de scènes naturelles. Ces travaux montrent ainsi qu'il est possible d'associer à chaque catégorie de scènes un spectre prototypique (figure 2.3).

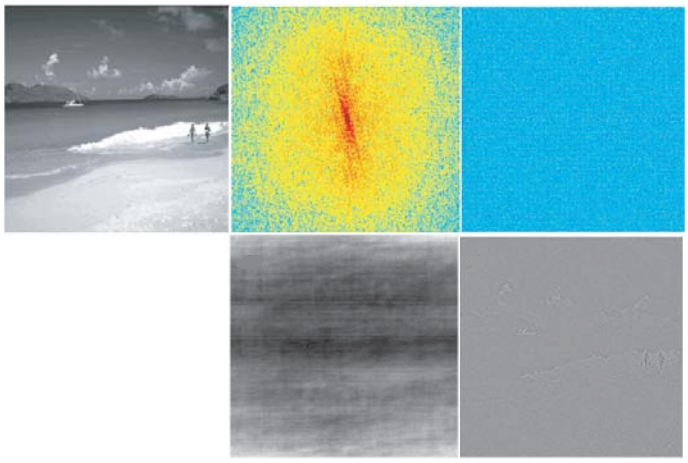


FIG. 2.2 – Analyse spectrale d'une scène naturelle (tiré de [Cha03]) ; de gauche à droite : scène de plage (image complète (amplitude+phase)) ; spectre d'amplitude de l'image (représenté dans le plan des fréquences (f_x,f_y)) ; spectre de phase de l'image (représenté dans le même plan) ; image reconstruite à partir du spectre d'amplitude original et d'un spectre de phase aléatoire ; image reconstruite à partir du spectre de phase original et d'un spectre d'amplitude aléatoire.

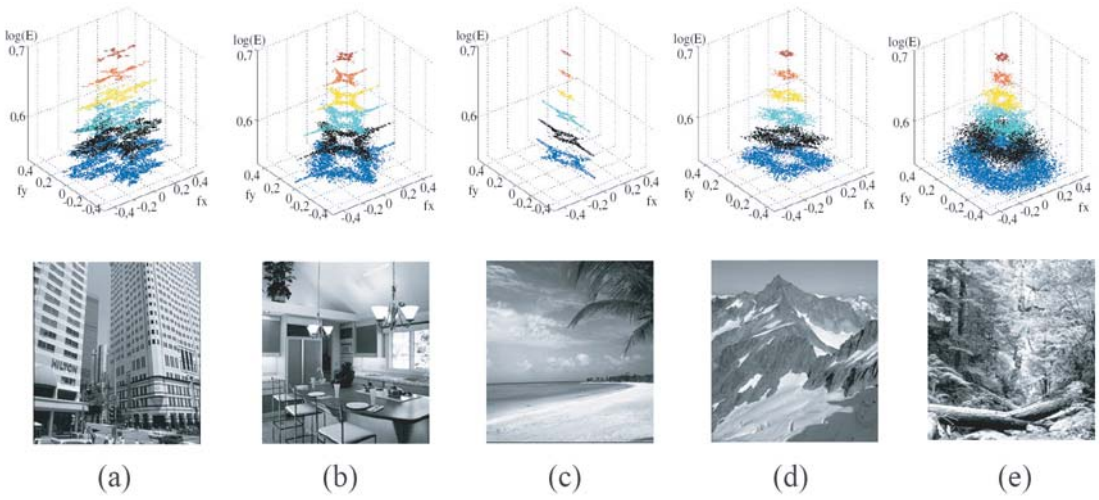


FIG. 2.3 – Exemples de spectres d'amplitude prototypiques de plusieurs catégories de scènes naturelles (tiré de [Leb04]) ; le spectre des scènes comportant des zones urbaines (a-b) est fortement marqué par la présence de fréquences horizontales et verticales ; au contraire, le spectre des scènes de paysages naturels tend à être le même selon toutes les directions (d,e), à l'exception des paysages comportant une ligne d'horizon bien marquée (c) favorisant les fréquences verticales.

2.1.2 Représentation par les spectres d'amplitude locaux

Cependant travailler uniquement sur le spectre d'amplitude global de l'image n'est pas suffisant pour rendre compte de toutes les propriétés de l'image analysée. Turiel et Parga

[TP00] montrent ainsi qu'une analyse locale associant une analyse fréquentielle (domaine de Fourier) et une analyse spatiale (position locale dans l'image) permet d'améliorer la précision de la description de la scène, notamment en découpant la scène en un ensemble de régions possédant des propriétés statistiques communes. Dans la transformée de Fourier, l'information de position spatiale est portée par le spectre de phase (figure 2.2). Différents travaux ont étudié l'importance relative de l'information portée par le spectre d'amplitude et de celle portée par le spectre de phase [GCP$^+$04]. Dans la figure 2.2 deuxième ligne, le spectre d'amplitude global de l'image a été associé à un spectre de phase aléatoire (par transformée de Fourier inverse) puis l'opération duale a été également effectuée en associant au spectre de phase de l'image, un spectre d'amplitude aléatoire. Il apparaît clairement la dominance de l'information portée par le spectre de phase. Il est également possible de construire des *chimères* en intervertissant les spectres d'amplitude et de phase de deux images (figure 2.4 première ligne). L'information portée par le spectre de phase apparaît également dominante. Cependant Morgan *et al* [MMH91] ont montré que cette dominance s'inverse si l'échange des spectres d'amplitude et de phase s'effectue par morceaux (figure 2.4). Si l'image est subdivisée en différentes parties dans lesquelles sont crées des *chimères* locales, il est possible d'observer qu'avec la réduction de l'échelle, l'information portée par le spectre d'amplitude domine celle portée par le spectre de phase. Les auteurs ont ainsi mis en évidence que l'importance relative de la phase et de l'amplitude s'inverse en fonction de la taille des images, et que la taille des régions pour laquelle l'inversion se produit pourrait être liée aux tailles des champs récepteurs des cellules visuelles du cortex visuel primaire (voir section 3.4).

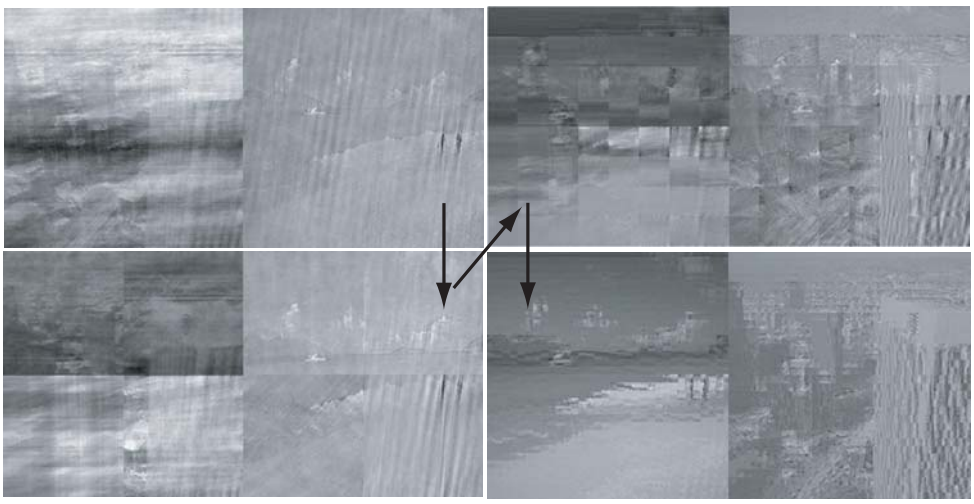


FIG. 2.4 – Reproduction de l'expérience de Morgan *et al* (tirée de [Cha03]) ; première ligne, à gauche : *chimères* réalisées en associant le spectre d'amplitude d'une image de plage et le spectre de phase d'une image de ville ; à droite *chimères* duale ; en suivant les flèches, subdivision successives des images et création de *chimères* par morceaux ; la taille des régions diminue entre les chimères des images globales (première ligne à gauche) et les chimères réalisée à la taille du pixel (dernière ligne à droite).

Shams et von der Malsburg [SvdM02] en s'inspirant des travaux de Morgan *et al* construisent un modèle simple utilisant des filtres localisés dans le domaine spatial et dans l'espace de Fourier. Ces filtres intègrent l'énergie locale du spectre d'amplitude et sont insensibles à l'in-

formation de phase. Les auteurs montrent, grâce à ce modèle simulant les cellules complexes de l'aire visuelle primaire (voir section 3.4), que l'information contenue dans les spectres d'amplitude correspondant à des régions locales de l'image couplée à l'information sur leur position spatiale relative est suffisante pour reconnaître une image.

Oliva et Torralba proposent également une méthode de reconnaissance des scènes ne faisant pas non plus appel à une étape préalable de segmentation des objets ou en régions [OT01]. Ils créent d'abord une représentation simplifiée de la scène, appelée *l'enveloppe spatiale* basée sur la projection de l'image sur une base de filtres permettant d'obtenir une réduction de dimension. Chaque filtre est associé à une dimension perceptive qui représente la structure spatiale dominante de la scène : naturelle (*naturalness*, par opposition à des structures urbaines), ouverture (*openness*), complexité (*roughness*), expansion (*expansion*, perspective), horizontalité (*ruggedness*)). A cette analyse globale est associée une analyse locale dans un ensemble de sous-régions décomposant la scène suivant un quadrillage régulier. La modèle de la scène ainsi obtenu se représente dans un espace multidimensionnel où les scènes se regroupent suivant leur catégorie sémantique (par exemple les rues, les plages). Cette représentation *holistique* de la scène apparaît ainsi suffisante pour retrouver la catégorie de la scène. Ces même auteurs ont également développé un modèle très similaire combinant analyse globale et locale permettant d'estimer la profondeur moyenne d'une scène [TO02b].

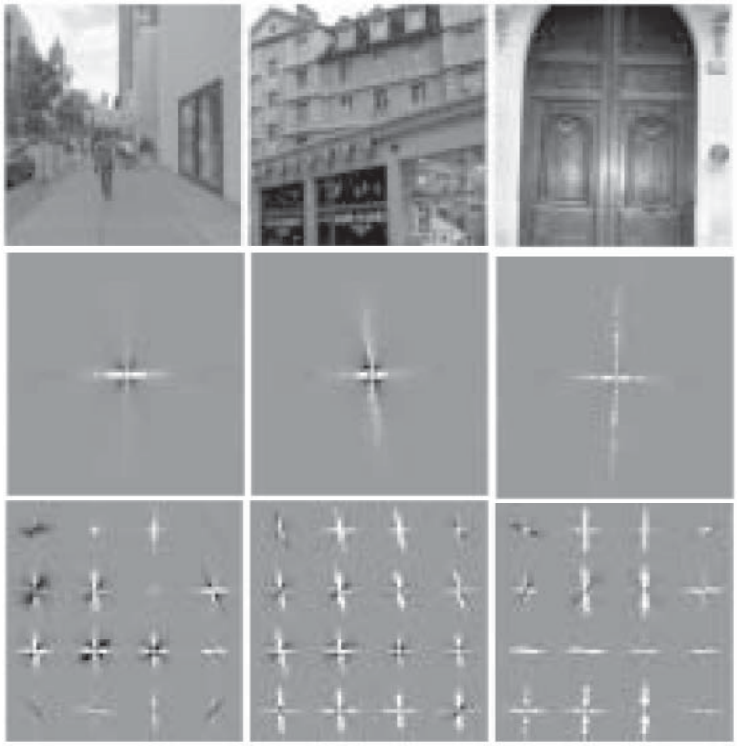


FIG. 2.5 – Exemple d'analyse globale et locale effectuée par la méthode de Oliva et Torralba (tiré de[OT01]) ; première ligne : images de scènes naturelles organisées suivant l'axe d'expansion (de gauche à droite : perspective forte, moyenne et nulle) ; deuxième ligne : spectre d'amplitude global de la scène projeté sur le filtre correspondant à l'axe d'expansion ; troisième ligne : spectres d'amplitudes de chaque sous-régions de la scène projetés sur le filtre correspondant à l'axe d'expansion des sous-régions correspondante.)

2.1.3 Représentation par la texture

Renninger et Malik [RM04] proposent qu'un modèle simple basé sur l'analyse des propriétés de la texture peut permettre l'identification rapide d'une scène. A l'aide d'une base de filtres (dérivées de gaussienne), la texture est analysée selon différentes orientations et fréquences spatiales. Ses propriétés sont ensuite projetées dans un espace de multi-dimensionnel où un apprentissage non-supervisé (KNN) permet d'obtenir une base de dimensions inférieures dont les axes sont appelés *textons généralisés*. Chaque scène est ensuite représentée par l'histogramme des textons généralisés qui la composent. Une expérience psychophysique de discrimination de la catégorie de la scène (temps de présentation, TP<60ms) permet de comparer les performances du modèle avec les performances chez les sujets humains. Les résultats obtenus sont significativement corrélés pour des TPs faibles. Les auteurs en concluent que l'identification rapide des scènes naturelles peut être expliquée par un modèle simple d'analyse des propriétés de la texture.

La texture peut être considérée comme un indice holistique, c'est-à-dire un indice extrait sur l'ensemble du champ visuel de manière extrêmement rapide sans avoir recours à des processus attentionnels pour analyser ses propriétés [BJ83] [BA88]. Les travaux de Renninger et Malik [RM04] montrent ainsi que l'indice de texture est un candidat possible de la représentation des scènes naturelles (ou du moins pour l'extraction de l'essentiel (*gist*) de la scène en vision pré-attentive). Dans leur travaux, Oliva et Torralba [OT01] n'explicitent pas le lien de leur modèle avec l'analyse de la texture. Cependant les informations véhiculées par les filtres composant leur modèle d'*enveloppe spatiale* de la scène sont également des combinaisons linéaires des caractéristiques de la texture (i.e des combinaisons de différentes fréquences spatiales et orientations). Ils ont pu développer un modèle à la fois pour catégoriser la scène mais également pour identifier son organisation spatiale (notamment suivant un axe perceptif représentant la perspective de la scène) et sa profondeur moyenne. En d'autres termes l'analyse locale des propriétés de la texture semblent contenir suffisamment d'information pour analyser les caractéristiques 3D des scènes naturelles.

2.1.4 Résumé

L'analyse des scènes naturelles est un problème complexe. Une approche consiste à étudier les caractéristiques du spectre d'amplitude global de la scène. Celui-ci permet d'obtenir de première informations sur la catégorie de la scène. Cependant ce modèle n'est pas suffisant, notamment pour représenter les caractéristiques locales de la scène. Pour cela une analyse des spectres d'amplitude régions localisées de la scène permet d'obtenir une représentation plus précise et notamment des informations sur les caractéristiques 3D de la scène (organisation spatiale, profondeur, perspective). Enfin l'analyse des caractéristiques locales de la texture peut servir de base au développement d'un modèle d'extraction des attributs 3D de la scène.

2.2 L'analyse de la texture

Cette section décrit succinctement les principales caractéristiques du domaine de l'analyse de texture.

2.2.1 Définition et propriétés

L'analyse de texture est un domaine général en vision par ordinateur qui a déjà fait l'objet de 30 années d'études. Cependant la texture reste un objet difficile à définir de manière précise et générale. Le nombre de définitions possibles données par différents auteurs en témoignent. Nous retiendrons celles données par Tamura *et al* [TMY78], Sklansky *et al* [Skl78] et Haralick [Har79], respectivement :

- *Nous pouvons considérer une texture comme ce qui constitue une région macroscopique. Sa structure correspond simplement à la forme répétitive dont les éléments ou primitives sont arrangés selon une règle de placement.*

- *Une région dans une image possède une texture constante si un ensemble de statistiques locales ou d'autres propriétés locales de l'image sont constantes, en variant lentement, ou approximativement périodique.*

- *Une texture d'image est décrite par le nombre et le type de ses primitives et leur organisation spatiale... Une caractéristique fondamentale de la texture : elle ne peut être analysée sans une référence correspondant à une primitive. Pour toutes surfaces , il existe une échelle (spatiale) telle que quand la surface est examinée, la texture n'existe pas. Ainsi lorsque la résolution augmente, elle fait passer d'une texture fine à une texture grossière.*

La texture correspond au *dessin* supporté par une surface (plane ou courbe). Les éléments structurants la texture possèdent des caractéristiques associées à des propriétés de répartition spatiale (périodicité, aléatoire, irrégularité, variations lentes) communes sur l'ensemble de la région considérée. Nous parlerons alors de texture homogène ou de région homogène. Enfin les caractéristiques (ou les statistiques) de la texture dépendent de la taille de la région spatiale utilisée pour l'analyser (échelle d'observation) : une taille petite peut être inférieure à la taille d'un élément de la texture ; une taille trop grande peut uniformiser les propriétés, les rendant inexploitables.

Deux classes de texture sont généralement distinguées (Figure 2.6) : les *macrotextures* présentant des relations spatiales relativement régulières entre les éléments (par exemple les briques d'un mur, les fleurs d'un champ de tournesol) ; les *microtextures* présentant des relations spatiales relativement faibles entre les éléments, souvent distribués de manière aléatoire (par exemple le grain du bois, les sillons sur une plage de sable)

Différents critères visuels peuvent être utilisés pour décrire qualitativement la texture tels que : le contraste, la granularité, l'orientation, la forme, la finesse, la régularité et la rugosité. Cependant l'objectif de l'analyse de texture est de trouver une manière simple et unique de représenter ses caractéristiques afin de pouvoir y appliquer des traitements mathématiques (par exemple pour la classification de texture). Beaucoup de techniques ont été développées pour décrire la texture. Tuceryan et Jain divisent les méthodes d'analyse de la texture selon quatre approches (voir [TJ98] pour une revue détaillée) : statistique (étude des moyennes et

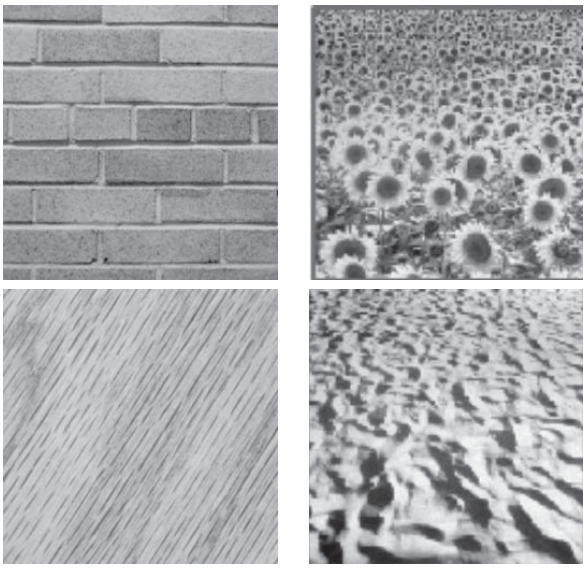


FIG. 2.6 – Exemples de textures macroscopiques (première ligne) et microscopiques (deuxième ligne).

variance d'outils tels que les matrices de co-occurences, les histogrammes de bords, la fonction d'autocorrélation, bien adaptée pour les textures microscopiques [Cru97]), géométrique (étude individuelle des éléments d'une texture macroscopique, théorie des *textons* de Julesz [Jul81]), basée sur des modèles (reconnaissance des formes, également pour des textures macroscopiques) et spatio-fréquentielle (caractérisation multi-échelle à base de transformée en ondelettes ou de filtrage (par exemple par filtres de Gabor), permettant le traitement de textures microscopiques et macroscopiques).

Différents aspects doivent être considérés au moment du développement d'algorithme d'analyse de la texture [OPM02] en fonction du problème abordé : l'invariance aux conditions d'illumination (textures en niveaux de gris) ; l'invariance au zoom (i.e à l'échelle spatiale ou échelle d'observation) et à la rotation (par exemple pour la classification de texture) ; l'invariance au paramètres de projection en 3D (dans le textures supportées par des surfaces planes inclinées dans l'espace ou de surfaces courbes vues en projection sur le plan de l'image) ; la taille de la fenêtre d'analyse (déterminer la largeur de la région locale analysée afin de pouvoir produire une description utile du contenu de la texture) ; l'invariance au bruit ; l'invariance aux paramètres propres de l'algorithme (les valeurs de ces paramètres ne doivent pas être critiques afin de pouvoir analyser le plus grand nombre possible de textures) ; la complexité calculatoire (relativement faible afin de pouvoir traiter des bases de texture de taille importante) ; la généralité de la méthode employée (afin de pouvoir adapter la méthode à différentes applications).

L'analyse de texture s'applique à quatre grands domaines en vision par ordinateur : la classification de texture (assignée une nouvelle texture à une classe de texture connue) ; la segmentation d'image (découpée l'image ou la scène en régions homogènes à l'aide d'algorithmes tels que *normalized cut* [SM00]) ; la synthèse de texture (génération artificielle de texture) ; l'extraction de la forme par la texture (retrouver la forme et l'orientation 3D de la surface

supportant la texture en étudiant les déformations de celle-ci lors de sa projection du monde 3D sur le plan de l'image 2D).

Dans notre étude nous distinguons les propriétés statistiques suivantes pour décrire les textures utilisées dans notre base de texture et notre base de scènes naturelles (voir Chapitre 6) (Figure 2.7) : l'homogénéité (voir définition plus haut, les textures ainsi que les régions extraites de scènes naturelles seront considérées homogènes et obtenues après une étape (supposée) de segmentation en régions) ; la régularité (par exemple une texture périodique, de manière générale les propriétés statistiques restent constantes par translation (stationnarité) à l'échelle d'observation utilisée) ; naturelle (par exemple un champ, texture obtenue à partir d'un environnement naturel) ; artificielle (par exemple une façade d'immeuble, texture obtenue à partir d'un environnement urbain ou crée par l'Homme) ; isotrope (par exemple : l'écorce d'un arbre,la texture ne présente d'orientation préférentielle) ; anisotrope (par exemple les sillons d'un champ, la texture possède une seule orientation préférentielle, les autres orientations n'étant pas présentes). Toutes ces caractéristiques ne sont pas exclusives et une texture peut présenter plusieurs d'entre elles (Figure 2.7) (par exemple la façade d'un immeuble est une texture homogène, artificielle, isotrope (2 orientations orthogonales) et régulière (mais avec des variations à une échelle d'observation petite).

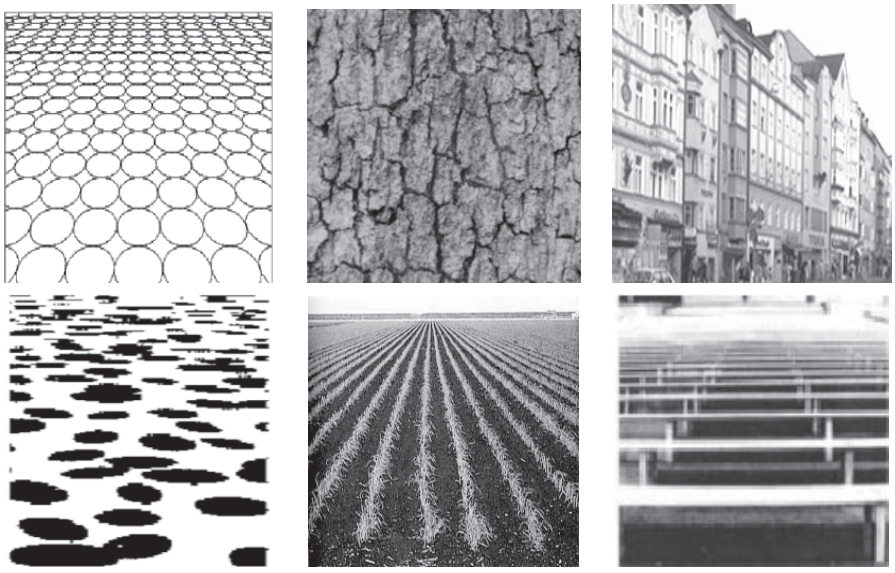


FIG. 2.7 – Exemples de textures homogènes ;de haut en bas, à gauche : texture régulière (stationnaire) ; texture irrégulière (non-stationnaire) ; au milieu : texture naturelle isotrope ; texture naturelle anisotrope ; à droite : texture artificielle isotrope ; texture artificielle anisotrope.

2.2.2 Résumé

Pour analyser les scènes naturelles, nous supposons qu'une étape préalable de segmentation en régions homogènes est effectuée. Il existe une grande variété de textures possible possédant différentes caractéristiques statistiques. Dans notre étude nous distinguons différentes caractéristiques. Les textures peuvent être : homogènes, régulières, irrégulières, naturelles,

artificielles, isotropes et anisotropes. Nous adopterons une approche spatio-fréquentielle pour l'étude des textures. dans le contexte du travail présenté nous nous intéressons plus à un type d'application de l'analyse de texture : l'extraction de la forme par la texture permettant de retrouver les paramètres 3D d'une surface texturée.

2.3 Codage 3D et projection perspective

Cette section présente le mode de projection utilisé permettant de coder les paramètres 3D d'une surface. Différents modèles de projection : projection parallèle (orthographique) [SB95b], paraperspective [Alo88], projection perspective [Gar92]. Nous considérons ici le modèle plus complet, la projection perspective, rendant compte de toutes les déformation produites, notamment, lors la formation de l'image sur la rétine de l'oeil.

2.3.1 Projection

Nous nous plaçons dans le cas d'une projection perspective (figure 6.8). Il s'agit d'obtenir la relation entre un point de la surface et un point de l'image en fonction des angles de slant et de tilt. L'angle de slant, appelé roulis et noté σ, correspond à l'inclinaison de la surface c'est-à-dire à l'angle formé par l'axe de vue et la normale à la surface. L'angle de tilt, appelé tangage et noté τ, correspond à la direction de l'inclinaison c'est-à-dire à l'angle de rotation appliqué à la surface après l'avoir inclinée. Dans la suite du document les anglicismes "perception du slant" pour la perception de l'inclinaison et "perception du tilt" pour la perception de la direction en profondeur seront employés pour des raisons de clarté induite par leur relation directe au codage en tilt/slant couramment utilisés dans la littérature en neurophysiologie, en perception visuelle et en extraction de la forme par la texture.

La projection perspective d'un objet peut se décomposer en une rotation de l'image (τ) et une compression dans cette direction (σ). Enfin une homothétie finale permet de rendre compte de la distance de la surface 3D à l'observateur et de la distance entre le plan 2D et la surface 3D. Une rotation initiale peut être également rajoutée permettant de contrôler l'orientation de la texture sur la surface (angle de *roll* non reporté ici).

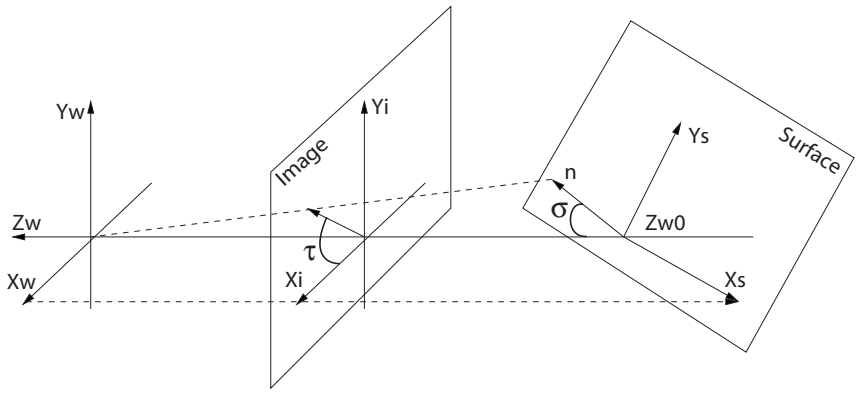


FIG. 2.8 – Modèle de projection perspective

La relation entre les coordonnées (x_s,y_s) de la surface vue et les coordonnées (x_i,y_i) de l'image projetée s'exprime par (voir également [SB95a]) :

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{\begin{bmatrix} cos(\sigma) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} cos(\tau) & sin(\tau) \\ -sin(\tau) & cos(\tau) \end{bmatrix}}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{A}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{2.1}$$

avec $a_i = \dfrac{-sin(\sigma)sin(\tau)x_i + cos(\tau)sin(\sigma)y_i + dcos(\sigma)}{d + dzw0}$ correspondant à un facteur de zoom en fonction de la position spatiale (x_i, y_i).

Les différents systèmes de coordonnées sont : le monde (x_w, y_w, z_w), l'image (x_i, y_i) et la surface plane (x_s, y_s). L'image est à une distance nodale f.

La projection perspective introduit trois types de distortion sur les surfaces :

- un effet dû à la distance (réduction de la dimension des éléments de la texture) (figure 2.9).

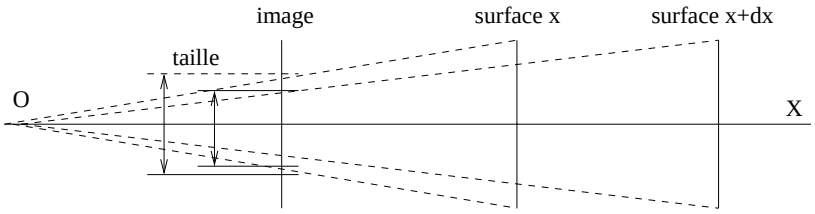


FIG. 2.9 – Effet de distance.

- un effet de position : compression due à l'angle entre la ligne de vue et la position du point sur la surface qui se traduit par un effet de torsion de la texture (torsion géodésique)) (figure 2.10).

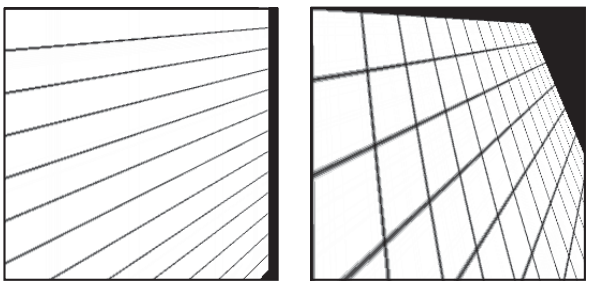


FIG. 2.10 – Effet de torsion géodésique ; de gauche à droite : texture orientées à $\tau = 0^o$ et $\sigma = 13^o$; texture orientées à $\tau = 60^o$, $\sigma = 23^o$; le centre de projection est situé au coin supérieur gauche.

- un effet de compression : compression qui dépend de l'angle entre la ligne de vue et la normale au plan (figure 3.3).

Un autre type de projection, la projection orthographique, est aussi beaucoup utilisée car elle permet des calculs plus simples. Cependant elle ne capture que l'effet de compression

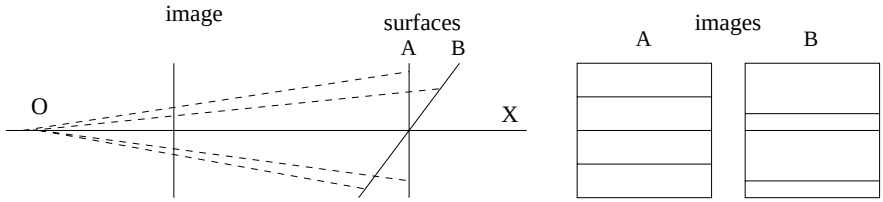


FIG. 2.11 – Effet dû à la compression en projection perspective.

(figure 2.19). De plus aucun effet de torsion de la texture n'apparaît excluant ainsi les effets de perspective linéaire sur des surfaces planes. Pour toutes ces raisons la projection perspective a été utilisée.

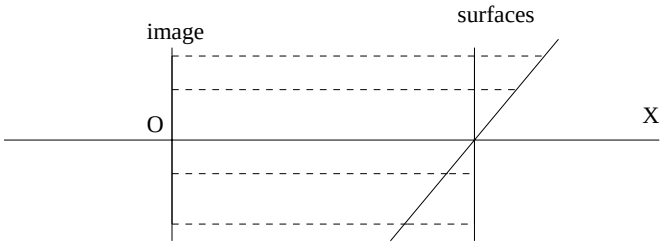


FIG. 2.12 – Pas d'effet de distance et de position en projection orthographique.

2.3.2 Résumé

Dans ce travail nous considérons les textures ayant subies une projection perspective lors du passage du monde 3D au plan de l'image. Les paramètres 3D d'une surface planes peuvent alors se réduire à deux angles : le tilt (direction en profondeur) et le slant (inclinaison dans la direction du tilt). L'objectif d'un modèle d'extraction de la forme par la texture pourra donc être de mesurer le plus précisément possible ces deux angles pour pouvoir extraire les informations 3D de l'image.

2.4 Modèles d'extraction de la forme par la texture

L'extraction de la forme à partir de la texture (ou *Shape from Texture* en anglais) consiste à obtenir les informations tridimensionnelles d'une image texturée par l'analyse des modifications subies par celle-ci lors de sa projection en deux dimensions.

Cette technique repose sur l'hypothèse suivante : les effets de profondeur ou de perspective de l'image ne sont dûs qu'au passage du monde 3D au plan de l'image 2D. Les déformations de la texture ne sont les reflets que de cette projection. La texture est donc supposée n'introduire aucun effet supplémentaire pouvant induire en erreur l'estimation des paramètres de la surface de départ. Si tel est le cas, cette information est alors interprétée comme une déformation due à la projection (ce qui correspond à l'effet *trompe l'oeil*).

Le développement d'une méthode d'extraction de la forme par la texture se divise en trois parties :

1. IDENTIFICATION : il s'agit d'identifier la caractéristique de la texture porteuse de l'information 3D (par exemple : les éléments de la texture, la fréquence moyenne locale).
2. ESTIMATION : il s'agit de mesurer la caractéristique retenue en un point ou en une zone de l'image (par exemple : mesurer la changement de densité des constituants de la texture, estimer le taux de compression ou mesurer la fréquence locale).
3. INTERPRÉTATION : il s'agit de traduire la quantité précédemment mesurée en terme de déformation de la texture en fonction des valeurs des paramètres 3D de la surface de départ (les angles codant l'inclinaison et l'orientation de la surface).

Nous distinguons deux grandes catégories d'algorithmes : celles basées sur l'analyse des éléments présents dans la texture (analyse spatiale) et celles qui effectuent une analyse dans le domaine de Fourier (analyse fréquentielle).

2.4.1 Analyse spatiale

L'un des premiers à s'intéresser à la perception de l'orientation des surfaces fut Gibson ([Gib50a])dans les années 50. Il introduisit le terme de *texels* pour désigner les éléments constituants la texture (par exemple : les briques d'un mur ou les graviers sur le sol). En faisant une hypothèse de distribution uniforme de ces *texels* sur la surface, il interprète toute modification de cette distribution dans l'image 2D comme provenant de la projection de la surface sur l'image en fonction de son orientation. Ces changements se traduisent par des gradients dans l'image (par exemple : des gradients de densité, de taille (les *texels* proches de l'observateur sont plus grands que ceux qui sont éloignés)), d'où le terme de *gradients de textures* (voir Chapitre 4 pour une explication détaillée).

Cette approche fut suivie par d'autres auteurs dont Aloimonos [Alo88] qui utilisa la somme des contours des *texels* comme gradient ou Stevens ([Ste84]) qui a repris les gradients les plus importants au sein d'un même formalisme pour l'étude des surfaces planes. Gårding ([Gar92]) a effectué un travail similaire pour les surfaces courbes. De nombreux travaux ont ensuite consisté à effectuer un traitement statistique de ces gradients.

Cependant l'approche gibsonienne présente de nombreux désavantages. Elle repose d'abord sur l'existence de *texels* dans la texture. Elle suppose également qu'il est possible de les identifier, de les évaluer et qu'ils sont répartis uniformément sur la texture. Or il existe de nombreuses textures qui ne contiennent pas de *texels* ou qui ne sont pas facilement identifiables (par exemple : si les briques du mur sont trop grandes, l'écorce d'un arbre). Cela oblige également à se limiter à l'information contenue dans ces éléments. En effet une fois extraits les textittexels, les reste de la texture et donc d'autres sources potentielles d'information sont écartées. Enfin Gårding a montré que le choix du type de gradient était important car celui-ci ne permet de considérer qu'une seule caractéristique de la déformation et qu'il était peut-être ainsi nécessaire d'en intégrer plusieurs afin d'obtenir une bonne estimation.

2.4.2 Analyse fréquentielle

Une autre approche a été initiée par Bajcsy et Liebermann ([eLL76]) en 1976 qui repose sur l'étude du spectre de la texture. En effet la variation en fréquence dans le spectre peut être interprétée comme la distortion d'une texture homogène (fig 2.13). Au premier ordre, une variation fréquentielle n'est induite que par une modification de la géométrie de la texture. Les travaux que nous présentons ensuite font tous une estimation dans le domaine fréquentiel.

FIG. 2.13 – Texture d'une surface courbe et traduction de la variation de celle-ci dans le domaine fréquentiel.

Depuis le début des années 90, l'utilisation de l'analyse spectrale a conduit à de nombreux algorithmes efficaces. Dans les méthodes développées par Super et Bovik [SB95b], Malik et Rosenholtz [MR97], Ribeiro et Hancock [RH01], Sakai et Finkel [SF95], Guerin and Elghadi [GDE99], Loh et Kovesi [LK05], la déformation de la texture est mesurée à l'aide de la distortion affine d'éléments du spectre (par exemple des pics d'énergie, les moments locaux, l'inertie). Ces méthodes sont précises mais nécessite la présence d'au moins deux composantes d'énergie distincts dans le spectre et ainsi souvent elles obtiennent leurs meilleurs résultats en présence d'au moins deux orientations dans des textures régulières ou faiblement irrégulières.

Des techniques alternatives ne font pas ce type d'hypothèse sur les composantes du spectre permettant ainsi de traiter des textures présentant ou non des orientations. Les méthodes développées par Super et Bovik [SB95a], Lindeberg et Gårding [GL96], Hwang *et al* [HLC98], Lelandais *et al* [LBP05], sont basées sur l'estimation locale de la fréquence (ou de l'échelle spatiale (l'inverse)) en utilisant différents types de filtres (par exemple des filtres de Gabor, des dérivée de gaussienne, des transformations en ondelettes). Elles sont basées sur le choix de la meilleur fréquence locale, en supposant explicitement qu'il n'y ait qu'une seule fréquence à la position locale considérée. Ceci n'est pas toujours vérifié dans les cas particuliers de textures multiples, d'occlusion ou de textures en transparence [BR95]. Afin de prendre en compte plus ou moins l'irrégularité de la texture, toutes ces méthodes requièrent une méthode intensive telle qu'une approximation parabolique ou une technique de minimisation de variance pour obtenir finalement les paramètres géométriques des surfaces (souvent limitées aux surfaces planes).

Dans la méthode développée par Clerc et Mallat [CM02], la migration locale des coefficients d'ondelettes est reliée aux paramètres de forme locale. Afin d'obtenir des résultats sur des textures irrégulières, le problème est traité comme un processus de stationnarisation ce qui permet d'imposer des contraintes fortes sur les variations des coefficients. Cependant due à une hypothèse d'ergodicité, chaque ondelette est prise autour d'une fréquence centrale relativement haute et donc ne tire pas explicitement avantage de l'ensemble des fréquences disponibles.

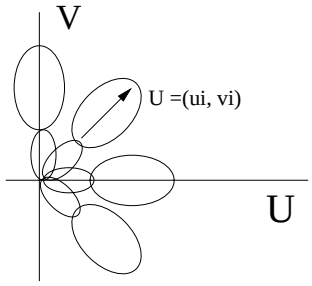
Enfin aucune de ces techniques ne fait un lien avec le fonctionnement du système visuel biologique, exceptée la méthode développée par Sakai et Finkel [SF95]. Cependant elle ne

fait pas intervenir explicitement de modèle des cellules corticales (par exemple un modèle des cellules complexes) qui est cependant un élément fondamental du fonctionnement du système visuel primaire (voir Chapitre 3).

2.4.3 Super & Bovik

La technique de Super et Bovik [SB95b] permet d'obtenir le tilt et le slant d'une surface courbe avec une bonne précision en supposant uniquement l'homogénéité de la surface. Leur approche cherche à caractériser la compression de la texture uniquement due à la courbure. Pour cela ils caractérisent les fréquences spatiales locales en un point particulier à l'aide des moments spectraux du second ordre (fig 2.14 (a)).

$$a(x) = \sum_i u_i^2(x) A_i^2(x)$$
$$b(x) = 2 \sum_i u_i(x) v_i(x) A_i^2(x)$$
$$c(x) = \sum_i v_i^2(x) A_i^2(x)$$



(a) (b)

FIG. 2.14 – La figure (a) donne l'expression des trois moments du second ordre. La figure (b) représente dans le domaine fréquentiel une rosace de filtres de Gabor.

L'estimation du moment maximum, du moment minimum et de l'orientation principale (orientation du plus petit moment du second ordre) permet à la fois de caractériser totalement la déformation locale et de s'affranchir de l'orientation du système de coordonnées au point étudié. Cette dernière propriété est importante car pour que l'hypothèse fondamentale du *Shape From Texture* soit validée, il faut que l'orientation des fréquences spatiales mesurées soit constante vis-à-vis du système de coordonnées, sinon cela induit une variation due à la rotation de la texture et non due à la forme de la surface.

Pour mesurer les moments locaux, Super et Bovik utilisent une batterie de filtres de Gabor disposés en rosace sur le spectre local (fig 2.14 (b)). Chaque filtre i permet d'obtenir l'amplitude du spectre d'énergie A_i suivant une orientation, définie par le vecteur u_i correspondant au vecteur central U_i du filtre de Gabor.

Ces moments sont ensuite normalisés par la somme des énergies au point considéré ($\sum_i A_i^2(x)$) afin de pouvoir les mettre en relation avec les moments d'autres points. Le passage aux moments canoniques est effectué : les auteurs utilisent trois expressions permettant de passer des moments estimer $a_i(x)$, $b_i(x)$ et $c_i(x)$ au moment maximum M, au moment minimum m et à l'orientation principale θ.

Il reste maintenant à interpréter ce résultat pour obtenir les deux angles d'orientation de la surface. Pour cela les auteurs utilise la projection inverse de la projection orthographique afin d'obtenir la relation entre les moments normalisés de l'image et ceux de la surface initiale.

$$a_s(x_s) = M(\sigma, \tau).a_i(x) \tag{2.2}$$

où $M(\sigma, \tau)$ est la matrice de passage des moments de l'image aux moments de la surface ; elle dépend des paramètres de la projection.

L'effet recherché est le taux de compression de la texture et la surface est supposée courbe. Les effets de position ou de distance ne sont donc pas utiles et désirés. Aussi utiliser la projection orthographique est suffisant et permet de simplifier les calculs sans influer sur les résultats. La résolution de l'équation précédente permet d'obtenir ainsi les expressions du slant et du tilt :

$$cos\sigma = \sqrt{\frac{M_s m_s}{Mm}}$$

$$\tau = \{\theta \pm \tfrac{1}{2} arccos\lambda, \theta \pm \tfrac{1}{2} arccos\lambda + \pi\}$$
$$\text{avec } \lambda = \frac{(cos^2\sigma+1)(M+m)-2(M_s+m_s)}{sin^2\sigma(M-m)}$$
$$sin\sigma \neq 0,\ M \neq m.$$

FIG. 2.15 – Expressions permettant d'obtenir le slant (σ) à gauche et le tilt (τ), à droite.

On remarquera l'indétermination sur le tilt due au signe $\pm$ et à l'ambiguïté sur sa direction (à $\pm\pi$). Une hypothèse sur la texture doit alors être faite pour pouvoir obtenir des expressions indépendantes des moments de la surface de départ non mesurables. Les auteurs utilisent alors la condition peu restrictive d'homogénéité qui permet de considérer les moments M_s et m_s de la surface constants. Le rapport entre les angles de slant de deux points distincts s'exprime alors :

$$\frac{cos\sigma_1}{cos\sigma_2} = \sqrt{\frac{M_1 m_1}{M_2 m_2}} \tag{2.3}$$

où σ_i représente le slant au point i.

Il reste enfin à obtenir une estimation du slant en un point afin de pouvoir ensuite le déduire pour tous les points de l'image. Un point frontal est un point dont la projection de la normale à la surface locale en ce point est colinéaire à la ligne de vue, c'est-à-dire dont le slant est nul. Pour des textures courbes et proches de l'observateur, un tel point existe toujours et correspond au point de minimum $\sqrt{Mm}$.

L'algorithme de Super et Bovik permet donc d'obtenir les angles d'orientation d'une surface directement par estimation de la déformation de sa texture. C'est l'un des premiers à être à la fois satisfaisant, efficace et pouvant traiter un très grand nombre de textures. Cependant la précision de l'estimation dépend de la précision à laquelle est déterminé un point frontal et ces moments. De plus ce point n'existe que si la surface est courbe. Enfin la méthode est sensible au bruit et à l'orientation de la texture sur la surface. En effet les moments prennent en compte toute l'information contenue dans la texture, l'estimation peut donc être biaisée. Cependant elle est insensible à une rotation interne de la texture puisque les opérateurs ont été choisis indépendants de l'orientation du système de coordonnées.

2.4.4 Sakai & Finkel

Les travaux de Sakai et Finkel [SF95] ont une approche plus cognitive. Ils se basent sur des tests psychophysiques de perception afin de trouver une caractérisation simple de la texture permettant d'obtenir l'information de forme et d'orientation des surfaces. Ces tests

font apparaître que les sujets semblent suivre soit la variation des pics de fréquences soit la variation de la fréquence moyenne de la texture. En effet les figures 2.16 (tirées de [SF95] représentent des textures générées à partir de bruit par les auteurs où ne varie que le pic de fréquence sans modification de la moyenne, ou que la moyenne sans variation de la fréquence du pic.

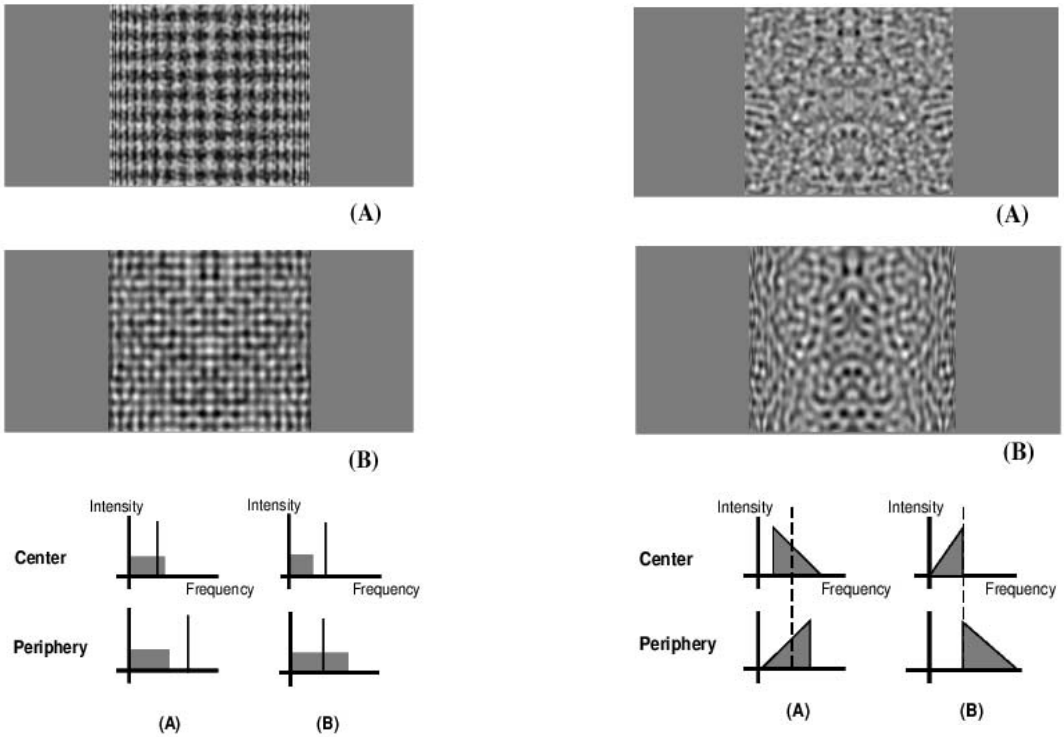


FIG. 2.16 – Textures crées avec soit une variation des pics de fréquence (à gauche) soit une variation de la fréquence moyenne (à droite).

Les auteurs remarquent ainsi qu'en présence de pics (barres verticales pleines), ce sont leur variation qui semblent donner l'information tridimensionnelle. Dans le cas contraire, c'est la variation de la fréquence moyenne (rectangles pleins) qui permettrait de distinguer une surface plane d'une surface courbe.

Afin de modéliser et de mesurer cette propriété, les auteurs introduisent un opérateur : *APF* (*Average Peak Frequency* ou pic moyen de fréquence). Il permet de caractériser la variation des pics s'ils existent ou la variation de la moyenne d'une texture au voisinage d'un point. L'expression formelle de l'APF est :

$$f_p(x_0, y_0) = \frac{1}{m} \sum_{I(x_0,y_0)} \sum_{f} prob(f_p(x, y) = f) * f \tag{2.4}$$

avec (x_0, y_0), les coordonnées du point ;
f_p, APF ;
$I(x_0, y_0)$, le voisinage du point ;

prob, distribution de probabilité que la fréquence soit un pic
(en pratique, c'est une gaussienne centrée sur la fréquence d'intensité maximale).

Les surface étudiées sont des surfaces courbes et la texture est supposée homogène. L'effet recherché est donc une variation de l'APF dans chacune des directions en un point donné. L'énergie et l'orientation des fréquences sont mesurées à partir d'un banc de filtres de Gabor. Comme pour la méthode de Super et Bovik, il est nécessaire d'obtenir ces mesures pour un point frontal qui sera pris comme référence (APF minimal dans toutes les directions). L'interprétation de la variation de l'APF est obtenue par sa normalisation (comparaison avec le point de référence) et l'expression du slant en projection orthographique :

$$\frac{(F^o(x,y)-F^o_{min})}{F^o_{min}}$$

avec $F^o(x,y)$, APF dans l'orientation o au point (x,y) ;
F^o_{min}, APF du point de référence (frontal) dans l'orientation o ;

$$tan(\sigma_o(x,y)) = \sqrt{(\frac{F^o(x,y)}{F^o_{min}})^2 - 1}$$

avec $\sigma_o(x,y)$, slant estimé dans l'orientation o au point (x,y) ;

FIG. 2.17 – Expressions de la normalisation à gauche et du calcul du slant à droite.

D'après la projection orthographique, l'APF varie maximalement dans la direction de la compression. L'orientation du tilt est donc obtenue en prenant l'orientation où la variation de l'APF normalisé est la plus grande.

Les auteurs ont également travaillé sur l'estimation de la profondeur de surfaces planes en projection perspective ([SF97]). La normalisation permet d'estimer la distance par rapport à l'observateur. Le point de référence n'est pas un point frontal ce qui empêche l'estimation du slant. L'étape de décision ne sert qu'à vérifier l'isotropie de la variation de l'APF suivant toutes les orientations due à la projection perspective.

Une analyse qualitative des résultats montre qu'ils sont cohérents avec la forme et la profondeur des surfaces utilisées. La méthode présente l'avantage d'être très peu sensible au bruit (travail sur les pics ou la moyenne). Par contre elle n'est valable que si la texture ne contient pas de rotation interne. Dans ce cas comme l'orientation change, l'évaluation de l'APF est complètement fausse car elle ne correspond plus à la même fréquence sur la texture. C'est notamment pour cette raison que les auteurs ont pris le soin de limiter l'effet de position de la projection perspective des surfaces planes en limitant l'amplitude du slant.

2.4.5 Gårding, Malik et Rosenholtz

Gårding ([Gar93] [Gar92]) s'est d'abord intéressé à réunir sous un même formalisme les gradients de texture les plus utilisés, de même que les moments du second ordre afin d'estimer la forme de surfaces planes ou courbes. Pour cela il introduisit une approximation de la projection perspective qui permet au niveau des calculs de se défaire de l'effet de position (fig 2.18).

La projection $F(p)$ est définie comme la projection inverse de la projection d'un point de la surface de départ sur une sphère unitaire (projection sphérique au lieu de plane). La

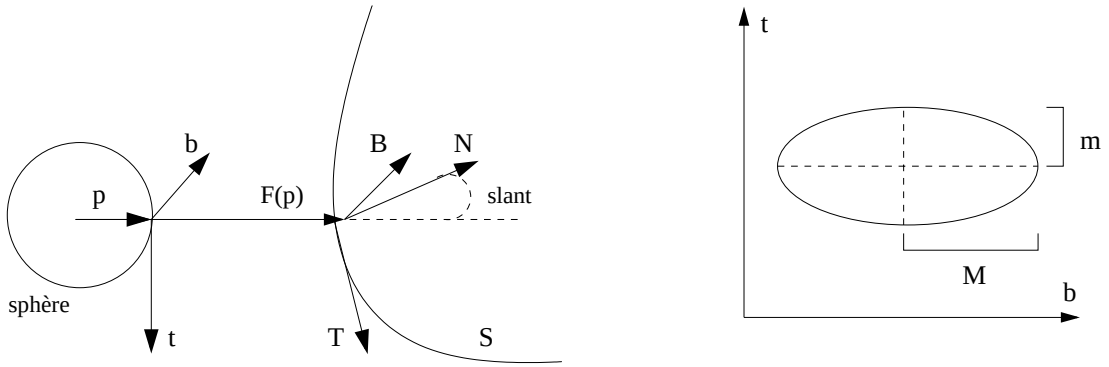


FIG. 2.18 – Projection de Gårding à gauche et image d'un cercle unitaire de la surface à droite.

matrice de projection du repère orthonormal (T, B, N) local à la surface (N, normale locale à la surface) au repère (t, b, p) (t colinéaire au gradient de distance de la sphère à la surface (direction du tilt), p colinéaire à la direction de la ligne de vue) s'exprime simplement de la manière suivante :

$$F(p) = \begin{pmatrix} \frac{r}{cos\sigma} & 0 \\ 0 & r \end{pmatrix} = \begin{pmatrix} 1/m & 0 \\ 0 & 1/M \end{pmatrix} \tag{2.5}$$

avec r, distance de la sphère à la surface ;
σ, le slant ;
m, compression suivant l'axe secondaire ;
M, compression suivant l'axe principal ;

Différentes combinaisons des gradients de M et de m permettent d'exprimer la compression de différentes caractéristiques d'un texel de la texture [Gar92] (par exemple ∇m permet d'obtenir le gradient de *compression*, ∇M, celui de perspective, ∇Mm, celui de *surface*).

L'approche de Malik et Rosenholtz [MR97] repose sur une analyse différentielle locale de l'image. Pour une texture homogène, il est possible de considérer deux imagettes (ou fenêtres) proches de l'image, comme identiques à une transformation géométrique près. Les auteurs proposent ainsi d'estimer une transformation affine entre ces imagettes en fonction des caractéristiques tridimensionnelles de la surface. L'idée est de voir la distortion de la texture comme un problème similaire à l'estimation d'un flot optique. Estimer une différence locale entre deux régions spatiales et estimer la variation locale temporelle d'une même région sont des problèmes qui peuvent être considérés comme équivalent et sur lesquels il est alors possible d'appliquer les mêmes techniques.

L'objectif est donc de retrouver les coefficients d'une matrice affine 2D de transformation entre deux imagettes d'une image. Ces coefficients dépendent du slant et du tilt de la surface considérée mais également du vecteur de déplacement. Les surfaces considérées sont des surfaces courbes et planes. L'estimation se fera sur les spectres de Fourier des imagettes.

L'interprétation va consister à mettre en relation la différence entre les deux spectres et la variation du spectre de la première imagette par rapport à la transformation affine. En faisant une approximation de Taylor au premier ordre, les auteurs obtiennent alors le système

d'équations linéaires suivant :

$$\begin{pmatrix} \frac{\partial F_1}{\partial f_x} f_x & \frac{\partial F_1}{\partial f_x} f_y & \frac{\partial F_1}{\partial f_y} f_x & \frac{\partial F_1}{\partial f_y} f_y \end{pmatrix} \begin{pmatrix} a_{1,1} \\ a_{1,2} \\ a_{2,1} \\ a_{2,2} \end{pmatrix} = F_2(\vec{f}) - F_1(\vec{f}) \tag{2.6}$$

avec F_i, spectre de l'imagette i ;
f_i, composante i de la fréquence ;
$a_{i,i}$, coefficients (i, i) de la matrice de l'affinité ;

Pour obtenir une expression formelle de la matrice, les auteurs utilisent le formalisme projectif introduit par Gårding. De plus il est nécessaire d'introduire une hypothèse supplémentaire à la condition d'homogénéité de la texture : l'*invariance par translation*. Cette supposition permet de considérer le changement de repère d'une imagette à l'autre de la surface comme une rotation.

Malik et Rosenholtz définissent la matrice affine comme étant la composée de la projection inverse d'un point p_1 de la sphère sur la surface au point P_1 ; une rotation d'angle δ_T pour passer au point P_2 ; de la projection sur la sphère au point p_2 ; et d'une rotation finale d'angle $\delta_t$pour se replacer dans le repère définit au point p_1 (fig 2.19).

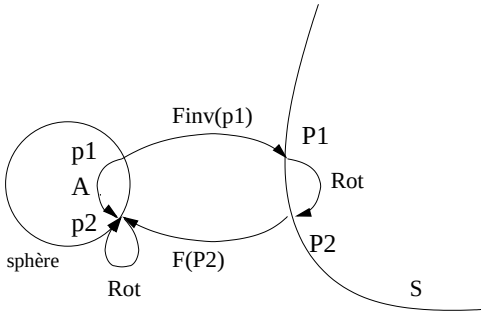


FIG. 2.19 – Etapes de l'estimation de la matrice affine.

En dérivant les calculs, les auteurs obtiennent une expression de la matrice affine où les coefficients dépendent du slant. Le tilt est obtenu en estimant l'angle entre le repère du point p_1 et un repère global (par exemple $(\vec{x},\vec{y})$).
Le passage du domaine spatial au domaine fréquentiel est obtenu en utilisant l'expression suivante :

$$F(g(Ax)) = \frac{1}{|A|} G(A^{-T} f) \tag{2.7}$$

avec F, la transformée de Fourier ;
g, la fonction spatiale (image) ;
$|A|$, le déterminant de la matrice de l'affinité.

Cette expression montre que si une matrice de transformation est estimée dans un domaine, elle le sera aussi dans l'autre.

Enfin les auteurs utilisent des algorithmes différentiels d'estimation du mouvement, notamment avec une évaluation de l'erreur de l'estimation. Ceci leur permet d'obtenir d'excellents résultats sur des textures artificielles et des textures d'environnements urbains. Cependant il faut remarquer qu'il est nécessaire d'avoir un spectre *riche* (par exemple pavage,briques) pour que l'estimation soit bonne.

2.4.6 Résumé

Parmis toutes les méthodes développées pour extraire la forme à partir de la texture, les méthodes qui estiment la fréquence locale en utilisant une décomposition en ondelettes ou des filtres spatio-fréquentiels sont celles qui imposant le moins d'hypothèses sur le contenu de la texture et sur ses composantes spectrales (par exemple les techniques de Clerc et Mallat, de Hwang *et al* et de Lelandais *et al*). De plus aucune méthode ne s'appuie sur un modèle biologiquement plausible de l'analyse de la texture, ce qui rend difficile le lien avec les études réalisées en psychophysique ainsi que les données recueillies en neurophysiologie.

2.5 Résumé

Nous avons introduit le problème de l'analyse de scènes naturelles. Nous avons présenté notre approche consistant à ne considérer que des régions locales recouvertes d'une texture homogène. La surface sous-jacente à cette texture peut être paramétrée en 3D par les angles d'inclinaison (le slant) et de direction en profondeur (le tilt) en projection perspective. L'extraction des informations 3D consiste donc à estimer ces deux angles à partir des déformations subies par la texture. Dans ce domaine de nombreuses techniques ont été développées. Les méthodes s'attachant à extraire la fréquence locale ne requièrent qu'une hypothèse d'homogénéité de la texture, sans contraintes sur les composantes du spectre de la texture. Enfin, excepté le modèle de Sakai et Finkel (qui n'intègre pas de modèle des cellules corticales), aucun modèle n'a été développé en s'inspirant des grandes étapes du fonctionnement du système visuel, rendant les liens avec la psychophysique et la neurophysiologie difficiles.

Afin de développer un modèle biologiquement plausible, nous devons tout d'abord connaître les grandes étapes du fonctionnement du système visuel. Ceci est l'objet du chapitre 3.

Chapitre 3

Neurophysiologie du système visuel

Ce chapitre décrit les caractéristiques principales des premières étapes du système visuel : la rétine, l'aire V1 du cortex visuel primaire et la réponse des cellules complexes. Notre objectif n'est pas de donner une description détaillée de la biologie du système visuel mais d'en dégager l'organisation générale et les structures principales impliquées dans la perception 3D ce qui permettra de comprendre les choix effectuer pour le développement du modèle informatique présenté au chapitre suivant. Le lecteur intéressé pourra se reporter aux ouvrages de références suivants : [VV90] [SW90] [KSJ91] [BI87] [McI96].

3.1 Architecture fonctionnelle du système visuel

Le système visuel représente l'une des aires les plus volumineuse du cortex de l'Homme (Figure 3.1). L'information lumineuse est captée par les photorécepteurs de la rétine, placée au fond de l'oeil. L'information visuelle subit alors une première série de filtrages. Elle transite ensuite par le nerf optique en séparant les deux hémi-champs visuels à travers le chiasma optique jusqu'aux corps genouillés latéraux. Le premier rôle du CGL (un CGL pour chaque latéralité du cerveau) est une fonction de relais et d'organisation des afférences rétiniennes avant leur projection sur l'aire visuelle V1, située dans le lobe occipital. Il joue également un rôle dans la fusion stéréoscopique des informations provenant de chaque hémi-champs visuels. Cependant il est à noter que ces fibres rétiniennes ne représentent que 10% des afférence totales du CGL. Les 90% restantes provenant directement de V1, c'est-à-dire des informations non-visuelles (ce qui impliquerait que le rôle du CGL est bien plus important et qu'il interviendrait notamment dans la contextualisation de la perception et la modulation des processus attentionnels). L'information se projette ensuite sur les aires du cortex visuel primaire V1 (V2,V3) où l'information visuelle est décomposée en un ensemble de caractéristiques basiques (par exemple les fréquences, l'orientation, la couleur). L'information se diffuse aux aires supérieures V4, MT et IT (cortex inféro-temporal). L'information se complexifie au fur et à mesure et chaque aire analyse des informations de plus en plus haut niveau (par exemple V4 analyserait les objets, MT, le mouvement, IT coderait des primitives complexes). L'information peut être hiérarchisée suivant deux voies : la voie du *Où* (MT) et le voie du *Quoi* (V4, IT).

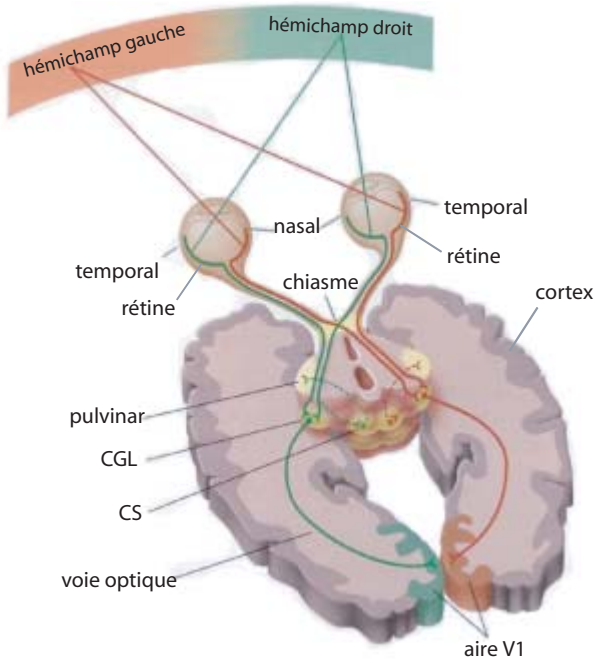


FIG. 3.1 – Les grandes étapes du système visuel (tiré de [Per03]) ; cette figure schématise les principales structures présentes dans le système visuel que nous considérons pour la perception 3D à partir de la texture ; l'information visuelle passe à travers l'oeil (A) et se projette sur la rétine ; elle transite ensuite par le nerf optique en séparant les deux hémi-champs visuels à travers le chiasma (B) jusqu'aux corps genouillés latéraux (CGL)(C) ; l'information se projette ensuite sur les aires du cortex visuel primaire V1 (V2,V3).

3.2 Prétraitements rétiniens

La rétine constitue le premier élément de la chaîne de traitement de l'information visuelle. Celle-ci est loin de se réduire à un simple capteur d'acquisition d'image mais au contraire réalise un ensemble de prétraitements spatio-temporels et chromatiques du signal lumineux avant que celui-ci ne soit analysé par les aires corticales supérieures. Nous présentons ici une description succincte de la physiologie de la rétine accompagnée par des simulations obtenues par le modèle de rétine de Hérault-Beaudot. Pour une description plus détaillée de la rétine le lecteur pourra se reporter à [BI87] et [MB99]. Le modèle présenté ici a déjà fait l'objet de nombreux travaux [Bea94] [H96] [All99] [TH99] [H99] [H01] [H05] [Dur05]. Dans ce travail nous nous intéressons plus particulièrement aux traitements effectués par la rétine sur une information statique et en vision achromatique.

La rétine se situe au fond de l'oeil et reçoit l'information lumineuse à travers la cornée et le cristallin qui réalisent les opérations de focalisations optiques. L'image formée sur la rétine est nette au niveau de sa partie centrale, appelée la fovéa. Celle-ci représente environ 1^o d'ouverture sur les 140^o couverts par le champ visuel au total (largeur d'un pouce vu lorsque le bras est tendu).

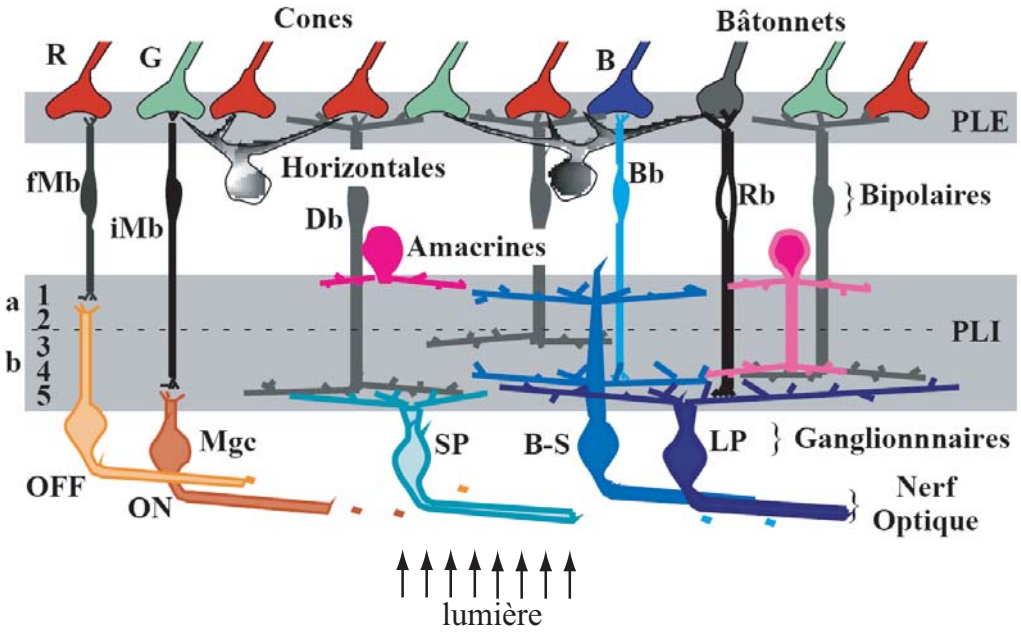


FIG. 3.2 – Diagramme simplifié de l'organisation de la rétine (tiré de [HÓ1]) ; les cellules de la rétine sont arrangées en différentes couches ; les photorécepteurs sont situés en haut, près de la paroi de l'oeil ; les corps des cellules horizontales et des cellules bipolaires constituent la couche des noyaux internes ; les cellules amacrines sont situées près des cellules ganglionnaires près de la surface de la rétine ; les connections neuronales axones/dendrites entre ces cellules permettent de les répartir en différentes couches plexiformes (PLE et PLI).

La rétine (Figure 3.2) est organisée en différentes couches de corps cellulaires. Entre ces couches se réalisent les connexions entre les cellules d'une couche et la suivante (connexions axones/dendrites). Elles constituent les couches plexiformes (couche plexiforme externe (PLE) et couche plexiforme interne (PLI)) et réalisent un ensemble de traitements. L'information lumineuse est captée par les photorécepteurs et est envoyée ensuite en sortie des cellules ganglionnaires aux corps genouillés latéraux (CGL) à travers le nerf optique.

3.2.1 Les photorécepteurs

Ils traduisent l'information lumineuse en potentiel électrique : c'est le phénomène de transduction. Il existe deux types de photorécepteurs : les cônes et les bâtonnets. Les cônes servent à la vision diurne (vision photopique, luminance élevée) et se divisent en 3 groupes sensibles au rouge (type L, grandes longueurs d'ondes), au vert (type M, longueurs d'ondes intermédiaires) et au bleu (type S, petites longueurs d'ondes). Ils permettent donc de distinguer la couleur et également les détails des objets ou de la scène. Les bâtonnets, beaucoup plus nombreux, servent à la vision nocturne (vision scotopique, faible luminosité). Dans une rétine il y a environ 5 millions de cônes répartis essentiellement dans la fovea et 120 millions de bâtonnets, répartis en dehors de la fovea. Pour prendre en compte la grande dynamique que couvre l'éclairement naturel, un mécanisme de compression adaptative est associée aux

photorécepteurs. La loi de Michaelis-Menten (Figure 3.3) permet de reproduire cette compression : en faible luminosité, une faible compression est appliquée afin d'étaler les valeurs ; en forte luminosité, une forte compression est appliquée afin de limiter la gamme des valeurs. Ce processus dynamique et relativement simple permet de manière efficace de réhausser les informations présentes dans des zones d'ombres tout en limitant les zones fortement éclairées. La figure 3.4 à droite montre un exemple de simulation de la compressions opérée par les photorécepteurs sur une scène. Nous noterons enfin que la répartition des photorécepteurs n'est pas uniforme. La densité est maximale au niveau de la fovea et décroit progressivement en périphérie.

Loi de compression:

$$I_{compressé}(p) = \frac{(255 + X_0(p))I_{entrée}(p)}{X_0(p) + I_{entrée}(p)}$$

Adaptation locale:

$$X_0(p) = 0.1 + \frac{410 I_{moyen}(p)}{105.5 + I_{moyen}(p)}$$

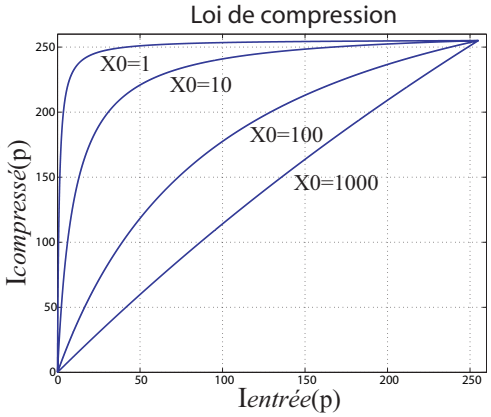


FIG. 3.3 – Loi de compression des photorécepteurs ; chaque position de l'image est notée p ; $I_{entre}(p)$ représente la valeur en p de l'image d'entrée ; $I_{compresse}(p)$ représente la valeur obtenue après compression ; $X_0(p)$ est le paramètre d'adaptation local à la luminosité moyenne autour de la position p et notée $I_{moyen}(p)$; à droite, le graphique représente la loi de compression pour différentes valeurs de $X_0(p)$; pour de faibles valeurs, un étalement des valeurs est obtenu ; pour de fortes valeurs, une forte compression permet de limiter la dynamique de sortie.

FIG. 3.4 – Exemples de scène naturelle ; à gauche : image d'entrée ; à droite : image compressée.

3.2.2 La couche plexiforme externe (PLE)

La PLE correspond à la couche où se réalisent la connexion entre les photorécepteurs, les cellules horizontales et les cellules bipolaires, formant ainsi une *triade synaptique* (Figure 3.5).

Les cellules horizontales interconnectent plusieurs photorécepteurs pour former une version passe-bas de l'image d'entrée (Figure 3.7 à gauche). Le couplage spatial des photorécepteurs est réalisé à l'aide de jonctions GAP permettant un lissage de leur réponses. De même le couplage spatial des cellules horizontales est réalisé à l'aide de ce type de jonctions. Les cellules bipolaires réalise la différence entre les réponses des photorécepteurs et les cellules horizontales. En fonction de la polarité de la triade synaptique, le champ récepteur des cellules bipolaires se décompose en deux parties : un centre excitateur et une périphérie inhibitrice (cellules ON) ; un centre inhibiteur et une périphérie excitatrice (cellules OFF). La figure 3.7 au milieu et à droite montre des exemples de simulations des sorties des cellules bipolaires ON et OFF. La différence ON-OFF permet finalement de filtrer les basses fréquences et de ne conserver que les hautes fréquences (Figure 3.6). Cette opération permet ainsi de ne récupérer que les contours présents dans l'image.

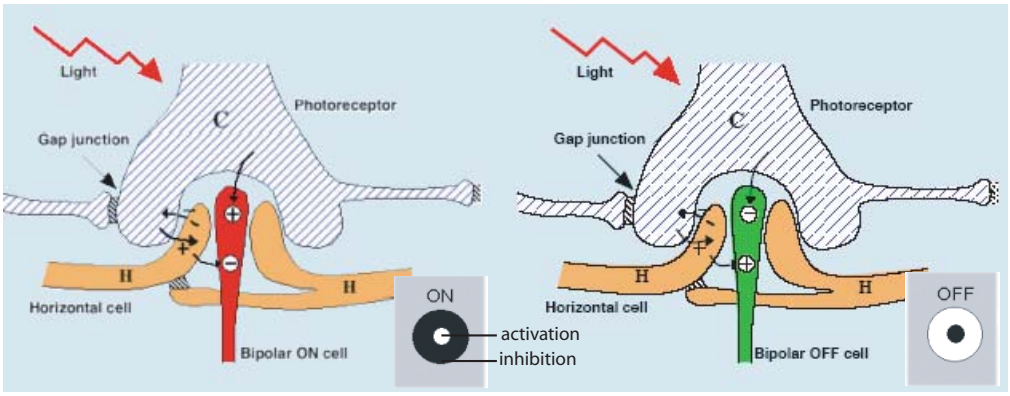


FIG. 3.5 – Schéma des connexions au niveau de la PLE (adapté de [HÓ1]) ; les polarités modifient les zones d'excitation ou d'inhibition des cellules bipolaires les séparant en 2 types : les ON et les OFF.

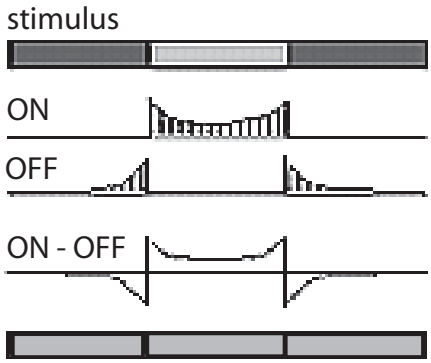


FIG. 3.6 – Réponse des ON et des OFF à un créneau lumineux (tiré de [Dur05]) ; la soustraction ON-OFF donne une version passe-haut de l'image.

FIG. 3.7 – A gauche : image de la sortie des cellules horizontales ; au milieu : image de la sortie des cellules ON ; à droite : image de la sortie des celules OFF.

3.2.3 La couche plexiforme interne (PLI)

La PLI correspond à la couche où se réalisent la connexion entre les bipolaires, les amacrines et les ganglionnaires. Les cellules amacrines sont réparties en près de 20 types différents dont les fonctions spécifiques ne sont pas encore bien connues. Elles ont une action essentiellement sur le gain des cellules bipolaires et des cellules ganglionnaires et sur le filtrage temporel de l'information. Les cellules ganglionnaires sont reliées aux sorties des cellules bipolaires et aux cellules amacrines. Elles sont nettement moins nombreuses que les photorécepteurs (environ 1,5 millions soit 80 fois moins nombreuses). Leurs axones correspondent à la sortie de la rétine et forment le nerf optique qui relie la rétine au corps genouillé latéral (CGL). Les cellules ganglionnaires se divisent en 3 types différents : P (80%), M (10%) et K (10%). Les cellules ganglionnaires P se projettent sur les couches parvocellulaires du corps géniculé latéral (CGL). Elles ont un petit champ récepteur, situé en général dans la fovéa et sont sensibles aux hautes fréquences spatiales et à la couleur. Les cellules M se projettent sur les couches magnocellulaires du CGL. Elles ont un champ récepteur large, essentiellement dans la parafovéa et véhiculent une information spatiale basse-fréquence achromatique. Ces cellules répondent plus rapidement que les cellules P et sont sensibles aux hautes fréquences temporelles. Les cellules ganglionnaires K ont des propriétés variées. Elles se projettent sur la voie koniocellulaires dont la fonction est encore mal définie et elles seraient relatives aux cônes bleus.

En sortie des cellules ganglionnaires apparaissent deux voies importantes du transfert de l'information visuelle vers les aires supérieures :

- **la voie magno** : elle est obtenue par combinaison entre la sortie des cellules bipolaires et les amacrines puis par différences entre les réponses ainsi obtenues (Figure 3.9) ; cette voie transmet des informations spatiales globales (régions ou blobs importants de l'image définissant un contexte global) (Figure 3.8 à gauche) et également les informations temporelles tel que le mouvement ; l'information est transmise très rapidement.
- **la voie parvo** : elle est obtenue par différence des sorties de cellules bipolaires (ON-OFF), celles-ci subissant une compression adaptative dans les cellules ganglionnaires ; cette voie transmet des informations de contraste spatial, les détails des formes et des contours des objets (Figure 3.8 à droite) et également l'information de couleur ; l'information est transmise plus lentement que la voie magno mais est porteuse d'une information détaillée sur la scène observée.

FIG. 3.8 – A gauche : image de la voie parvo ; à droite : image de la voie magno.

3.2.4 Résumé du prétraitement rétinien

Le schéma 3.9 représente l'architecture simplifiée de la rétine de l'image d'entrée aux voies parvo et magno.

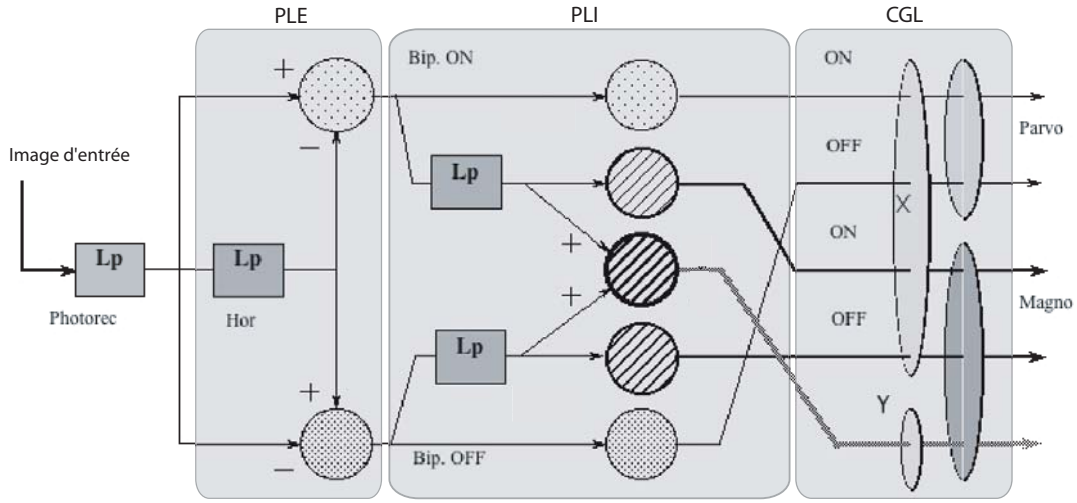


FIG. 3.9 – Schéma du modèle de rétine (tiré de [Dur05] et adapté de [HÓ1]) ; LP (low-pass) correspond au filtrage passe-bas opéré par les cellules ; la voie konio, mal définie, n'est pas reportée.

Le filtrage opéré par les photorécepteurs est à large bande conduisant à l'élimination des très hautes fréquences spatiales (bruit). Le filtrage opéré par les cellules horizontales est passe-bas. L'inhibition des cellules horizontale sur les photorécepteurs conduit à un filtrage passe-bande. Celui-ci est centré sur les fréquences utiles présentes dans l'environnement naturel. Or le spectre d'amplitude des scènes naturelles présente la propriété de décroître en moyenne en $1/f^2$ correspondant à la présence de beaucoup d'énergie en basses fréquences [AR92]. Ainsi le préfiltrage de l'image d'entrée par la rétine conduit à un rehaussement de l'énergie des hautes fréquences spatiales, c'est-à-dire à un blanchiment du spectre(Figure 3.10).

Ce prétraitement rétinien s'avère donc très important pour le traitement de l'information de texture dans les scènes naturelles car il permet de rehausser le contraste des hautes fréquences (rehaussement de l'information de texture proprement dite) tout en réduisant les différences d'intensité uniquement dues à des conditions d'illumination variables dans les

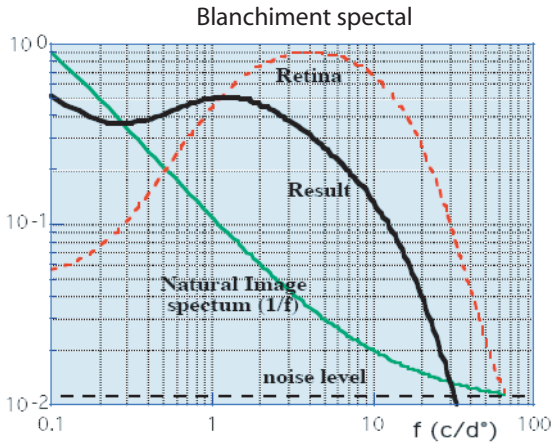


FIG. 3.10 – Graphique (tiré de [HÓ1]) montrant le spectre des images naturelles (en $1/f^2$), la fonction de transfert spatial de la rétine (en pointillé) et le produit des deux (Result) ; Cette dernière courbe est à peu près plate sur deux ordres de grandeurs de la fréquence spatiale, le spectre de l'image a donc été blanchi.

différentes zones de l'image. La sortie de la voie parvo permet ainsi de récupérer essentiellement l'information portée par la texture qui peut alors être utilisée par les aires corticales supérieures (Figure 3.11). La figure 3.12 montres des exemples de résultats obtenus après prétraitement et correspondant à l'information d'entrée du modèle de V1 décrit au chapitre 6.

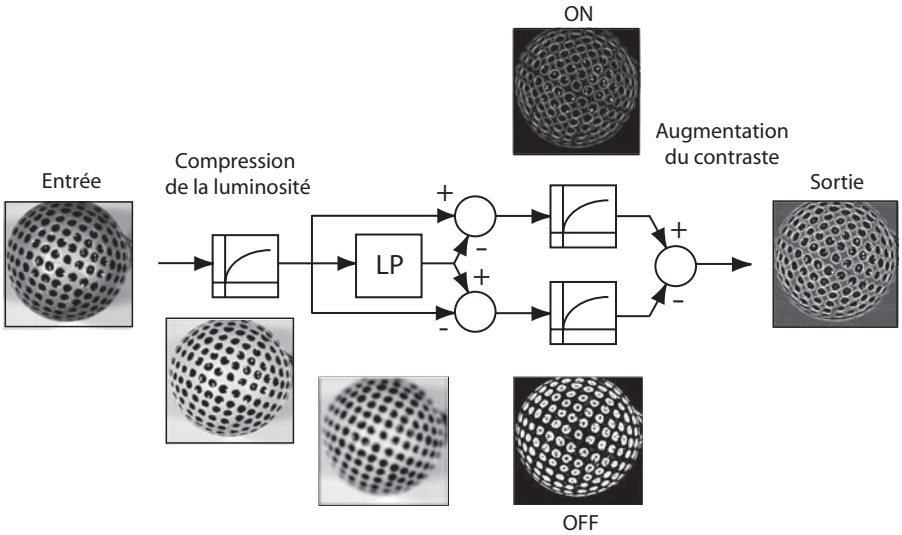


FIG. 3.11 – Etapes du modèle de la voie parvo utilisé pour l'analyse de la texture : compression de la luminosité de l'image d'entrée ; filtrage passe-bas et combinaison avec l'entrée compressée ; seconde compression conduisant à un rehaussement de contraste ; sortie finale où les ombres ont été enlevées et l'information de texture, rehaussée.

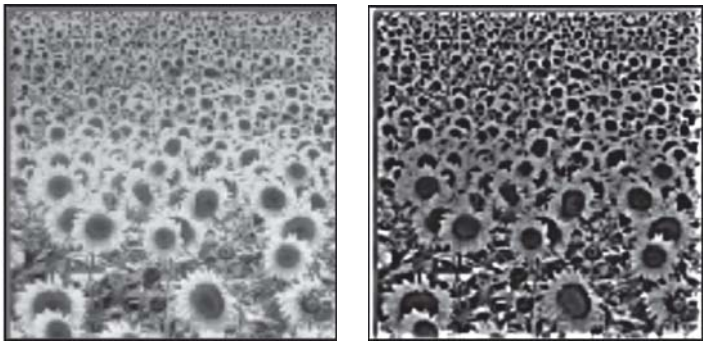


FIG. 3.12 – Exemples de résultats obtenus après pré-traitements par la rétine ; à gauche : image initiale ; à droite : image prétraitée par la voie parvo de la rétine.

3.3 L'aire corticale V1

L'aire V1 (ou 17 selon la classification de Brodmann) est l'aire principale dévolue à la vision. Toute lésion dans cette structure conduit à une perte complète de la perception visuelle consciente. Le nombre de neurones de cette aire (environ 350 millions) est beaucoup plus élevé que le nombre de photorécepteurs et à chaque axone afférent correspondent plusieurs centaines de neurones. La figure 3.13 montre un schéma de l'organisation complexe de V1 suivant un principe de regroupement entre neurones participant à la même tâche [Bul98]. Pour réaliser l'analyse fine et complète de l'image rétinienne ces regroupements se font suivant la structure cible (V2, autres aires corticales, structures sous-corticales), la voie afférente provenant du CGL (P,M,K), la dominance oculaire (certaines neurones reçoivent également des signaux binoculaires), la sélectivité à l'orientation, la sélectivité aux fréquences spatiales et l'organisation spatiale (représentation du champ visuel sur la rétine ou *rétinotopie*). V1 est divisée en différentes couches cytologiques parallèles chacune dédiées soit aux traitements des signaux afférents provenant du CGL (couches $4C\alpha$ et $4c\beta$), soit à la projection sur le colliculus supérieur et le CGL (couches 5 et 6), soit à la projection vers les autres aires corticales (couches 2, 3 et 4B).

En plus de cette structure laminaire, V1 est également organisée en colonnes fonctionnelles ([HW74]) (organisation qui se retrouve dans toutes les structures du cortex). Il est possible de distinguer les *macro-colonnes* fonctionnelles se regroupant autour de *blobs* de cytochrome oxydase (dédié au signal chromatique) pour une dominance oculaire particulière. Chaque macro-colonne possède une structure en *pinwheel* de colonnes d'orientation et de fréquence située dans la zone située entre les blobs (interblobs). Les colonnes d'orientation sont notamment formées de cellules très sensibles à la direction du stimulus visuel [DeV91] et couvrent entre 15 et 20 orientations. Les colonnes de fréquence réalisent également une décomposition des signaux afférents autour de 6 à 10 fréquence centrales ([SHS+97], [SW90]).

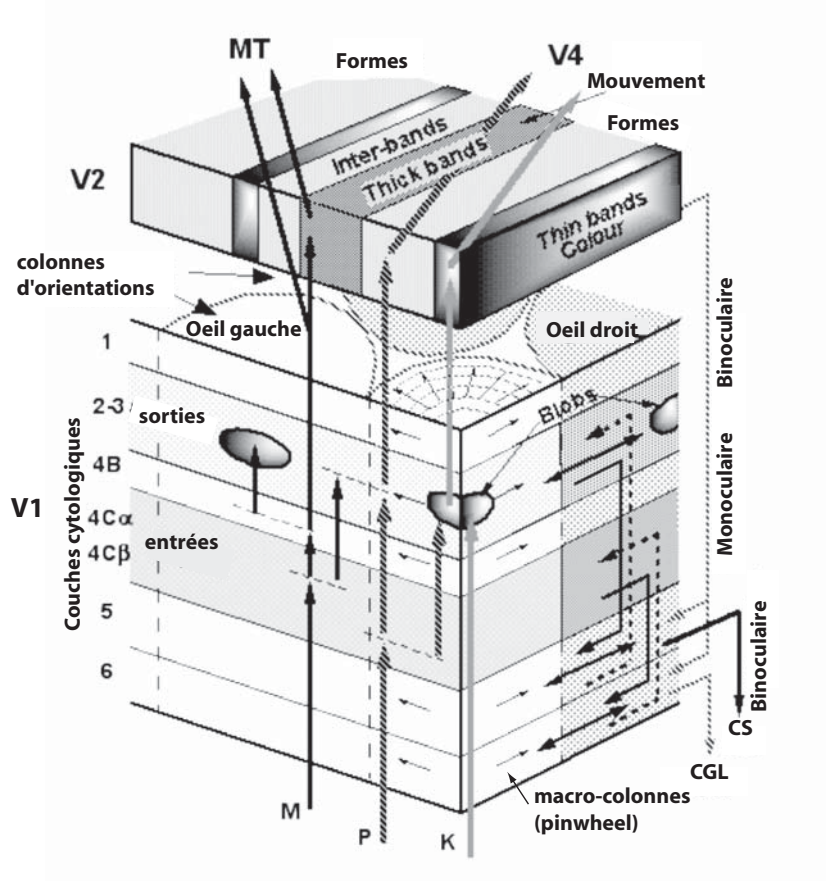


FIG. 3.13 – Schéma représentant l'organisation de V1 et ses connections avec le CGL et V2 ; à remarquer la structure en couches avec le nombre important d'intraconnections par couche et d'interconnections entre les couches ; à remarquer également l'organisation en macro-colonnes et en colonnes d'orientation et de fréquences organisées en *pinwheel*.

3.4 Les cellules corticales

Les neurones de l'aire visuelle V1 possèdent des caractéristiques liée à leur organisation spatiale et à leur mode d'activation. Les propriétés décrites ici ne sont pas exhaustives mais elles correspondent à celles qui sont impliquées dans la perception 3D en vision monoculaire.

3.4.1 Rétinotopie et pavage du champ visuel

La majorité des neurones participant à la structure du système visuel peuvent être activés par une stimulation lumineuse quand celle-ci est appliquée sur une petite région du champ visuel. Cette région définit le *champ récepteur* du neurone considéré. Nous considérons ici que que le champ visuel est défini comme une petite région située sur un plan dans la région de la fixation oculaire.

Dans beaucoup des aires impliquées dans le traitement visuel, les champs récepteurs des neurones ne sont pas disposés aléatoirement mais au contraire respecte la topologie locale des

cellules de la rétine, c'est-à-dire que ces aires contiennent une représentation rétinotopique du champ (ou de l'hémichamp) visuel. Ainsi les neurones du CGL et de V1 formes des *cartes corticales* qui respectent les relations spatiales entre les champs récepteurs. Cette représentation est relativement linéaire pour les neurones du CGL et présente une forte non-linéarité pour ceux de V1 conduisant à une sur-représentation de la partie centrale de la rétine par rapport à la périphérie (80% de V1 est dédié au traitement à la vision centrale limitée à 10^o de champ visuel sur un total approchant 140^o, Figure 3.14 à gauche). Elle peut être modélisée par une fonction logarithmique complexe (Figure 3.14 à droite) se caractérisant par : la partie centrale correspondant à la fovea est représentée sur un repère cartésien ; la partie périphérique est représenté sur un repère logpolaire, l'axe des abscisses correspondant au facteur de zoom et l'axe des ordonnées, à l'orientation. Ce type de représentation présente notamment la propriété intéressante de coder les variations de zoom et d'orientation par deux translations (horizontale pour le zoom et verticale pour la rotation) facilement détectables par comparaison entre les réponses de neurones voisins.

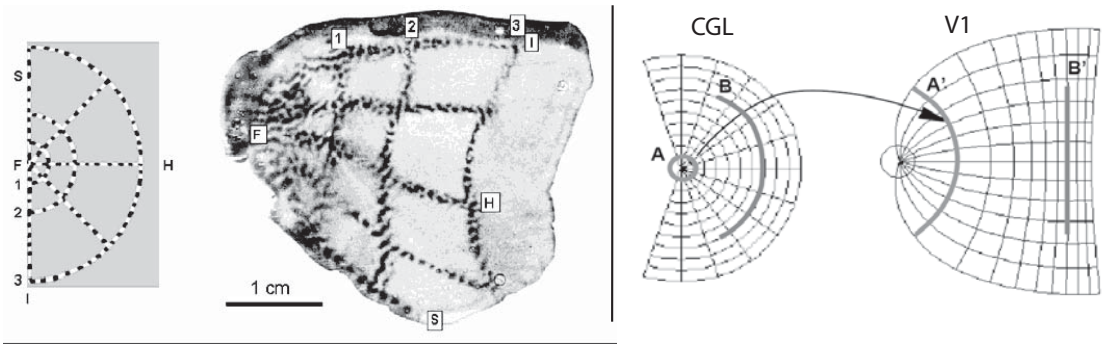


FIG. 3.14 – Rétinotopie des neurones de l'aire visuelle V1 ; à gauche : image de l'activité neuronale dans le cortex du singe produit par le stimulus visuel A ; B est la photographie du cortex strié déplié où l'activité de la deoxyglucose a imprégné les neurones stimulés [TSSV82] ; à droite : modélisation du champ visuel dans le CGL (à gauche) et sa projection sur l'aire V1 modélisé par une fonction logarithmique [Sch80].

Les taille des champs récepteur de V1 et V2 sont relativement petites et sont susceptibles d'indiquer un changement dans une petite région du champ visuel. Contrairement aux neurones des aires supérieures, telles que IT, dont le champ récepteur plus large permet d'intégrer des stimuli plus larges. Aussi comme le note Bullier [Bul98], la vraie différence entre les neurones des différentes aires n'est pas leur complexité mais la taille de la région spatiale sur laquelle ils peuvent intégrer de l'information. En progessant dans les aires corticales supérieures, la taille des champs récepteurs des cellules associées augmente progressivement jusqu'à recouvrir l'ensemble du champ visuel pour certains d'entre eux. Cependant la complexité de la réponse de ces cellules augmente également et permettent aux différentes aires corticales d'extraire une information de plus plus en complexe et précise de la scène observée. La figure 3.15 présente schématiquement le pavage réalisé par les neurones des différentes aires corticales. Les champs récepteurs des photorécepteurs de la rétine, relativement petits, pave avec un fort résolution l'image d'entrée. Les neurones des aires suivantes (V1 et V2) avec des champes récepteurs plus larges intègrent une information spatiale plus étendue.

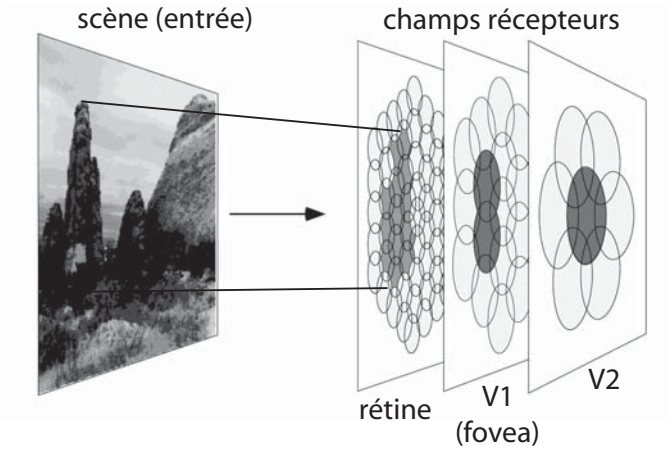


FIG. 3.15 – Schéma montrant la répartition des champs récepteurs des neurones dans les différentes couches corticales ; les neurones de V1 et V2 sont arrangés en respectant la rétinotopie avec beaucoup de neurones pour chaque position locale du champ visuel ; les neurones répondant à l'un des rochers de la scène présentée à gauche sont coloriés en gris.

3.4.2 Taille des champs récepteurs

L'une des caractéristiques importantes du cortex cérébral est sa capacité à se réorganiser en réponse à une altération à long-terme des signaux afférents. Ce mécanisme s'avère crucial notament pour le rétablissement après un dommage au niveau du cortex (lésion). Au contraire l'existence d'une plasticité à court-terme (de minute en minute) et son rôle dans la représentation de l'information visuelle dans les aires corticales reste mal définie. Un changement important de la taille du champ récepteur aurait pour conséquence une modification des propriétés de filtrage du neurone, notamment sa sélectivité fréquentielle.

Le champ récepteur de chaque neurone est traditionnellement considéré comme ayant une structure fixe. En d'autres termes sa réponse et sa taille restent identiques une fois le période de développement du cortex achevée (par exemple la taille des champs récepteurs des neurones de V1 est de l'ordre de 1^o). Cependant une plasticité à long-terme a été observée chez le chat : une lésion est opérée dans la fibre géniculocorticale (du CGL à V1) entraîne l'arrêt de l'activation des neurones récepteurs de V1 correspondant à la région spatiale affectée (par la rétinotopie) créant un *scotome* ; cependant des champs récepteurs des neurones voisins a augmenté sans modifier leur sélectivité à l'orientation et à la fréquence en comblant en grande partie le scotome [ES99]. Pour étudier l'existence d'une plasticité à court-terme, Pettet et Gilbert [PG92] et DeAngelis *et al* [DAOF95] ont effectué des expériences en créant des scotomes artificiels (un stimulus est présenté dans le champ visuel et est placé dans la zone couverte par le champ récepteur de la cellule étudiée ; il correspond à une image formée d'un fond texturé avec une carré gris superposé pour désactiver la région spatial) (Figure 3.16 à gauche).

Dans les deux travaux, effectués dans l'aire V1 chez le chat, une modification de l'activité des neurones voisins a été observée. Alors que Pettet et Gilbert conclurent à une augmentation d'un facteur 5 de la taille de leur champ récepteur, DeAngelis *et al* montrèrent de manière plus précise que la taille des champs récepteurs et leur propriétés de filtrage n'étaient pas

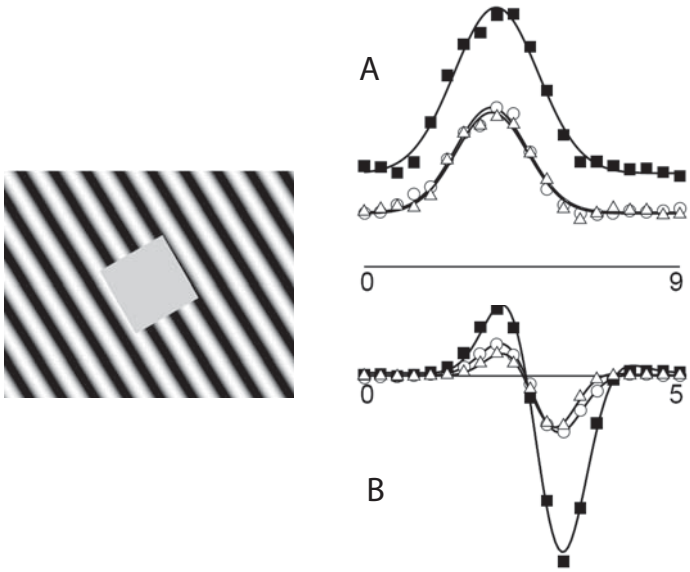


FIG. 3.16 – A gauche : exemple de stimulus permettant de simuler un scotome (scotome artificiel) ; à droite : effets d'un scotome artificiel sur les profils 1-D d'une cellule complexe (A) et d'une cellule simple (B) obtenus par corrélation inverse [DAOF95] ; les carrés pleins représentent les données en présence du scotome artificiel ; les cercles et les triangles blancs représentent les données avant et après l'application du scotome ; les données sont reportées en réponse absolue (spike/stimulus).

modifiées, uniquement l'amplitude de la réponse de certaines cellules augmentait sensiblement et revenait à leur niveau initial après retrait du scotome (Figure 3.16 à droite). Les auteurs conclurent que la taille des champ récepteurs des cellules corticales restait donc fixe dans une perception à court-terme. Plus précisément ils observèrent une augmentation généralisée à l'ensemble des cellules et non pas uniquement de celles situées dans la région du scotome, laissant supposer l'existence de connexions horizontales à longue portée entre les cellules corticales. Ces interactions pourraient se modéliser par une normalisation divisive appliquée à chaque cellule (shunting inhibition) à partir d'un ensemble de cellules dont le champ récepteur couvre d'autres régions spatiales [Hee93]. Cela s'accompagnerait d'une adaptation temporelle de la réponse de cet ensemble de cellules conduisant à la réduction progressive de l'inhibition provoquée par le scotome. Ce mécanisme conduirait à une augmentation du gain de la réponse de la cellule étudiée (gain multiplicatif) sur une courte durée (1 à 2 seconde).

3.4.3 Les cellules simples et complexes

L'ensemble des propriétés du stimulus (par exemple la taille, la couleur, la forme, l'orientation, le mouvement) définissent la *sélectivité* du champ récepteur du neurone considéré. Il est à noter qu'il peut être aussi temporel (champ récepteur spatio-temporel, [DOF95]) à variables séparables ou inséparables en espace et en temps. C'est le réseau complexe des connexions depuis les photorécepteurs jusqu'aux cellules des aires supérieures et les interactions locales qui permet d'isoler et de spécifier la sélectivité du champ récepteur. Dans la rétine, nous avons

décrit un modèle permettant d'obtenir un champ récepteur concentrique inhibiteur/excitateur pour les cellules bipolaires et de définir les propriétés de sorties des cellules ganglionnaires. La modélisation est plus difficile pour les cellules du CGL et des aires supérieures car le codage opéré dans le nerf optique pour la transmission reste encore mal connu (voir à ce sujet [Per03] pour un modèle de codage épars dynamique faisant notamment émerger des réponses de filtres similaires aux cellules de V1). Il faut alors avoir recours à des enregistrements neurophysiologiques des groupes de cellules dans l'aire étudiée et faire varier les caractéristiques des stimuli pour pouvoir modéliser leur réponse. Cette approche étant analogue à l'obtention de la réponse impulsionnelle d'un filtre en analyse du signal, le terme de *filtrage cortical* est employé. Les travaux de Hubel et Wiesel [HW62] [HW68] [HW74] ont permis de mettre en évidence l'existence de deux types de cellules dans l'aire V1, les cellules *simples* et les cellules *complexes*, répondant à des stimulations suivant une orientation et une bande de fréquence donnée. La figure 3.17 montre la structure des champs récepteurs des cellules du CGL (similaire à celle des cellules ganglionnaires de la rétine) et des cellules corticales (simples et complexes).

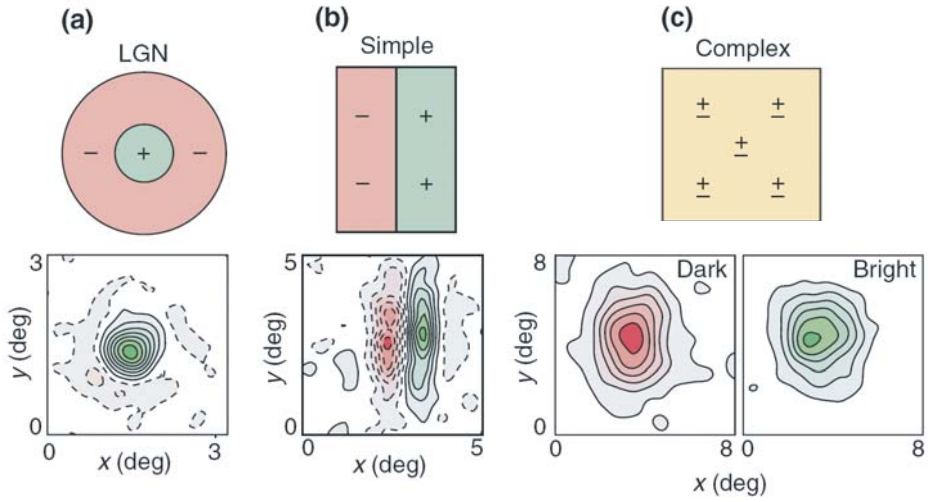


FIG. 3.17 – Structure des champs récepteurs des classes principales de neurones de la voie geniculo-striée (du CGL à l'aire V1) [DOF95].

Les cellules simples répondent linéairement à une stimulation et leur réponse est en général modélisée par un filtre de Gabor ([Dau80] [Mar80] [JP87] [Rin02]) (Figure 3.17 B). Ce filtre correspond à une sinusoïde modulée par une fenêtre gaussienne. Il existe deux types de cellules simples : les cellules en phase et en quadrature (ce qui correspond à prendre un filtre de Gabor basé respectivement sur un cosinus ou un sinus). Leur réponse dépend ainsi de la position du stimulus dans le champ récepteur de la cellule.

Wallis remet en question la modélisation des cellules simples par des filtres Gabor [Wal01]. En effet la plupart des données neurophysiologiques sur les profils de réponses de ces cellules montrent que les données mesurées sont souvent mal modélisées par une fonction gaussienne. Notamment ces profils sont symétriques sur une échelle log-fréquence. Or cette propriété n'est pas vérifiée par les filtres de Gabor qui sont assymétriques en log-fréquence (Figure 6.1). Wallis

montre que des différences de gaussiennes ou des modèles de Cauchy sont mieux adaptés car ils respectent cette propriété.

Les cellules complexes répondent non-linéairement à une stimulation. Leur réponse ne dépend pas de la position du stimulus dans le champ récepteur de la cellule et elle est bien simulée par la somme des carrés des réponses des cellules simples prise en phase et en quadrature (Figure 3.17 C) (voir [CDM+05] pour une revue des différents modèles des cellules complexes).

3.5 Correlas neuronaux de la perception 3D basée sur la texture

L'une des questions importantes en neuroscience est de savoir comment le système visuel reconstruit une représentation tridimensionnelle de son environnement à partir des informations bidimensionnelles projetées sur la rétine. De nombreuses études se sont portées sur le traitement de la disparité binoculaires mais peu sur l'analyse des gradients de texture malgré leur importance dans la perception 3D (voir Chapitre 4).

Gallant *et al* [GEN95] ont étudié la réponse de cellules dans V4 chez le singe. Ils ont ainsi trouvé des cellules répondant à différentes configurations de slant et de tilt. Cependant ils n'arrivèrent pas à mettre en évidence l'influence de l'information de profondeur donnée par les indices de texture (par exemple le changement de taille des texels), ces derniers ayant peu d'influence sur les résultats. Ils conclurent ne pas avoir trouvé de neurones de V4 répondant spécifiquement au slant et au tilt, mais ils n'exclurent pas la possibilité de l'existence de tels neurones dans d'autres aires corticales.

Tsutsui *et al* ont découvert un correla neuronal de la perception de la profondeur à partir de l'information de texture dans la partie caudale du sulcus intrapariétal (aire CIP, Figure 3.18 à droite). Pour les expériences, des singes sont entraînés à percevoir des surfaces texturées et inclinées dans l'espace (Figure 3.18 à gauche). Ils doivent indiquer si la surface présentée possède une orientation (tilt) identique à une surface de référence présentée auparavant (tâche go-nogo). Dans le même temps l'activité des neurones situés dans l'aire CIP sont enregistrées. Différentes textures sont présentées ainsi que des stimuli présentant un indice de disparité (des stéréogrammes composés de points aléatoires). Les résultats montrent que les singes utilisent l'indice de texture et l'indice de disparité de manière équivalente pour juger la configuration spatiale des surfaces. Les neurones de cette aire semblent donc à la fois sélectifs à différentes configurations d'orientation de surfaces planes indépendamment des indices contenus dans la texture. Cette étude montre ainsi à l'échelle de la réponse individuelle de neurones de l'aire CIP, l'existence d'une base neuronale du codage de l'information 3D à partir des gradients de texture.Ces auteurs ont également montré la sensibilité des neurones de cette aire à un gradient de disparité et faiblement à un gradient d'orientation induit par la perspective linéaire [TYST01].

Sereno *et al* ont mené des en imagerie fonctionnelle également au niveau de l'aire CIP [STAL02]. Ils ont trouvé, à l'instar de Tsuitsui *et al*, à la fois chez l'être humain et chez le singe, qu'une activation des neurones survient à la présentation de formes 3D définies par un gradient de texture et par un mouvement parallaxe.

Ces résultats suggèrent donc l'existence d'un correla neuronal à la perception 3D. Ils indiquent également la construction d'une représentation unifiée de l'orientation 3D des surfaces

par l'intermédiaire d'une structure corticale dédiée à la fusion des différents indices visuels tels que la disparité, les gradients de texture, la perspective linéaire et le mouvement parallaxe.

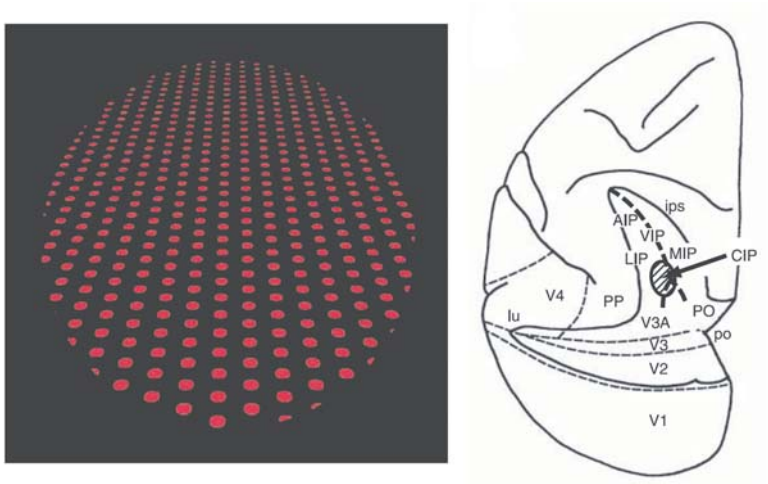


FIG. 3.18 – A gauche : exemple de texture représentant une surface inclinée utilisée par Tsutsui *et al* [TSNT02] lors d'une tâche d'appariement du tilt entre la surface présentée et une surface de test (l'enregistrement des cellules est effectué sur un singe) ; à droite : schéma indiquant la position de l'aire CIP entre l'aire LIP et V3A.

3.6 Résumé

Le système visuel est un mécanise complexe, encore largement inconnu. Cependant beaucoup de connaissances ont été accumulées et ont permis de bien définir les fonctions réalisées par les premières étapes. Dans ce travail nous nous intéressons à une version simplifiée du système visuel en considérant les étapes suivantes : le prétraitement rétinien dont la voie parvo est bien adaptée à l'analyse de la texture ; la décomposition du champ visuel en régions locales correspondant aux champ récepteurs des cellules de V1 (de taille fixe et en interaction avec les cellules voisines (normalisation locale)) ; la projection sur V1 et la décomposition de l'information visuelle en fréquence et orientation par les cellules complexes ; une analyse des gradients de texture par des aires située à la suite de V1.

A partir de ces données est-il possible de construire un modèle de perception 3D à partir des indices contenus dans la texture ? Ceci est l'objet du chapitre 6. La réponse à cette requière tout d'abord l'identification des indices de texture réellement extrait par le système visuel. Ceci est l'objet des deux prochains chapitres 4 et 5 décrivant notre étude réalisée en psychophysique.

Chapitre 4

Perception 3D : les indices de texture

Le système visuel est capable d'obtenir une information 3D aussi bien à partir de son environnement qu'à partir de l'analyse d'une image bidimensionnelle. Les recherches en psychophysique posent les questions fondamentales relatives aux types d'informations et de mécanismes mis en jeu par le système visuel pour percevoir la forme et l'orientation des surfaces. Cette étude permet de mieux cerner les possibilités et les limites du système visuel. Il s'agit d'en décrire au mieux le fonctionnement à partir de la description des performances visuelles des sujets obtenus sur différentes tâches.

Tout d'abord ce chapitre présente les différents indices classiquement étudiés en perception 3D sur des images monoculaires appelés les gradients de texture. Les hypothèses d'homogénéité et d'isotropie associées aux gradients de texture sont ensuite présentées et discutées. Les principales caractéristiques de la perception de la forme et de l'orientation de surface à partir de l'analyse de la texture sont décrites. Enfin deux indices supplémentaires, le gradient de fréquence et la perspective linéaire, sont plus particulièrement décrits et mis en relation avec les travaux de Li et Zaidi sur la caractérisation spectrale de l'information de forme par la texture.

4.1 Les gradients de texture

Au chapitre 2 nous avons vu que la texture est une source importante d'information pour la perception 3D. Cependant contrairement aux autres indices visuels (par exemple la disparité, la parallaxe de mouvement, l'ombre), il est difficile de caractériser mathématiquement les indices visuels liés à la texture. Ces indices apparaissent lors de la projection de la surface présente dans le monde 3D sur le plan de l'image 2D. Celle-ci induit un changement de taille et une déformation des éléments de la texture en fonction de l'orientation et de la forme de la surface initiale (Section 2.3). Cependant extraire une variable physique directement reliée à cette déformation est un problème difficile. Cela est notamment dû au fait que le concept même de texture ne repose pas sur une définition formelle mais au mieux sur une description statistique (voir Chapitre 2.2). Le problème encore non résolu et qui va être abordé est de savoir quels sont les indices de texture utilisés par le système visuel humain pour la perception 3D. Pour cela nous allons décrire les différents indices de texture classiquement étudiés.

Dans les années 50, Gibson a été le précurseur de l'étude des déformations de la texture induites par la projection sur le plan de l'image en introduisant le concept de *gradients de texture*. Il décrivit la texture comme *la structure d'une surface, à distinguer de la structure de la substance sous-jacente à la surface* ([Gib79]). Il sépara ainsi la texture de la surface sur laquelle celle-ci est plaquée. Pour caractériser la texture il introduisit alors le terme de *texels* pour désigner les éléments constituants la texture (par exemple : les briques d'un mur, les graviers sur le sol ou les fleurs d'un champ de tournesols)(Figure 4.1).

FIG. 4.1 – Image d'un champ de tournesols ; chaque fleur peut être considérée comme un texel selon la définition de Gibson.

En faisant une hypothèse de distribution uniforme de ces texels sur une surface, il interpréta toute modification de cette distribution dans l'image 2D comme provenant de la projection de la surface sur l'image. Cette variation reflète alors les caractéristiques géométriques (i.e la forme) et l'orientation (i.e la direction et l'inclinaison dans l'espace) de la surface initiale. Ces changements peuvent se traduire par des variations continues des caractéristiques des texels d'où le terme de *gradients de textures*. Ces gradients constituent la mesure physique caractérisant la déformation de la texture. Quatre types de gradient de texture sont en général distingués :

- le gradient de taille
- le gradient de compression
- le gradient de densité
- le gradient de perspective

La figure 4.2 montre ces différents gradients. Le gradient de taille est dû au fait que les *texels* proches de l'observateur apparaissent plus grands que ceux qui sont en profondeur. Il est ainsi inversement proportionnel à la distance de l'élément sur la surface par rapport à l'observateur. Le gradient de compression correspond à la déformation subie par le texel dans la direction de l'inclinaison. Celui-ci est compressé proportionnellement au cosinus du slant. Cette déformation fait apparaître une direction préférentielle et dans le cas d'une texture isotropique, cela introduit une anisotropie locale (voir 4.2). Le gradient de densité correspond à l'augmentation systématique du nombre de texels par unité de surface avec la profondeur. En général ce n'est pas la densité mais les positions relatives des texels qui sont calculées. Aussi ce gradient est parfois également appelé gradient de position. Le gradient de perspective correspond à un changement de position et à une variation de l'orientation des texels suivant des lignes de fuite et se croisant en un point appelé le point de fuite.

Les gradients de taille et de compression sont des gradients qui peuvent être calculés individuellement pour chaque texel en rapport à une référence (par exemple la compression

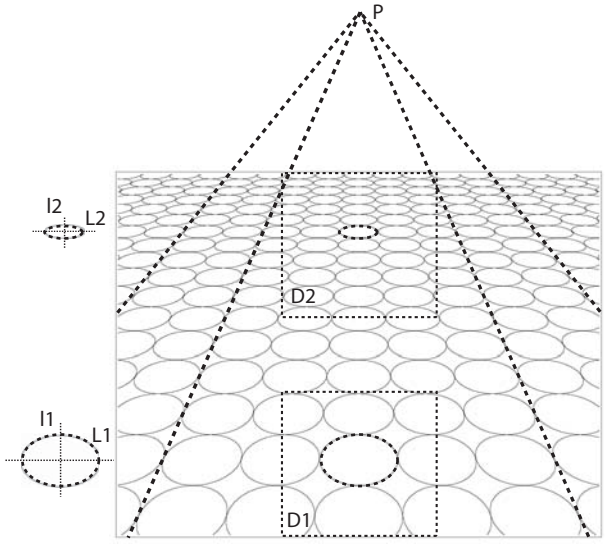


FIG. 4.2 – Projection d'une texture composée de cercles ; les 4 types de gradients sont reportés : le gradient de taille ($L1\mathrm{X}l1 > L2\mathrm{X}l2$) ; le gradient de compression ($L1/l1 < L2/l2$, les ellipses s'aplatissent avec la profondeur) ; le gradient de densité ($densite_{D_1} < densite_{D_2}$) ; la perspective linéaire : les lignes parallèles deviennent des lignes de fuite qui se coupent en un point de fuite (P).

d'une ellipse peut être mesurée en supposant celle-ci parfaitement circulaire à une inclinaison nulle). Dans ce cas ces gradients sont dits locaux. Les gradients de densité et de perspective sont des caractéristiques au contraire ne pouvant s'appuyer que sur la relation entre les texels. Ces gradients relèvent d'une intégration spatiale sur une région regroupant plusieurs texels et sont dits globaux.

Certains de ces gradients ne sont pas indépendants les uns des autres. Tandis que les travaux de Gibson se basaient sur l'hypothèse de densité uniforme, Stevens [Ste84] indiqua ensuite que le gradient de densité n'était pas la meilleure mesure de l'orientation d'une surface car celui-ci dépend directement des gradients de taille et de compression. Ces deux derniers peuvent cependant être en partie dissociés et être étudiés séparément. Le gradient de taille peut permettre de retrouver l'orientation de la surface au signe du tilt près. Le gradient de compression, lui, définit de manière unique l'orientation d'une surface sans ambiguïté sur le signe du tilt.

Les études sur les gradients de texture en psychophysique se sont développés parallèlement aux modèles en traitement d'images et en analyse de forme. Ces modèles se sont à la fois appuyés sur les indices psychophysiques (segmentation et analyse des texels [Alo88] [Gar92]) mais ont également montré l'existence d'autres codages possible de la déformation de la texture (notamment par l'étude fréquentielle [Wit81] [SB95b] [MR97]). Le lecteur trouvera au chapitre 2.4 une revue détaillée de ces modèles.

4.2 Analyse locale ou globale ? (isotropie ou homogénéité ?)

En parallèle avec le développement des modèles basés sur l'analyse des gradients, une autre approche s'est basée sur l'étude de la déformation locale de la texture. En psychophysique, ces deux modèles représentent deux stratégies possible employées par le système visuel. La question s'est posée en ces termes : la perception 3D monoculaire relève-t-elle d'une analyse locale ou d'une analyse globale de la surface perçue dans le champ visuel ?

Ces deux alternatives se sont traduites par deux types d'hypothèses sur les caractéristiques statistiques de la texture : l'hypothèse d'isotropie basée sur une analyse locale de chaque texel et l'hypothèse d'homogénéité basée sur une analyse globale par intégration sur l'ensemble de la surface.

4.2.1 Homogénéité

L'homogénéité d'une texture peut se traduire formellement comme la réalisation d'un processus stochastique spatialement stationnaire. Pour des surfaces avec une courbure gaussienne nulle (notamment les plans), l'homogénéité se traduit par une invariance des statistiques de la texture par translation sur la surface. Après projection, la variation de ces statistiques peut alors être utilisée pour retrouver le tilt et le slant en chaque position de la surface car elles ne dépendent que des paramètres de la projection et non pas des caractéristiques de la texture.

L'hypothèse d'homogénéité couvre tous les types de texture composées d'un seul type de texel. Elle est donc très générale mais en contrepartie elle n'induit pas une paramétrisation directe de la texture. Pour retrouver l'orientation de la surface il est d'abord nécessaire de définir des gradients mesurés à partir de la texture, comme nous l'avons décrit précédemment. Il s'agit, comme l'indiquent Malik et Rosenholtz dans [MR97], de considérer la déformation locale entre des régions voisines de la surface. Ainsi ce n'est pas la distribution des statistiques locales qui est importante mais leur modification d'une position à l'autre, donc à travers une analyse globale de la surface.

De nombreux modèles ont été développés en s'appuyant sur une hypothèse d'homogénéité de la texture [SB95b] [LK05] [MR97] [SB95a] [HLC98] [CM02] (Chapitre 2.4).

4.2.2 Isotropie

Au début des années 80, Stevens [Ste81] suggéra que l'estimation de la forme à partir de la texture pouvait être calculée en chaque position spatiale sous l'hypothèse de texels approximativement circulaires avant leur projection. Plus formellement, Witkin [Wit81] introduisit l'hypothèse d'*isotropie directionnelle* comme indice de texture pour la perception 3D, ceci constituant une alternative à la définition d'homogénéité de Gibson. Witkin montra que si une texture est initialement constituée d'une distribution d'orientations uniforme (i.e une distribution isotropique), la projection de la surface introduit un biais dans la répartition des orientations qui devient non-uniforme (i.e anisotropique) et cela en chaque point. L'anisotropie locale est donc une mesure de la quantité de déformation locale subie par la texture, ce qui permet de calculer en chaque point le tilt et le slant correspondant. L'hypothèse d'isotropie permet d'obtenir la forme à partir d'une série d'analyses locales de la texture, sans calcul de gradients par intégration sur l'ensemble de la surface.

La figure 4.4 représente une texture formée de cercles projetés en perspective. Les éléments sont des cercles avant projection et donc sont parfaitement isotropes. Après projection, ces

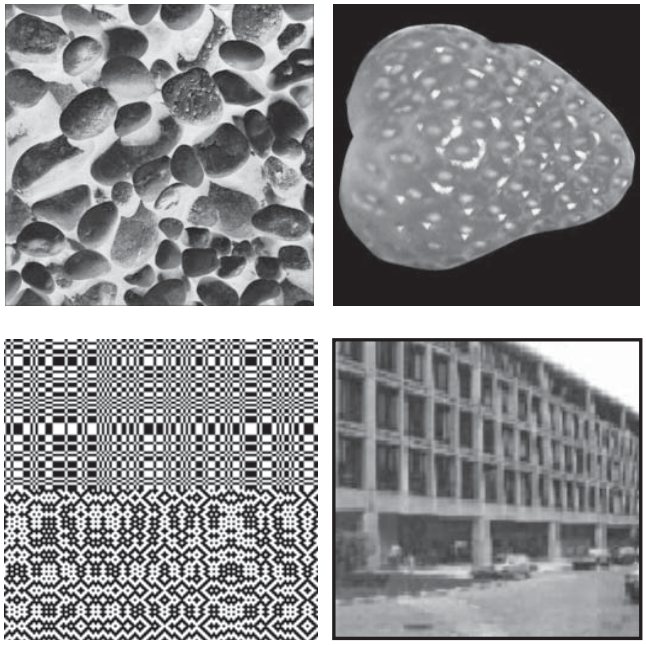


FIG. 4.3 – Exemples de textures homogènes en haut (texture de Brodatz [Bro66] et une fraise tirée de [LK05]) et non-homogènes en bas (texture artificielle et façade d'immeuble tirée de [HLC98]).

cercles sont déformés et cette déformation se traduit par une rupture de l'isotropie. Cette déformation peut être mesurée en chaque position spatiale de manière indépendante. La figure 4.5 montre des exemples de textures où l'hypothèse d'isotropie peut s'appliquer soit par l'analyse de la déformation des éléments (par exemple sur la boule de golf) soit par l'analyse de la distribution des orientations (par exemple sur les cylindres).

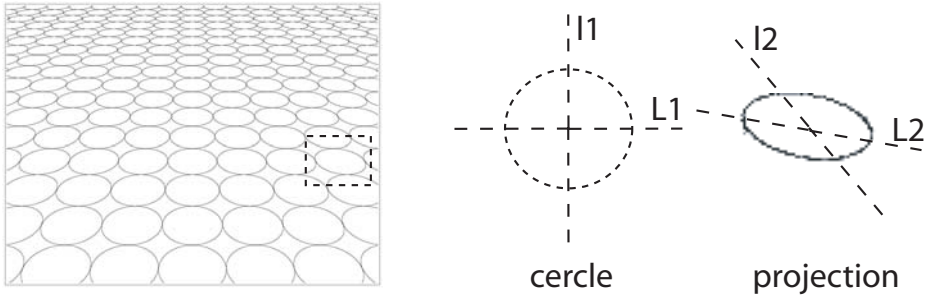


FIG. 4.4 – Exemple d'anisotropie lors de la projection en perspective d'un cercle.

Par rapport aux gradients de texture, Knill dans [Kni98c] indique que l'hypothèse d'isotropie est reliée au gradient de compression. En effet ni le gradient d'échelle, ni le gradient de position (dans la cas de textures irrégulières) n'introduisent de changement dans les statistiques des orientations et requièrent une hypothèse d'homogénéité pour être utilisable. La perspective linéaire n'est présente que si la texture exhibe des alignements. Le gradient de compression peut par contre être utilisé sous les deux hypothèses : la projection introduit une compression progressive des texels avec l'éloignement en profondeur et cette compression

s'effectue dans une direction déterminée en fonction du tilt et de la position spatiale sur la surface.

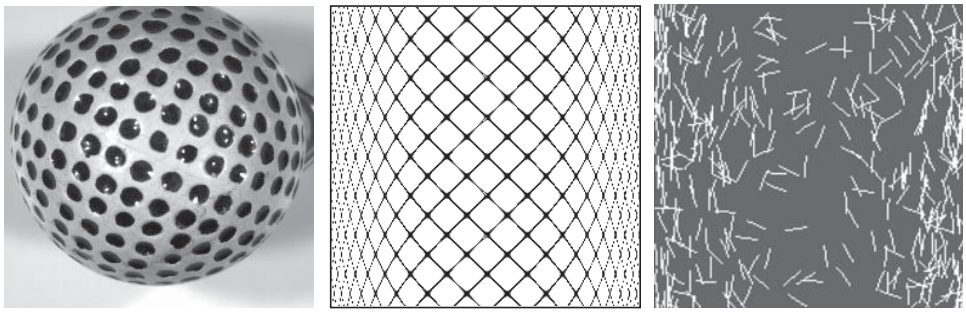


FIG. 4.5 – Exemples de textures isotropiques où l'hypothèse d'isotropie peut s'appliquer : une balle de golf tirée de [CM02], une texture plaquée sur un cylindre et une texture composée de segments courts disposés aléatoirement sur un cylindre (tirée de [AM04]) ; pour ces deux dernières textures le biais sur les orientations verticales est nettement visible sur les bords.

Malgré l'élégance de sa formulation, l'hypothèse d'isotropie ne couvre pas tous les types de texture, notamment les textures *directionnelles* (Figure 4.6).

FIG. 4.6 – Exemples de textures directionnelles où l'hypothèse d'isotropie n'est pas vérifiée : un champ, une texture anisotropique tirée de [TTD05] et une autre tirée de [CM02].

Stone [Sto93] suggéra une autre hypothèse : l'homotropie (Figure 4.7). Elle correspond au fait que la distribution de l'orientation des vecteurs locaux tangents aux contours dans l'image soit invariante avec la position. Cette hypothèse est une forme d'anisotropie avec une contrainte supplémentaire.

Différents modèles se sont basés sur cette mesure de l'anisotropie locale pour retrouver la forme par la texture [BM90] [BS90] [Gar93] [LK05] notamment en utilisant les statistiques d'ordre deux de petits éléments dans le domaine spatial ou en travaillant sur les caractéristiques des spectres locaux (Chapitre 2.4).

4.2.3 Analyse locale ou globale ?

Comme le note Knill dans [Kni98c] la connaissance *a priori* de l'isotropie de la texture augmente considérablement l'information sur celle-ci. Cependant une large classe de textures (dans l'environnement *naturel*) est homogène et seulement une partie est également isotropique. Si le système visuel dispose d'un moyen de tester la validité des hypothèses, comme

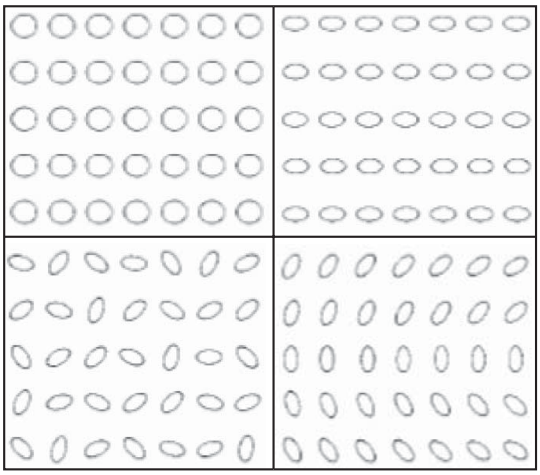


FIG. 4.7 – Résumé des différentes hypothèses proposées sur les statistiques de la texture pour la perception 3D (tirée de [Sto93]) ; première ligne de gauche à droite : texture isotropique ; texture anisotropique et homotropique ; deuxième ligne : texture isotropique et homotropique ; texture non-homotropique.

cela est le cas pour d'autres contraintes naturelles telles que la rigidité ou la symétrie, l'homogénéité, l'isotropie ou les deux hypothèses à la fois peuvent être utilisées pour percevoir la forme de surfaces texturées. L'identification de l'hypothèse utilisée par le système visuel est importante car cela donnerait des indications plus précises sur l'organisation des mécanismes corticaux sous-jacents participant à la vision 3D.

L'hypothèse d'homogénéité est très générale et ne donne aucune indication sur l'indice utilisé ni sur sa modélisation. Comme l'a montré Gibson, la seule contrainte imposée est l'analyse d'une texture composée d'une seul type de texel. Cette hypothèse s'applique à toutes les textures ainsi définies, même celles difficiles à modéliser. Par exemple les études de Todd *et al* [TOKK04] montrent que même sur des surfaces doublement incurvées possédant une texture anisotropique les sujets sont capables de percevoir correctement la forme de l'objet (Figure 4.8). Cette hypothèse est simplement trop faible pour en déduire des informations sur le type d'indice extrait.

Tester l'hypothèse d'homogénéité n'est pas directement possible car il faut faire intervenir des gradients de texture, eux-mêmes mal identifiés. Ainsi plusieurs études ont plutôt cherché à savoir si le système visuel est sensible à l'anisotropie induite par la projection de la texture (dans le cas où cette hypothèse est valide).

Todd et Akerstrom dans [TA87] ont construit des stimuli en prenant pour texels des rectangles variables répartis sur une sphère. Les sujets doivent juger de la profondeur des surfaces pour différentes conditions (Figure 4.9). Les auteurs observèrent que la profondeur est complètement éliminée si les texels ne sont pas suffisamment allongés ou s'ils ne sont pas approximativement alignés les uns par rapports aux autres orthogonalement à la direction du tilt, indépendamment de la compression. Ils conclurent que si la compression est importante, elle n'est cependant pas suffisante et les statistiques des orientations influencent également la perception. Comme le note Cumming *et al*, ces résultats pourraient s'expliquer par une prise en compte de l'isotropie. Cependant les auteurs concluent également, à partir de leurs résultats, que les observateurs ne perçoivent pas la forme des surfaces en attribuant localement

FIG. 4.8 – Exemples de surfaces doublement incurvées recouvertes d'une texture anisotropique (tirée de [TOKK04]) ; à gauche l'orientation de la texture suit une direction verticale constante ; à droite l'orientation de la texture suit la direction de la plus petite normale à la courbure ; la forme est parfaitement perceptible malgré l'anisotropie de la texture.

une valeur de profondeur à partir de la longueur des texels ou une orientation à partir de leur compression. Les auteurs se basent sur le fait qu'ils n'ont trouvé aucune différence significative entre des textures à base de texels identiques ou avec des tailles initiales différentes, ce qu'ils dénomment textures régulières et irrégulières. Cependant comme le note Rosenholtz dans [RM97], perceptuellement les deux types de texture apparaissent assez *irrégulières* et les textures dites régulières apparaissent déjà anisotropiques.

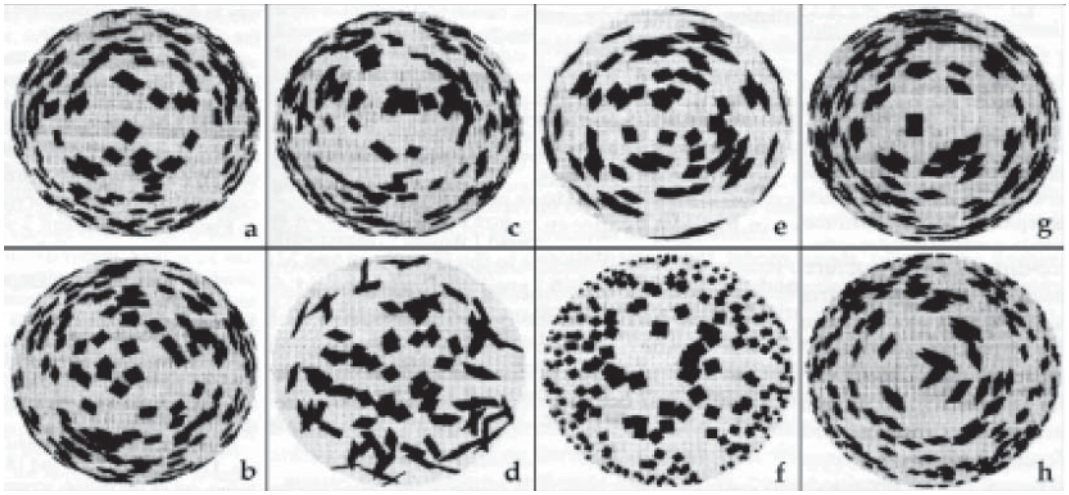


FIG. 4.9 – Différents stimuli utilisés par Todd et Akerstrom dans [TA87] montrant différentes caractéristiques importantes pour la perception de la courbure de la surface :(a) projection perspective ;(b) projection parallèle ;(c) formes irrégulières ;(d) orientations aléatoires ;(e) surface constante ;(f) sans compression ;(g) étiré ;(h) étiré sans compression.

Cumming *et al* [JCP93] ont utilisé des cylindres couverts d'une texture composée d'ellipses très étirées (anisotropiques). Lors d'une tâche de jugement de la courbure de cylindres , ils ont comparé les performances obtenues avec des textures formées de cercles ou d'ellipses orientées aléatoirement (isotropiques). Ils ont observé une diminution de la perception de la courbure des cylindres possédant déjà une forte anisotropie. Ils ont ainsi avancé l'idée que le système visuel

est sensible à l'anisotropie. Cependant, comme le note Rosenholtz *et al* [RM97], leur stimuli ne sont pas suffisants pour écarter d'autres explications telles qu'une texture anisotropique contienne simplement moins d'information et/ou que les éléments subissent moins d'effet de compression, entraînant une diminution du poids de cet indice (hypothèse d'homogénéité).

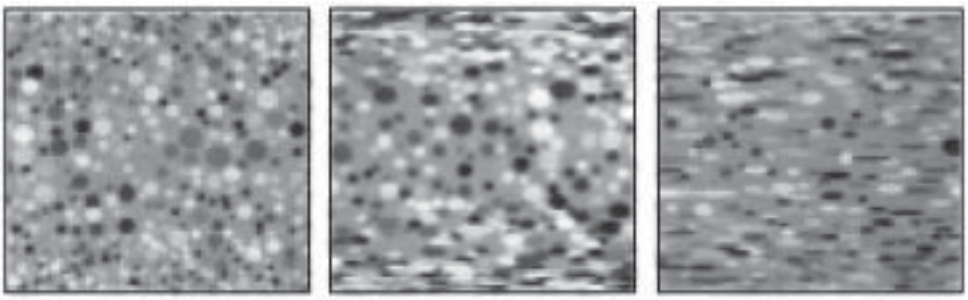


FIG. 4.10 – Exemples de stimuli utilisés par Cumming *et al* tirée de [JCP93] ; le cylindre de gauche est recouvert d'une texture composée de cercles ; sur le second cylindre, le gradient de compression a été accentué ; sur le dernier cylindre le même degré de compression est appliqué après avoir préalablement compressé l'ensemble des cercles dans la direction de la courbure (i.e en introduisant une anisotropie) ; ce dernier est jugé avoir une courbure moins importante que le deuxième cylindre.

Les travaux précédents de Todd et Akerstrom et de Cumming *et al* indiquent la possibilité de l'influence de la rupture d'isotropie comme indice de texture, leur stimuli ne permettent pas cependant de conclure définitivement.

Rosenholtz et Malik dans [RM97] ont construit des stimuli leur permettant de contrôler facilement la quantité d'anisotropie présente initialement dans la texture. Ces stimuli sont crées à partir de textures de Voronoi polygonales permettant d'obtenir une texture présentant des motifs irréguliers et positionnés aléatoirement sur la surface (évitant l'introduction d'autres indices comme la perspective linéaire). L'anisotropie est contrôlée par compression ou étirement préalable de la texture dans la direction du tilt (l'effet obtenu dans une direction différente est également étudié). La compression augmente la quantité d'anisotropie dans la texture après sa projection. Au contraire l'étirement diminue la quantité d'anisotropie présente dans la texture après projection. En adaptant une gauge sur la surface, les sujets doivent indiquer l'inclinaison qu'ils perçoivent pour les différents types de textures (isotropiques et anisotropiques). Si un biais apparaît dans les réponses par rapport aux cas où la texture est isotropique alors l'anisotropie est utilisée comme indice de texture sinon seuls les gradients de texture sont pris en compte.

Pour toutes les inclinaisons testées, une surestimation du slant est bien obtenue lorsque la texture est pré-compressée et une sous-estimation lorsqu'elle est étirée. Ce biais signifie que les sujets utilisent bien l'information de déviation par rapport à l'isotropie pour estimer l'inclinaison de surfaces planes. Cependant le biais obtenu est inférieur à celui prédit par une estimation uniquement basée sur la mesure de l'anisotropie. Le système visuel utilise donc également d'autres indices tels que les gradients de texture. Ces résultats permettent néanmoins de rejeter l'hypothèse d'une estimation uniquement basée sur les gradients de texture ou uniquement basée sur la déviation par rapport à l'isotropie. Les auteurs optent pour une combinaison des deux hypothèses (à la manière d'une combinaison d'indices). Cependant Rosenholtz et Malik observent également des variations importantes dans les estimations

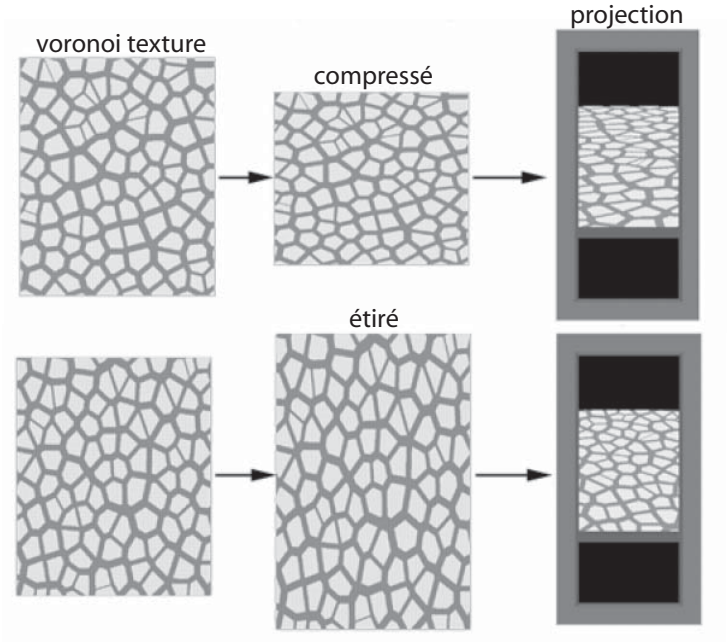


FIG. 4.11 – Exemple de textures de Voronoi utilisées par Rosenholtz et Malik dans [RM97] ; deux déformations sont réalisées conduisant à une texture compressée ou étirée dans la direction du tilt, modifiant ainsi la déviation par rapport à l'isotropie ; une surestimation de l'inclinaison est obtenue sur la texture compressée.

lorsque les surfaces sont courbes, selon l'anisotropie initiale de la texture, sa régularité ou encore selon la largeur du champ visuel.

Knill dans [Kni98c] [Kni98a] [Kni98b] a réutilisé les stimuli de Rosenholtz et Malik pour tester l'influence de l'isotropie sur l'indice de compression. Pour cela il a comparé les performances d'un observateur idéal utilisant l'indice de compression avec et sans l'hypothèse d'isotropie. Il observa que, sans hypothèse d'isotropie, les performances de l'observateur idéal diminuent d'un facteur 4 montrant ainsi que l'isotropie peut être un indice important pour juger la 3D. Dans [Kni98a], une des expériences consiste à manipuler la fiabilité des indices d'échelle et de compression. Cela est fait en manipulant indépendamment les variances des longueurs des texels (pour l'indice d'échelle) et des formes des texels (pour l'indice de compression). L'auteur compare ensuite les performances des sujets avec un observateur idéal uniquement basé sur l'indice de compression. Les résultats montrent clairement que les performances des sujets dépassent celles de l'observateur idéal n'intégrant pas d'hypothèse d'isotropie, ce qui permet d'affirmer que les sujets intègrent une autre information en plus de l'indice de compression.

L'importance (théorique) de l'hypothèse d'isotropie et le fait que toutes les textures ne soiet pas isotropiques amène à penser que le système visuel pourrait avantageusement combiner les deux hypothèses. Ainsi Knill et Rosenholtz et Malik envisagent un système effectuant une combinaison (complexe) des deux hypothèses : par défaut l'hypothèse d'homogénéité est utilisée, si une rupture d'isotropie est décelable (dans le cas où une isotropie initiale de la texture est identifiable) alors cette contrainte peut également s'appliquer pour augmenter la précision de l'estimation. Ce modèle soulève cependant la question de savoir si le système

visuel peut effectivement déterminer dynamiquement l'applicabilité de contraintes telles que l'isotropie à partir des informations contenues dans l'image.

Le problème initial de déterminer si le système visuel effectue une analyse locale ou globale de la surface pour en déterminer son orientation et sa forme semble avoir obtenu des réponses par les études sur l'influence de la taille du champ visuel (Chapitre 4.3). Les résultats mettent en évidence qu'une analyse dans une zone trop étroite diminue les performances. Au contraire il semble plutôt qu'une analyse mettant en jeu une intégration de l'information sur une région suffisamment grande soit nécessaire.

Si aucune conclusion n'est encore permise il est cependant possible de relever que l'hypothèse d'isotropie n'est pas suffisante pour rendre compte de l'ensemble des résultats et son association avec d'autres indices doit être envisagée. Au contraire l'hypothèse d'homogénéité est très générale mais repose sur l'identification d'un gradient de texture. Une approche va donc consister à chercher les gradients de texture susceptibles de rendre le mieux compte des performances visuelles des sujets en supposant l'homogénéité de la surface et d'en analyser les combinaisons possible.

4.3 Caractéristiques de la perception 3D

Les différents travaux basés sur l'étude des gradients de texture ont permis de tracer les caractéristiques principales et les limites de la perception des formes à partir de l'information de texture. Celles-ci peuvent se diviser en 4 types d'études analysant : les configurations géométriques préférentielles, l'influence de la régularité de la texture, l'influence du champ visuel et les gradients de texture. Nous allons décrire chacune de ces caractéristiques qui représentent les connaissances accumulées sur la perception 3D à partir de la texture.

4.3.1 Configurations géométriques préférentielles

Dès les premières recherches menées par Gibson, il a été observé que les performances d'estimation et de discrimination de surfaces inclinées varient avec les valeurs du tilt et du slant.

Pour le tilt, de très bonne performances sont obtenues pour des valeurs autour de 90^o correspondant à des surfaces de sol. Gibson [Gib79] suggéra l'existence d'un mécanisme spécialisé adapté à ce type particulier d'orientation qui serait lié au déplacement (i.e lié à l'évolution). De bonnes performances sont aussi obtenues pour des valeurs autour de 0^o correspondant par exemple à une paroi de mur verticale. Les données physiologique peuvent aussi expliquer en partie ces performances du fait d'une plus grande distribution des cellules corticales autour de ces deux orientations avec une distribution minimale autour de 45^o.

Les performances s'améliorent également avec l'augmentation du slant. Knill dans [Kni98c] dérive un observateur idéal et montre que la variance de l'estimateur diminue quelque soit le gradient de texture montrant que théoriquement l'information portée par les gradients augmente la précision de l'estimation de l'inclinaison plus celle-ci est importante. Dans [Kni98a], les sujets exhibent également cette non-linéarité dans leurs réponses (pour un slant $> 30^o$) et les courbes sont parallèles à celles des observateurs idéaux. Ces résultats répliquent déjà ceux obtenus par Blake *et al* dans [BBS93], ce qui montre bien que plus le slant augmente, plus l'information de texture est fiable et plus la perception 3D est précise.

4.3.2 Influence de la régularité de la texture

Différents travaux ont cherché à déterminer les caractéristiques de la texture influençant l'estimation de la 3D. Dans cette étude, il s'agit de caractériser à la fois les performances du système visuel mais également de déterminer quelles sont les caractéristiques qui font qu'une texture facilite ou non la perception 3D. Cette analyse est rendue difficile par le fait qu'elle dépend du modèle pris pour caractériser la texture (voir Chapitre 2.2).

Pour étudier les performances des sujets, Gibson a utilisé une tâche bimodale combinant la perception visuelle et la perception sensori-motrice consistant à faire incliner aux sujets un panneau représentant la surface plane [Gib50b]. Gibson observe ainsi qu'en l'absence d'autres indices, l'inclinaison de la surface est sous-estimée. Cette erreur systématique a été confirmée plus tard par d'autres auteurs notamment Braunstein [Bra68]. Gibson observe également que cet effet augmente avec ce qu'il appelle *l'irrégularité* de la texture. La figure 4.12 reproduit les textures présentant les irrégularités définies par Gibson [FM64].

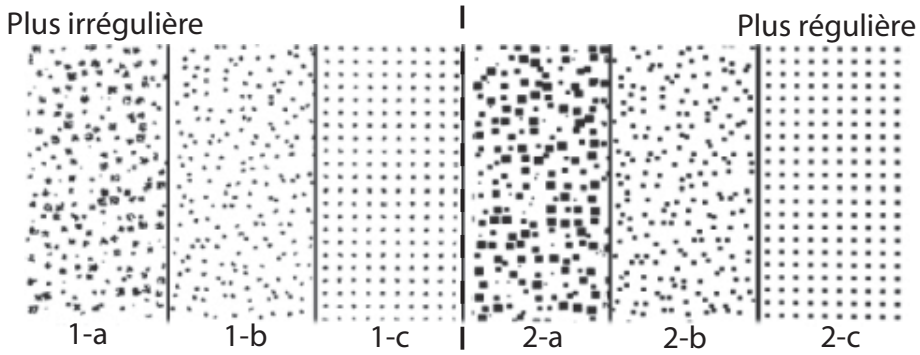


FIG. 4.12 – Exemples d'irrégularités définies par Gibson [FM64] : les irrégularités de forme (les texels peuvent être de formes différentes) (1-a, 1-b, 1-c) ; les irrégularités de taille (texels de tailles différentes) (1-a, 2-a), les irrégularités de position (non-alignement) (1-b, 2-b) ; 2-c représente la texture la plus régulière où l'ensemble des texels sont des carrés de même taille et alignés ; par opposition 1-a est la texture la plus irrégulière au sens de Gibson.

Turner *et al* dans [TGB91] ont tenté de définir cette irrégularité dans le domaine fréquentiel pour obtenir une description plus précise de la texture que celle de Gibson. Ils indiquent ainsi qu'elle peut être caractérisée par un spectre *large* c'est-à-dire un spectre comportant plusieurs composantes fréquentielles (par exemple un quadrillage serait une texture régulière tandis qu'une texture d'osier serait irrégulière).

Rosas *et al* ont adopté une approche exploratoire afin de tester la qualité de différentes textures [RWW04]. La figure 4.13 reproduit les différentes textures utilisées. Elles ont été choisies afin de couvrir un large ensemble de caractéristiques statistiques avec notamment la présence ou non de texels clairement identifiables. Lors d'une tâche de discrimination entre deux surfaces planes inclinées, suivant différentes valeurs de slant, les auteurs observent une influence importante de la texture sur les performances de discrimination. Leurs résultats montrent également que plus l'inclinaison est forte, plus les performances sur l'ensemble des textures augmentent et les différences entre les textures s'estompent, conformément aux précédents résultats sur la nonlinéarité des performances avec le slant. Les auteurs établissent à partir des différences obtenues entre les textures un ordre reflétant la qualité de l'information pour réaliser la tâche de discrimination (*facilitation*). La figure 4.13 reproduit l'ordre des

textures selon leur facilitation respective : la texture composée de points Polka permet les meilleurs performances de discrimination tandis que les textures proches du bruit donnent les plus faibles.

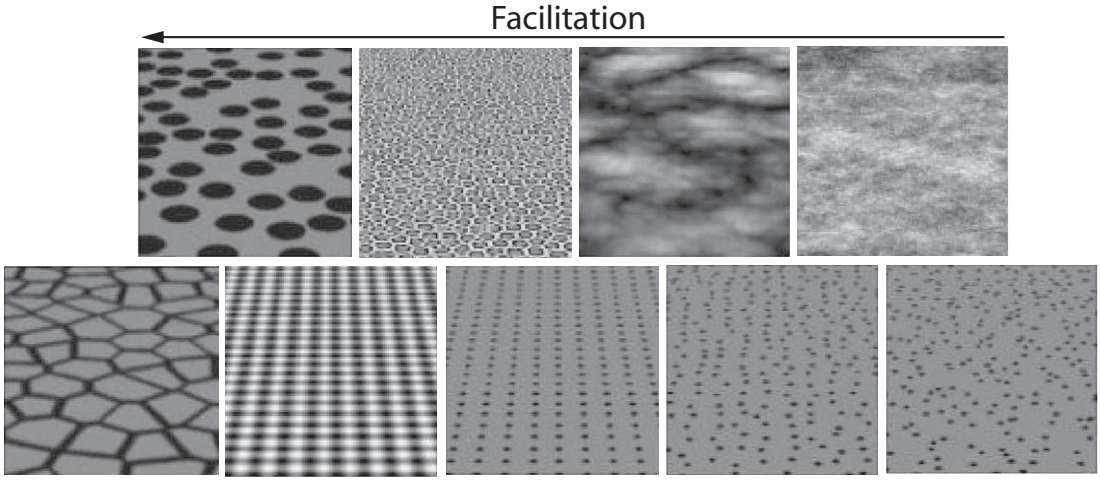


FIG. 4.13 – Textures utilisées par Rosas *et al* (tiré de [RWW04]) ; la première ligne reproduit l'ordre de facilitation des textures pour la discrimination du slant trouvés par les auteurs : points Polka, texture de léopard, bruit cohérent et bruit en $1/f$; la seconde ligne reporte les autres textures utilisées : texture de Voronoi, réseaux, treillis avec des points répartis de plus en plus irrégulièrement.

4.3.3 Influence du champ visuel

Blake *et al* [BBS93] furent parmis les premiers à mesurer spécifiquement l'influence la taille du champ visuel sur la surface sur la fiabilité de l'indice de texture. Ils simulèrent un observateur idéal sur des textures où le nombre de texels reste fixe tout en modifiant l'angle d'ouverture. Les auteurs montrent ainsi l'influence de l'ouverture du champ visuel sur les performances des sujets : l'indice de densité est le plus fiable pour de larges ouvertures ($> 20^o$) tandis que l'indice de compression reste également fiable pour de petites ouvertures ($< 20^o$).

Knill dans [Kni98a] mesure les seuils de discrimination entre des surfaces inclinées pour différents champs visuels. La figure 4.14 reproduit la texture formée d'ellipses aléatoires et les différentes configurations étudiées d'ouverture verticales : symétrie par rapport à la verticale centrale ; ouvertures horizontales située en haut et en bas (pour un tilt à 90^o). Le nombre de texels est maintenu constant pour les différentes ouvertures pour maintenir le même nombre d'estimations indépendantes locales si cette stratégie est employée. Les résultats montrent qu'une ouverture verticale permet d'obtenir toujours de bonnes performances ; une ouverture horizontale située en haut diminue la fiabilité des gradients de texture (taille et compression [Kni98c]) mais la discrimination est toujours possible ; les moins bonnes performances sont obtenues pour une ouverture horizontale située en bas (proche de l'observateur). Ces résultats montrent la nécessité d'une ouverture du champ visuel suffisante pour pouvoir intégrer une information sur une région suffisamment large dans la direction du tilt (i.e direction correspondant aux gradients maximums) ce qui conduit à supposer que le système visuel n'utilise

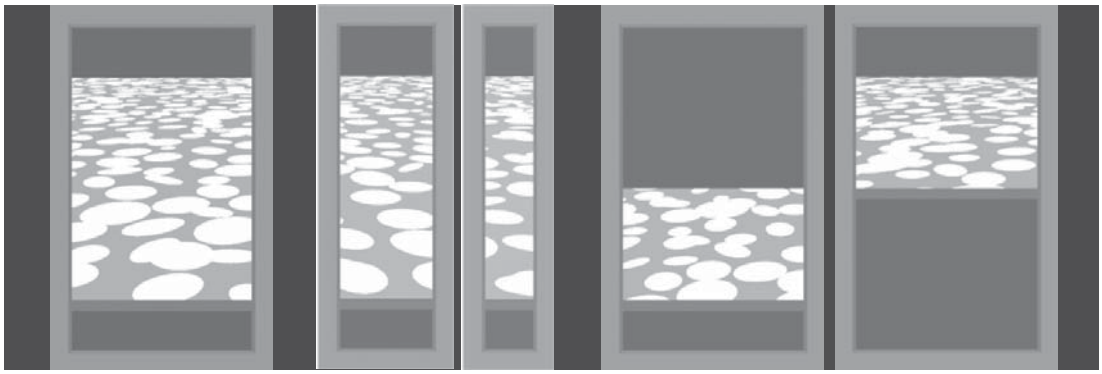


FIG. 4.14 – Exemples de stimuli où le champ visuel a été modifié (tiré de [Kni98a]) ; de gauche à droite : stimulus initial ; réductions du champ visuel horizontal ; réductions du champ visuel vertical (en bas et en haut de la texture initiale).

que faiblement une analyse locale et utilise plutôt les gradients de texture pour estimer l'inclinaison des surfaces. L'auteur observe également que la fiabilité des indices augmentent avec la densité des texels (meilleurs performances dans les zones éloignées de l'observateur) ce qui augmente l'information portée par la texture. La région contenue dans le champ visuel doit donc être suffisamment large dans la direction du tilt et contenir suffisamment d'information de texture pour estimer les gradients associés. Malgré ces résultats Knill indique que la comparaison avec un observateur idéal basé uniquement sur les indices de densité et de compression montre que les sujets n'utilisent pas spécifiquement ces indices, ou au mieux une combinaison des deux.

Todd *et al* dans [TTD05] ont montré également l'influence combinée du champ de vision et des caractéristiques de la texture. Leur stimuli sont composées de deux plans présentés sous différents angles d'ouverture du champ visuel avec un slant constant. Les sujets indiquent la forme (concave ou convexe) qu'ils perçoivent et réalisent une tâche d'ajustement pour estimer la profondeur perçue.

La figure 4.15 montre les différentes textures utilisées, dans l'ordre de difficulté de la perception 3D. Les résultats obtenus montrent que pour une ouverture large, la forme est toujours relativement bien perceptible, tandis qu'elle diminue jusqu'à disparaître pour des ouvertures faibles. Les auteurs concluent également à l'utilisation des gradients de texture plutôt qu'à une analyse locale (par exemple la rupture de l'isotropie) et indiquent l'importance des informations d'orientation pour percevoir la 3D.

4.3.4 Les gradients de texture

L'ensemble des résultats précédents montre l'implication et l'importance des gradients de texture pour la perception 3D, comme suggéré par Gibson. Cependant trouver le ou les gradients de texture utilisés par le système visuel représente un véritable enjeu pour comprendre son fonctionnement interne.

Cutting and Millard [CM84] ont étudié les indices de taille, de densité et de compression lors d'une tâche de différenciation entre une surface *plane* et une surface *courbe*. Pour cela ils créèrent des stimuli à base d'octogones leur permettant de manipuler chaque gradient indépendamment. En utilisant un paradigme de conflit d'indices, les résultats montrent l'indice de

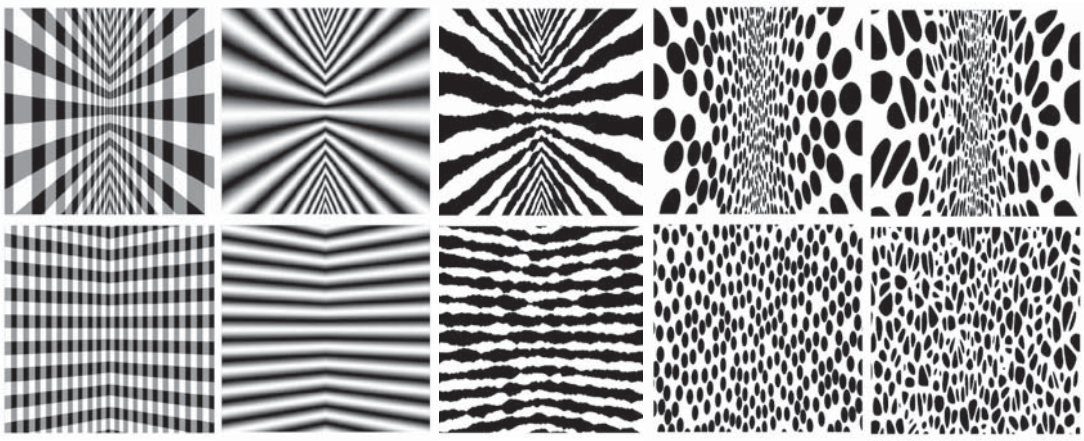


FIG. 4.15 – Ensemble des textures utilisées par Todd *et al* (tiré de [TTD05]) ; toutes les surfaces sont concaves ; de gauche à droite : réseaux ; contours réguliers ; contours irréguliers ; texels réguliers ; texels irréguliers ; les images de la première ligne sont affichées sous un champ visuel de 60^o, celles du bas, avec un champ visuel de 10^o.

taille est celui qui contribue le plus à la perception d'une surface plane. Les performances diminuent lorsque cet indice est incorrect et cela même en présence des autres indices. Le gradient de compression est l'indice qui contribue à l'essentiel de la perception des surfaces courbes. Cependant, comme le note Knill dans [Kni98a], l'importance minimale de la compression est peut-être due à la tâche qui ne mesure pas directement la précision de la perception 3D.

Todd et Akerstrom [TA87], sur leur stimuli à base de sphères 4.9, conclurent que si la compression est importante pour la perception de la forme, il est nécessaire de rajouter une contrainte d'alignement.

Blake *et al* [BBS93] furent parmis les premiers à développer un modèle d'observateur idéal permettant de comparer directement les performances des sujets avec les performances théoriques optimales obtenues en s'appuyant sur des indices spécifiques. Les auteurs ont ainsi étudiés les indices de densité et de compression sur une tâche d'estimation de la courbure de cylindres texturés. La texture est représentée par des segments de droites, orientés aléatoirement, dont la longueur est modifiée. Les résultats montrent que les sujets ont des performances supérieures à celles qu'ils devraient obtenir uniquement en se basant sur l'indice de densité. Les auteurs concluent ainsi que le système visuel doit combiner d'autres indices, tel que l'indice de compression. Cependant il est à noter que si ces résultats montrent que la densité n'est pas suffisante, ils ne prouvent pas non plus qu'elle soit réellement utilisée.

Knill a conduit une étude relativement complète et systématique sur l'importance relative de la densité, du changement de taille et de la compression. Il a développé pour cela un observateur idéal associé à chaque indice [Kni98c]. Il a comparé ses performances théoriques avec les résultats obtenus sur des sujets [Kni98a]. Finalement il a étudié la mise en conflit des indices [Kni98b]. La figure 4.16 montre les stimuli utilisés dans ces trois études composés de textures à base d'ellipses aléatoires et à base de pavage de Voronoi.

Knill observe tout d'abord que la texture donne une indication plus fiable pour le slant que pour le tilt, ce qui indiquerait une première dichotomie pour l'estimation des deux angles, confirmant les résultats de Rosenholtz et Malik dans [RM94]. Pour les deux types de texture

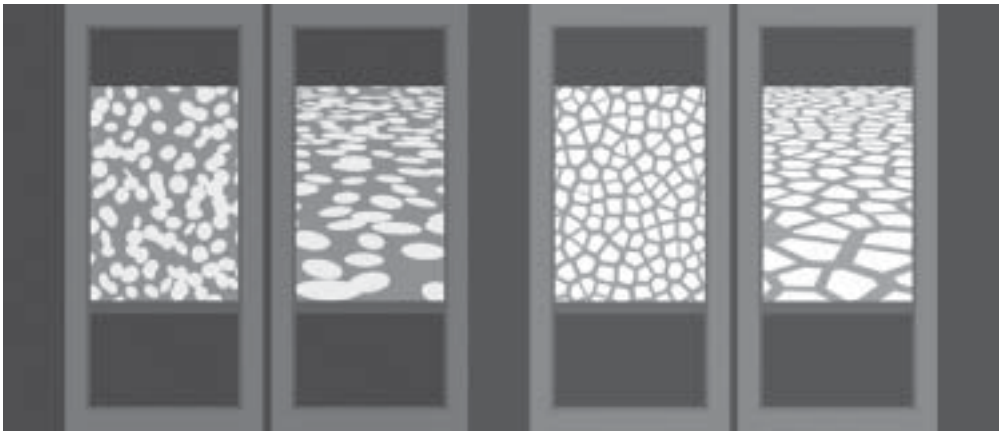


FIG. 4.16 – Stimuli utilisés par Knill (tiré de [Kni98c] [Kni98a] [Kni98b]) ; à gauche, texture à base d'ellipses aléatoires ; à droite, pavage de Voronoi correspondant à une texture d'aspect plus *naturelle* tout en partageant les mêmes statistiques que les ellipses.

de la figure 4.16, l'indice de compression est le plus fiable, suivi par l'indice de taille et enfin celui de densité.

Ces résultats sont en conflit avec les résultats précédents sur la prépondérance de l'information de taille. Cela est certainement dû à l'utilisation par Blake *et al* de textures planes régulières où apparaît l'information de perspective linéaire, un indice qui n'est pas présent dans les texture irrégulières (définies statistiquement).

Knill remarque finalement que l'ensemble des gradients de texture définis précédemment représente une décomposition particulière de l'information de texture. Ainsi il ne faut peut être pas dissocier l'information de compression et de taille. Cette décomposition est naturelle au sens où elle suit la définition des différents indices de texture. Ainsi par exemple les stimuli utilisés par Todd et Akerstrom (Figure 4.9) montrent qu'il est possible de manipuler l'information de compression tout en conservant l'indice de taille constante (et inversement). Cependant il est possible de modéliser la texture en confondant la compression et la taille. Ainsi Malik et Rosenholtz dans [MR97] par exemple caractérise la texture comme une relation affine entre des régions voisines de l'image en se basant sur le specre d'amplitude local de la texture. Ainsi il est possible d'envisager de changer d'espace de représentation (par exemple l'espace de Fourier) pour pouvoir analyser l'information contenue dans la texture au delà des gradients *classiques* de texture.

4.4 Gradient de fréquence et perspective linéaire

Cette section présente une autre approche de la caractérisation de la texture basée sur les informations de variation de fréquence et de variation d'orientation (la perspective linéaire). Il décrit les études réalisées sur ces deux informations de texture et leur lien avec les gradients de texture décrits précédemment. Ces indices sont également mis en relation avec les travaux de Li et Zaidi. Les auteurs proposent un modèle permettant la caractérisation spectrale de la texture pour la perception 3D. Ces travaux ont servi de point de départ à nos expériences (Chapitre 5) et à notre modèle d'extraction de la forme par la texture basé sur l'analyse de la fréquence (Chapitre 6).

4.4.1 Gradient de fréquence

La notion de fréquence spatiale dans une image n'est pas une information intuitive. La figure 4.17 montre un stimulus simple fabriqué à partir d'une seule composante fréquentielle. La variation entre les zones de grande intensité (blanc) et de faible intensité (noir) permet de caractériser la texture par une fréquence. Une variation rapide (spatialement) correspond à une haute fréquence. Une variation lente correspond à une basse fréquence. Cette fréquence est constante lorsque la texture est projetée sur une surface plane (Figure 4.17(a)). Projetée sur un cylindre, cette fréquence subit une variation proportionnelle à l'inclinaison locale de la texture (Figure 4.17(b)). Ceci permet d'obtenir à la fois une caractérisation de la texture (fréquence moyenne) et une mesure de la déformation de la texture par l'étude de sa variation (Figure 4.17(c)). Nous parlerons ainsi de gradient de fréquence comme d'un indice de la perception 3D à partir de la texture de manière similaire aux gradients de texture introduits à la section 4.1.

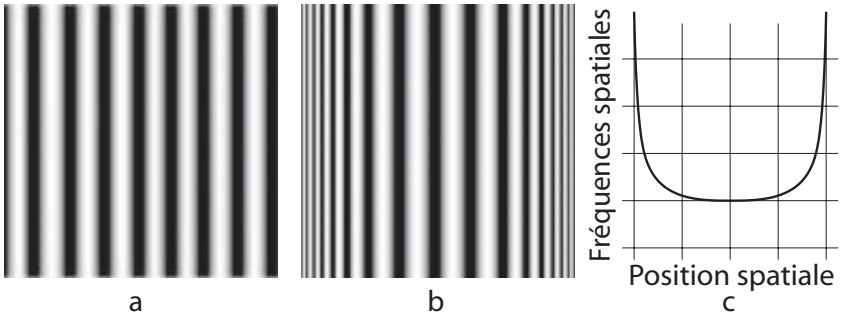


FIG. 4.17 – Texture sinusoidale ; (a) surface plane ; (b) projection sur un cylindre ; (c) traduction de la déformation subie par la texture dans le domaine fréquentiel.

Différents modèles ont été développés en vision par ordinateur en se basant sur l'analyse des composantes fréquentielles des textures (Section 2.4). Ainsi Malik et Rosenholtz [MR97] ont introduit la notion de *distortion affine de la texture*, qui est très proche de la notion de gradient de texture. Le changement local de texture est modélisé comme une transformation affine locale ce qui présente l'avantage de *résumer* l'ensemble des gradients de texture et de contenir assez d'information pour retrouver l'orientation et la forme de la surface. Ce modèle intègre un calcul de l'intervalle de confiance dans l'estimation de la forme d'une surface permettant d'obtenir un observateur idéal afin de prédire les réponses de sujets humains sur la même tâche.

L'analyse de la fréquence présente également l'intérêt d'être plus proche du fonctionnement du système visuel comme décrit au chapitre 3. En effet d'après les connaissances acquises en neurophysiologie sur la structure du système visuel humain, il est relativement improbable que les éléments de texture puissent être analysés en les comptant au sein d'une région locale ou en mesurant précisément leur longueur, largeur ou compression. La décomposition de l'information visuelle en bande de fréquence et en orientations par les cellules simples et complexes laisserait plutôt envisager l'analyse des gradients de texture dans le domaine de Fourier (par exemple suivant les modèles de Malik et Rosenholtz [MR97] ou de Sakai et Finkel [SF95]) ou par des filtres spatiofréquentiels (par exemple suivant le modèle de Clerc et Mallat [CM02] ou celui que nous proposons (voir Chapitre 6)).

Peu de travaux ont étudiés spécifiquement l'indice de variation de fréquence. Nous citerons ainsi les travaux de Prins et Kingdom [PK02] basés sur l'utilisation de stimuli composés de masques de Gabor dont l'orientation et la fréquence sont contrôlés de manière à obtenir l'impression d'une surface avec une courbure sinusoïdale (Figure 4.18). Cependant les masques apparaissent comme des éléments individuels posés sur la surface. La variation de fréquence n'est donc pas continue et l'information de forme peut être aussi bien portée par un gradient de texture tel que la variation de taille du masque de Gabor. De plus les deux variations n'apparaissent pas sur la même texture ne permettant pas d'étudier leur contribution relative.

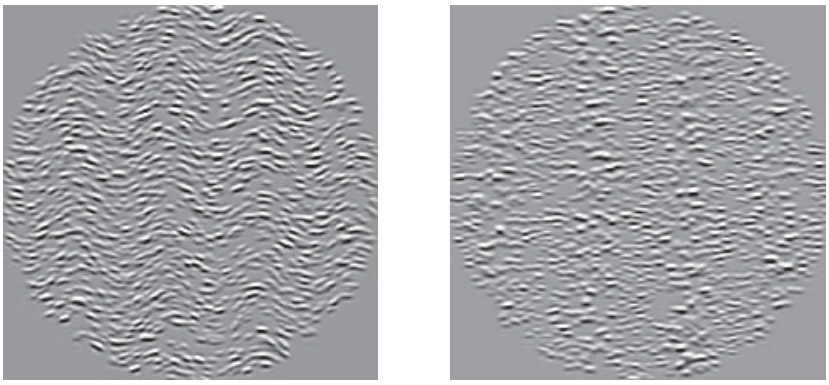


FIG. 4.18 – Exemples de stimuli utilisés par Prins et Kingdom (tiré de [PK02]) ; chaque texture est composée d'un ensemble de masques de Gabor ; à gauche : application d'une variation des orientations correspondant à une surface sinusoïdale ; à droite : application d'une variation de fréquence correspondant à la même surface ; seule la texture de gauche transmet bien une impression de courbure sinusoïdale.

4.4.2 Perspective linéaire

La perspective linéaire correspond à l'effet obtenu lors de la projection (perspective) de lignes parallèles sur le plan de l'image. Les lignes tendent à converger vers un point unique, nommé le *point de fuite*, et deviennent des *lignes de fuite* (Figure 4.19).

Différents travaux ont étudiés la perspective linéaire comme indice de la perception 3D et sa combinaison avec les autres indices de texture ([TTD05], [ABS98], [OML03]).

Todd *et al* [TTD05] observèrent une amélioration significative de l'estimation de l'inclinaison de surfaces lorsque celles-ci sont recouvertes d'une texture à carreaux (*plaid*) par rapport à une texture sans aucun alignement (texture isotropique) (Figure 4.15, notamment les deux textures de droite de la deuxième ligne ne transmettent pas d'impression de surface concave).

Andersen *et al* [ABS98] étudièrent la perception de la profondeur et de l'inclinaison (slant) dans des scènes crées artificiellement. Ils utilisèrent des textures composées de grilles pour analyser l'utilisation des indices de compression et de perspective linéaire. Ils conclurent que la perspective linéaire est bien utilisée pour estimer la profondeur d'une scène. Cependant l'indice de compression est dans certain cas plus fiable notamment pour des surfaces fortement inclinées comme des surfaces de sol.

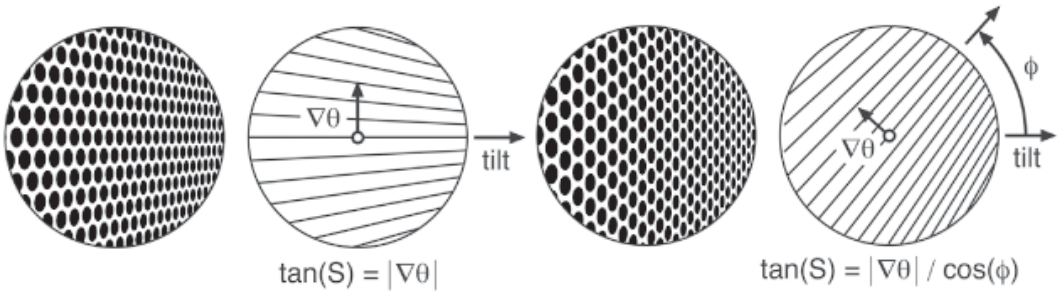


FIG. 4.19 – Exemple de perspective linéaire (tiré de [SBft]) ; à gauche : surfaces texturées inclinées ; à droite : exemples de lignes de fuite associées aux exemples ; la quantité de convergence ($V0$ et VH) permet de retrouver la valeur du slant (S) de la surface connaissant l'angle de roll (ϕ).

Oruc *et al* [OML03] ont étudié la corrélation des gradients de texture et de la perspective linéaire afin d'analyser l'influence relative de ces deux indices dans une tâche d'estimation d'une surface plane inclinée. La figure 4.20 présente les différents stimuli en présence de chaque indice séparés et en combinaison. Les deux indices participent bien à l'estimation de la surface et ils se combinent de manière optimale en fonction de la fiabilité des indices (indiqué par la densité des lignes ou des diamants ; une densité faible conduit à une diminution de la fiabilité de l'indice).

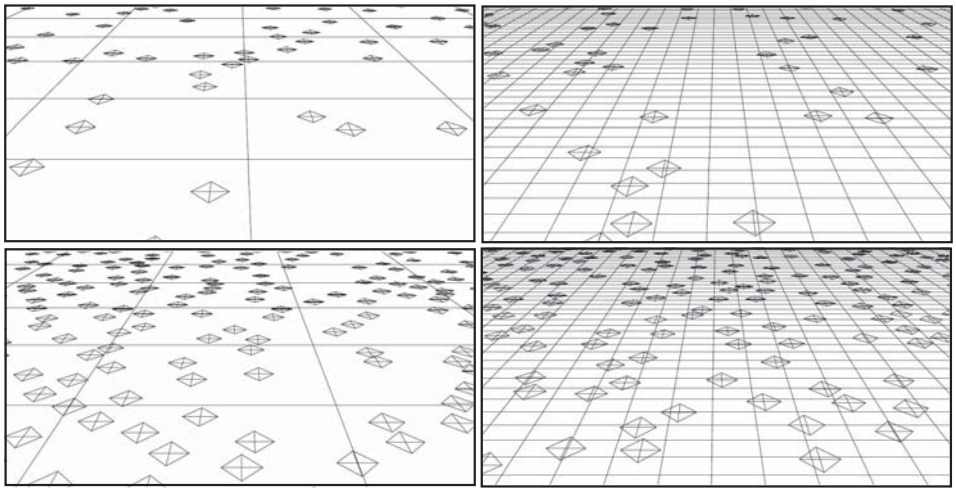


FIG. 4.20 – Exemples de stimuli utilisés par Oruc *et al* dans [OML03] ; chaque stimulus présente une condition différente de combinaison des indices de perspective linéaire (indiquée par le quadrillage formé de lignes horizontales et verticales) et de texture (indiqué par les diamants répartis de aléatoirement sur la surface) ; les différences de densité représentent permettent de faire varier la fiabilité de chaque indice.

Cependant il est à noter que dans les travaux précédemment décrits, la texture utilisée représente souvent un quadrillage qui contient à la fois un indice de perspective linéaire et un gradient de texture (par exemple changement de taille des carreaux ou une variation de fréquence). Ainsi dans ces expériences, la perspective linéaire et les gradients de texture ne

sont pas séparés et présentent une forte corrélation. Ainsi les stimuli utilisés dans ces études ne permettent pas d'évaluer la contribution relative de la perspective linéaire seule face aux autres indices de texture.

4.4.3 Modèle spectral d'extraction de la forme par la texture

Li et Zaidi ont proposé une approche originale basée sur l'analyse du spectre d'amplitude global de la texture ([LZ00] [LZ01c] [LZ01b] [LZ01a] [LZ03] [LZ04]). Leur modèle conduit à une description plus précise de l'information contenue dans la texture permettant de transmettre l'information de forme de la surface sous-jacente.

Les auteurs ont mesurés les performances de sujets à estimer la courbure 3D relative le long d'une surface texturée. Ils ont ensuite analysé les différences entre les spectres d'amplitude de plusieurs textures présentant des motifs différents. Ils ont observé que la présence d'énergie dans le spectre localisée à l'orientation correspondant à la direction de la ligne de courbure maximale de la surface est crucial pour transmettre l'information de forme. D'après le modèle de Malik et Rosenholtz [MR97], la présence d'un pic d'énergie dans le spectre à une orientation correspondant au maximum de variation d'inclinaison de la surface est suffisante pour retrouver l'orientation d'imagettes locales (i.e l'orientation de petites régions locales supposées planes).

Pour pouvoir effectuer cette analyse, les auteurs ont crées différents types de texture à partir de réseaux, de bruit filtré, en manipulant directement le spectre d'amplitude [LZ00] et à partir de textures naturelles de Brodatz [LZ01c] (Figure 4.21 et Figure4.22). Les textures sont ensuite projetées sur des surfaces ondulées [LZ00] [LZ01c] [LZ01b] [LZ01a], sur des surfaces courbes développables [LZ03] ou sont déformées (par étirement ou creusement créant de inhomogénéités locales) [LZ04] suivant une ou deux directions spatiales. La courbure de la surface perçue par les sujets est reconstruite par une succession de mesures de la profondeur relative locale suivant une direction (Figure 4.21).

D'après leurs résultats, les auteurs concluent : qu'il n'est ni nécessaire ni suffisant d'identifier les éléments de texture individuels ou les gradients de texture pour extraire la forme de la surface ; une variation de fréquence en une seule dimension est insuffisante pour transmettre une information de forme complexe ; une bonne perception de la profondeur n'est perçue que lorsque la texture projetée contient de l'énergie dans le spectre localisée à l'orientation correspondant à la direction de la ligne de courbure maximale de la surface ; la présence de cette composante spectrale crée un motif dans la texture faisant apparaître un effet de perspective linéaire uniquement dans le cas d'une projection perspective (la texture est donc un indice de forme uniquement pour ce type de projection) ; seules certaines textures naturelles peuvent transmettre une information correcte de forme, ce qui peut être prédit à partir des caractéristiques du spectre d'amplitude global.

Une autre manière d'interpréter l'hypothèse de Li et Zaidi est de considérer l'utilisation par le système visuel d'un indice de variation de fréquence et d'un indice de perspective linéaire. Comme le suggère les auteurs, un modèle analysant séparément ces deux indices peut permettre de reproduire les résultats obtenus sur leur texture.

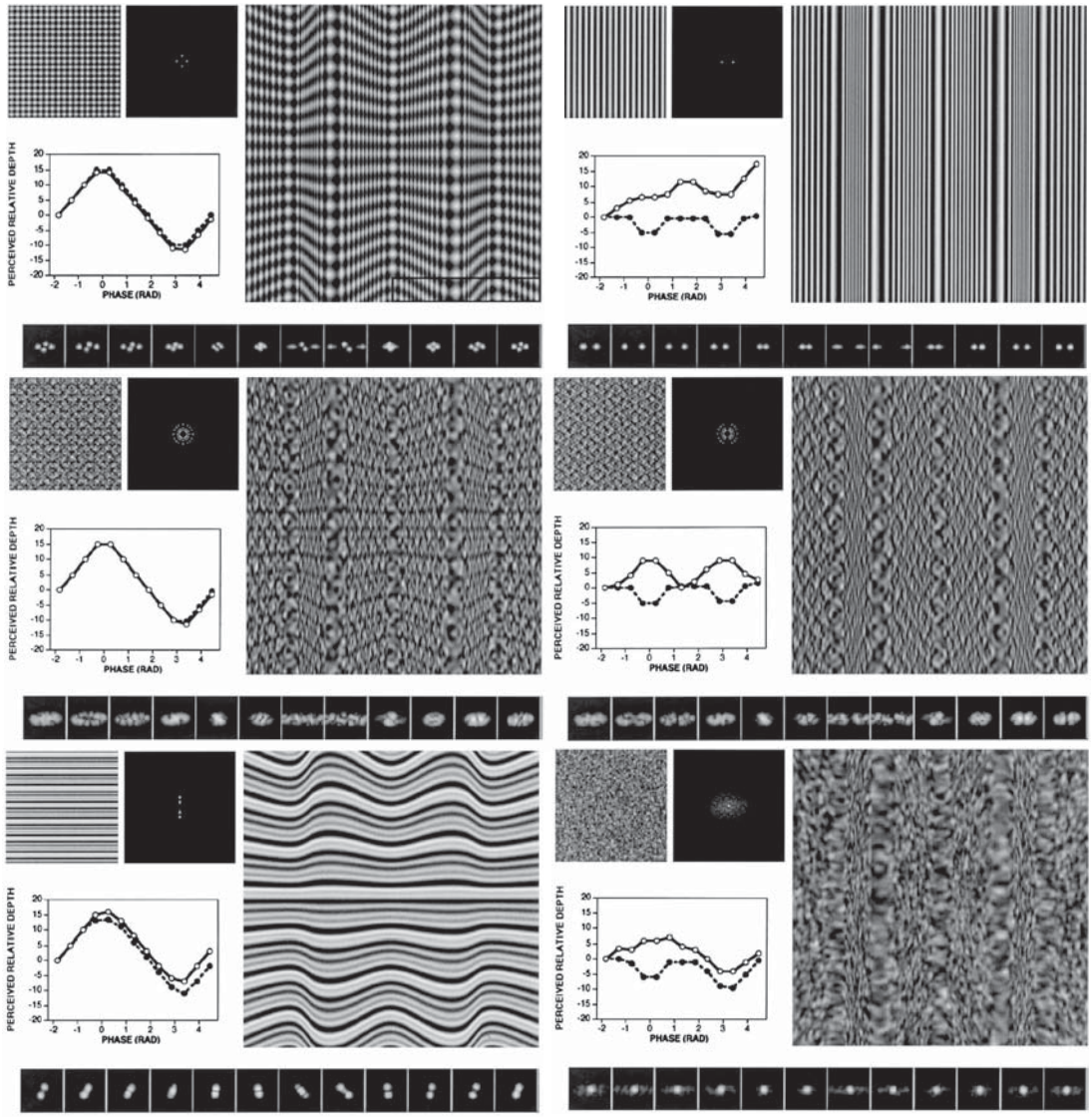


FIG. 4.21 – Exemples de texture utilisées par Li et Zaidi (tiré de [LZ00]) ; pour chaque image : en haut de gauche à droite : texture initiale, spectre d'amplitude global de la texture ; texture après projection sur une surface ondulée ; en bas à droite : profondeur relative perçue par deux sujets sur ligne correspondant à un période d'ondulation de la surface ; en bas : spectres locaux d'imagettes de taille 32X32 pixels partant du centre de l'image jusqu'à son extrémité droite ; les textures utilisées sont, de haut en bas et de gauche à droite : un réseau horizontal et vertical ; un réseau vertical ; une texture octotropique (formée de 8 composantes d'énergie réparties sur les orientations du spectre) ; la même texture octotropique moins la composante horizontale (correspondant aux éléments horizontaux dans la texture) ; la composante horizontale de la texture précédente ; un bruit isotropique ; la colonne de gauche présente les textures où la perception de l'inclinaison est bonne ; la colonne de droite présente les textures où la perception de l'inclinaison est mauvaise dû au manque d'une composante d'énergie dans la direction de l'inclinaison (tilt) dans le spectre global de la texture.

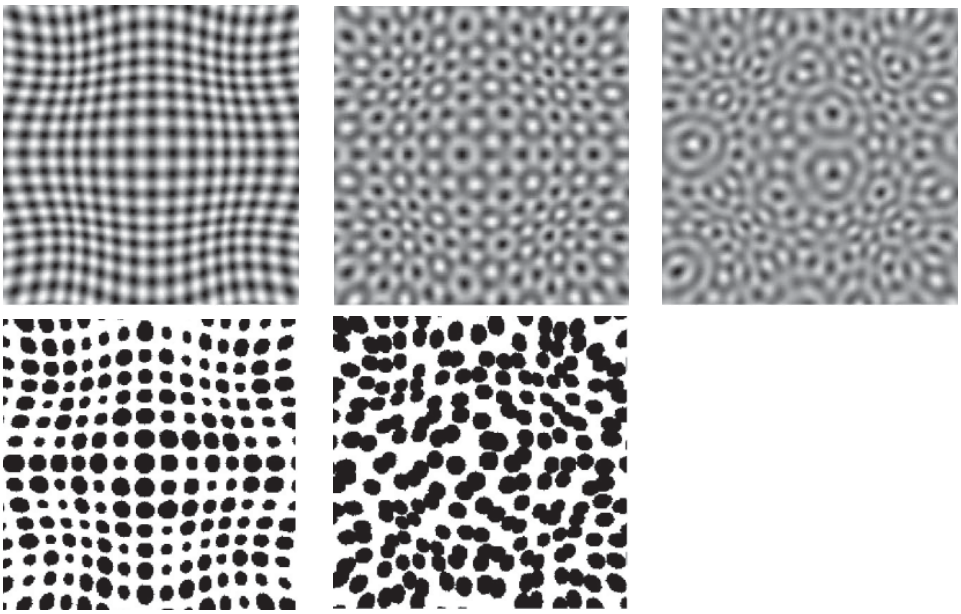


FIG. 4.22 – Exemples de textures déformées par étirement ou creusement (créant de inhomogénéités locales avec localement des courbures gaussiennes non nulles) suivant les axes verticaux et horizontaux (tiré de [LZ04]) ; la forme obtenue est concave avec une partie centrale convexe ; les variations combinées d'orientation des composantes verticales et horizontales le long des axes principaux transmet bien l'information de forme pour la texture basée sur des réseaux (première ligne à gauche), la texture octotropique dont les 4 composantes proches de l'horizontale et de la verticale ont été supprimées (première ligne au milieu) et la texture formée de points régulièrement répartis mais de taille variable (deuxième ligne à gauche) ; la surface apparaît plate pour la texture octotropique (première ligne à droite) car les 4 composantes proches de l'horizontale et de la verticale masquent les variations des composantes exactement horizontales et verticales et la texture formée de points aléatoirement répartis (deuxième ligne à droite) dont le spectre global est isotropique.

Saunders et Backus [SBft] ont étudié la contribution de la perspective linéaire pour la perception de l'inclinaison de surfaces plane texturées afin notamment de tester l'hypothèse de Li et Zaidi (i.e l'estimation du slant à partir de la perspective linéaire dépend de la présence d'une composante spectrale orientée). Pour cela ils ont crée des textures formées de points Polka répartis selon une grille alignée avec le tilt, selon une grille avec un angle de roll de 30^o et de manière aléatoire (spectre isotropique) (Figure 4.24). Les sujets doivent juger le signe du slant (inclinaison vers la droite ou vers la gauche). Les résultats montrent que la perspective linéaire est bien utilisée comme indice 3D. Les auteurs confirment ainsi l'importance de la présence d'une composante spectrale orientée avec la direction du tilt pour percevoir efficacement l'inclinaison. Ils indiquent néanmoins que cela n'est pas nécessaire car une perception correcte est déjà obtenue avec les texture présentant un spectre isotropique.

Todd *et al* dans [TTD05] confirment l'hypothèse émise par Li et Zaidi mais essentiellement pour des stimuli vus sous une petite ouverture du champ visuel (5^o). Dans ces conditions, les sujets jugent parfaitement le signe du slant de surfaces recouvertes par une texture basée sur des réseaux ou possédant des contours réguliers (Figure 4.15) et la perception est plus

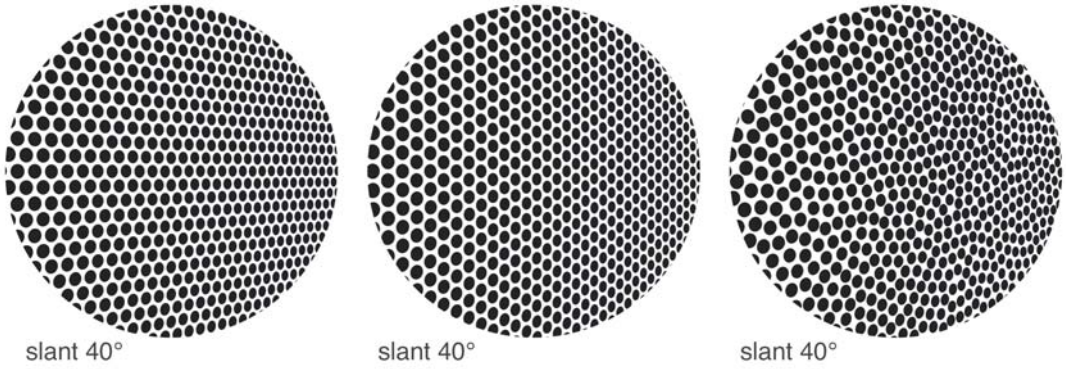


FIG. 4.23 – Exemples de textures utilisés par Saunders et Backus (tiré de [SBft]) ; de gauche à droite : les points sont répartis selon une grille alignée avec le tilt ; les points sont répartis selon une grille orientée avec un roll de 30^o par rapport au tilt ; les points sont répartis aléatoirement ; les seuils de discrimination augmentent successivement pour les textures de gauche à droite (les performances diminuent).

difficile pour des textures isotropiques. Cependant en considérant des ouvertures du champ visuel plus larges, les résultats de todd *et al* montrent que les sujets perçoivent bien le signe du slant (avec un taux de bonnes réponses à 90% pour 20^o d'ouverture et 99% à partir de 40^o) sur toutes les textures et notamment les textures isotropiques formées de points Polka. Dans [TO02a],

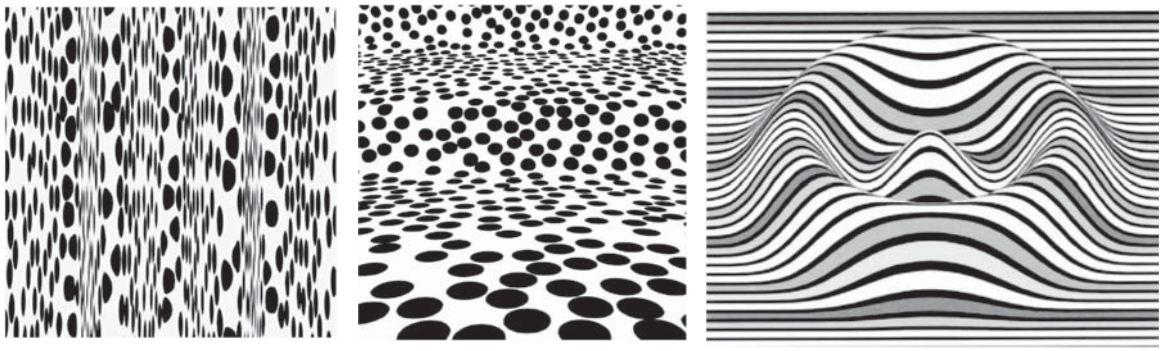


FIG. 4.24 – Exemples de textures présentées par Todd et Oomes (tiré de [TO02a]) afin de nuancer l'hypothèse de Li et Zaidi ; de gauche à droite : surface ondulée équivalente à celles de la figure 4.21 (les auteurs insistent sur le point de vue particulier imposé par la courbure de cette surface) ; surface courbe recouverte d'une texture isotropique (les différents plans, s'éloignant en profondeur, sont parfaitement visibles) ; surface volumétrique composée de lignes parallèles vue en projection orthographique suivant un angle oblique (les auteurs montrent que ce type de projection transmet également une information complexe de forme).

4.5 Résumé

Beaucoup de travaux ont étudié les gradients de texture tels que la variation de taille, de densité et de compression. L'utilisation de ces gradients de texture induit l'application d'une

hypothèse d'homogénéité et un traitement global de la texture. Cependant différents auteurs ont également appuyé l'application d'une hypothèse d'isotropie induisant un traitement local (au niveau de chaque élément individuel de la texture). Bien que beaucoup de résultats tendent vers l'utilisation de gradients de texture (hypothèse d'homogénéité plus générale que l'isotropie), la question reste ouverte.

Les performances du système visuel ont été évaluée en fonction des configurations géométriques de la surface (les performances s'améliorent pour des valeur de slant importantes et des tilt proche de 90^o (sols) et 0^o (murs)), en fonction de la régularité de la surface (les performances diminuent avec l'irrégularité, cependant celle-ci reste difficile à définir précisément), en fonction du champ visuel (une ouverture proche de 20^o Le gradient de compression semble être celui sur lequel le système visuel se base le plus pour percevoir une surface inclinée. Cependant plusieurs travaux montrent qu'il n'est pas suffisant et qu'il est possible d'envisager son interaction avec d'autres indices, tel que la perspective linéaire pour des surfaces planes, ou le passage à un autre espace de représentation tel que le domaine de Fourier.

Peu de travaux ont été menés sur la variation de fréquence comme indice de texture pour la perception 3D. Les auteurs ont beaucoup étudié les gradients de texture, tels que la compression, qui possèdent néanmoins des similarités avec le gradient de fréquence. Celui-ci représente une mesure statistique générale qui peut être effectuée sur tous les types de textures (macrotextures et microtextures, voir Chapitre 2.2). Ceci n'est pas le cas des gradients de texture nécessitant la segmentation individuelle des éléments de la texture, s'ils existent. Enfin l'utilisation par le système de la variation de fréquence n'a pas été démontrée car les stimuli utilisés pour les différentes expérimentations n'isolent pas cette information des autres gradients.

L'ensemble des travaux présentés sur l'étude de la perspective linéaire ont mis en évidence l'utilisation de cette information par le système visuel pour estimer l'inclinaison d'une surface texturée. Cependant cet indice est toujours introduit dans les stimuli par des lignes de fuites ou par des groupes d'éléments isolés qui forment une ligne de fuite subjective (par exemple des points Polka alignés). Ces éléments introduisent également une variation de fréquence (la convergence des lignes produit un effet de resserrement et les éléments subissent une déformation (compression) due à la projection ce qui provoque une variation vers les hautes fréquences spatiales) et ne permettent donc pas d'étudier l'utilisation de la perspective linéaire seule. D'après les travaux de Li et Zaidi, il apparaît cependant important de distinguer ces deux indices pour pouvoir estimer correctement l'orientation d'un plan.

Il est à noter enfin que dans la plupart des expériences, notamment celles de Li et Zaidi et de Saunders et Backus, la tâche consiste à estimer le signe du slant. Or cela est équivalent à estimer la direction en profondeur, c'est-à-dire la valeur du tilt. Ces expériences ne procurent donc pas de mesure quantitative des performances de discrimination entre des inclinaisons (slant) et des directions (tilt) différentes. Elles ne donnent pas non plus une mesure de la fiabilité relative entre l'indice de variation de fréquence et l'indice de perspective linéaire.

Le chapitre suivant 5 présente nos travaux sur ces deux indices et leur combinaison. Nous présentons de nouveaux stimuli permettant de s'abstraire des limitations rencontrées par les études précédentes. Ces stimuli permettent notamment la séparation complète des deux indices. De plus la contribution individuelle de chaque indice à la perception 3D est évaluée sur des tâches de discrimination de l'inclinaison et de l'orientation.

Chapitre 5

Perception 3D : gradient de fréquence et perspective linéaire

Ce chapitre présente nos travaux en psychophysique sur la perception 3D à partir de l'information de texture en vision monoculaire. Nous nous intéressons plus particulièrement aux indices de fréquence et de perspective linéaire tels qu'ils ont été introduits au chapitre précédent 4.4. Pour évaluer plus précisément la contribution des chacun de ces indices nous avons crée des stimuli composés d'une texture homogène artificiellement construite à partir d'un ensemble de masques de Gabor paramétrables en fréquence et en orientation. Ils forment une surface plane inclinée vue en projection perspective. Nous présentons tout d'abord la méthode permettant de générer les stimuli et leurs caractéristiques. Ensuite nous décrivons les expériences psychophysiques permettant d'étudier la perception de textures présentant uniquement un des deux indices, les deux indices en combinaison et les deux indices en conflit. L'influence des indices a été analysée en fonction des performances obtenues sur deux tâches perceptives distincts : la discrimination du *slant* et la discrimination du *tilt*. Les résultats obtenus sont commentés et mis en relation avec les travaux antérieurs. Nous discutons notamment de la validité de l'hypothèse d'isotropie et mettons en relation ce travail avec les travaux de Li et Zaidi présentés au chapitre 4.4. Nous discutons enfin d'une description plus précise des indices de texture en considérant séparément le gradient de fréquence, la perspective linéaire et la courbure.

Ce travail a été réalisé dans le cadre du programme EURODOC, programme d'aide à la mobilité des doctorants à l'étranger. J'ai ainsi pu effectuer un séjour de 6 mois à l'Université de Glasgow, dans le laboratoire de Pascal Mamassian. Il a donné lieu à plusieurs communications [8] [7] et d'un article en cours d'écriture à soumettre à Vision Research.

5.1 Génération des stimuli

5.1.1 Masques de Gabor

Un masque de Gabor correspond à la réponse impulsionnelle d'un filtre de Gabor. Ici, nous n'utiliserons pas ce masque comme filtre, mais au contraire comme générateur de texture

élémentaire qui se trouve localisé à la fois dans le domaine spatial et fréquentiel. C'est le motif de base des textures que nous allons générer.

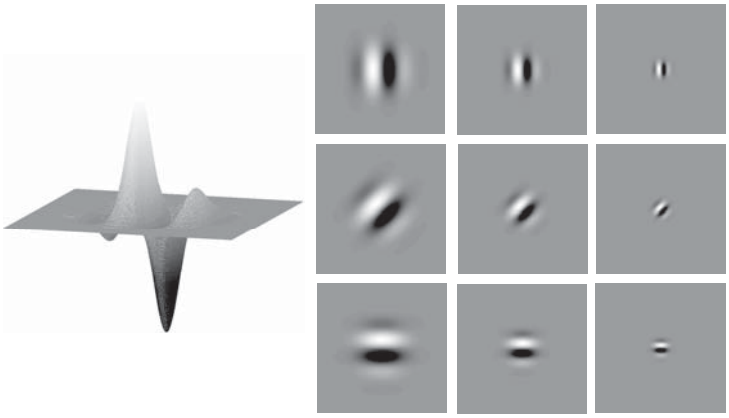


FIG. 5.1 – Exemples de masques de Gabor présentant une largeur de bande de 1.5 octave ; à gauche : profile 3D d'un masque ; à droite : masques présentant différentes fréquences spatiales (basses, moyennes et hautes fréquences) et différentes orientations (90^o,45^o,0^o).

Un masque de Gabor est formé d'un signal sinusoïdal orienté dans l'espace modulé par une enveloppe gaussienne en deux dimensions. En chaque position spatiale de coordonnées (x,y), la luminance est définie par :

$$\begin{aligned} I(x,y) =& L_o + L_m cos(2\pi f((x-x_o)sin\theta + (y-y_o)cos\theta) + \varphi) \\ & \times exp(-(\frac{(x-x_o)^2}{2\sigma^2} + \frac{(y-yo)^2}{2\sigma^2})) \end{aligned} \tag{5.1}$$

où L_o et L_m représentent les luminances moyennes et la modulation de contraste, f, la fréquence spatiale du signal sinusoïdal, θ, son orientation spatiale, φ, sa phase (fixée à $\pi/2$), (x_o, y_o), les coordonnées du centre du masque, et σ, la largeur de l'enveloppe gaussienne. Pour nos stimuli nous reprenons la convention utilisée par Prins et Kingdom [PK02] consistant à faire varier σ avec la fréquence afin de garder constante la largeur de bande à mi-hauteur, notée Δf, et fixée à 1.5 octave par la formule :

$$\sigma = \frac{1}{f\pi}\sqrt{\frac{ln2}{2}}\frac{2^{\Delta f}+1}{2^{\Delta f}-1} \tag{5.2}$$

La figure 5.1 présente des exemples de maques de Gabor ainsi obtenus pour différentes valeurs de fréquence centrale et d'orientation.

Les masques de Gabor sont positionnés sur la surface suivant une distribution uniforme (Figure 5.3). Chaque masque est placé successivement en respectant la contrainte que le centre de chaque nouveau masque ne doit pas être situé à une distance inférieure déterminée du centre des autres masques déjà présents sur la surface. Cette distance est choisie égale à 1.6 σ permettant un compromis entre une couverture maximum de la surface et un recouvrant minimum entre les masques. Tous les recouvrements sont effectués sur les valeurs des contrastes des masques (intensité moyenne nulle) afin de maintenir une intensité moyenne uniforme sur la surface. Une couverture complète de la surface est obtenue après un grand nombre de tirages (fixé à 500000).

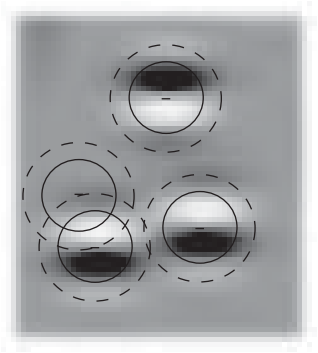


FIG. 5.2 – Exemples de placement des masques de Gabor ; les cercles pleins représentent la largeur à mi-hauteur (rayon σ) ; les cercles en pointillées représentent la zone d'inhibition (rayon 1.6 σ).

Les stimuli considérés dans ce travail représentent des surfaces planes inclinées dans l'espace. Ils sont présentés à travers une ouverture circulaire afin de supprimer la bordure rectangulaire. En effet les verticales et les horizontales produites par les bords représentent des repères pouvant introduire un autre type d'indice non contrôlé et pouvant rentrer en conflit (par exemple le contour indique que la surface de l'écran sur lequel est projeté le stimulus est une surface plane fontoparallèle). La bordure de l'ouverture est adoucie avec une fonction sigmoidale afin de réduire la transition brutale entre le fond noir et la texture. Pour l'ensemble des textures générées, le contraste rms (*root mean square*), mesuré dans différentes sous-régions de la surface, reste constant.

Les masques sont indépendants les uns des autres. Leur fréquence et leur orientation peuvent être modifiées séparément. Il est alors possible d'appliquer des variations de fréquence ou d'orientation en fonction de leur position spatiale.

FIG. 5.3 – Exemples de texture homogène obtenue avec une fréquence fixe pour tous les masques de Gabor et deux orientations possibles (0^o et 90^o) ; de gauche à droite : après tirage de 50, 500 et 500000 masques.

5.1.2 Variation de fréquence

Comme nous l'avons vu à la section 4.4, la projection d'une surface plane induit un gradient de fréquence dans la texture. Nous calculons ici la variation théorique de la fréquence

en fonction du tilt et du slant afin d'obtenir la simulation d'une surface plane inclinée dans l'espace.

5.1.2.1 Calcul de la variation de fréquence

Soit f_s la fréquence d'un masque de Gabor plaqué sur la surface 3D avant projection, la fréquence de sa projection sur le plan de l'image 2D f_i s'exprime par :

$$f_i \approx \frac{1}{a_i}\left(I - \frac{\nabla a_i x_i^t}{a_i}\right) A^t f_s \tag{5.3}$$

où, en reprenant les notations de l'équation 2.1, (x_i, y_i) représentent les coordonnées du centre du masque de Gabor dans l'image ; A^t est la matrice correspondant à la projection perspective et ne dépend que du slant et du tilt ; $\frac{1}{a_i}$ est un facteur d'échelle dépendant de la position spatiale (x_i, y_i) ; $\left(I - \frac{\nabla a_i x_i^t}{a_i}\right)$ est un facteur de correction. Le lecteur trouvera en annexe A.1 le calcul complet de la variation de fréquence spatiale.

5.1.2.2 Vérification et résultats

Pour vérifier la validité de l'équation 5.3, nous effectuons une analyse expérimentale consistant à comparer le résultat obtenu en appliquant cette équation avec une projection réelle.

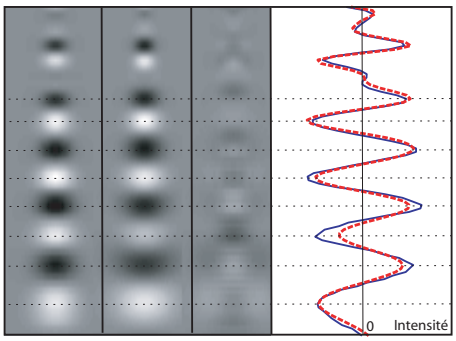


FIG. 5.4 – Vérification de l'équation 5.3 ; de gauche à droite : projection réelle ; simulation de l'inclinaison d'un plan à l'aide de l'équation ; différence entre les deux images obtenues ; superposition des variations d'intensité de la ligne centrale verticale de la projection réelle (courbe bleue en trait plein) et de la simulation (courbe rouge en trait pointillé).

La figure 5.4 présente un ensemble de masques de Gabor identiques orientés horizontalement ayant subit deux types de transformation : à gauche l'image formée de l'ensemble des masques a été inclinée dans l'espace induisant leur déformation due à la projection perspective ; à droite la position et la fréquence de chaque masque ont été adaptées individuellement pour simuler une inclinaison équivalente ; la troisième colonne correspond à la différence entre la projection réelle et la version simulée. Nous observons que la différence est quasiment nulle. La dernière colonne présente la superposition des variations d'intensité des lignes verticales passant par le centre des images de la première colonne et de la seconde colonne. Malgré quelques différences d'amplitude dans les hautes fréquences (dues à la limite imposée par la résolution), les deux variations se superposent très bien. Le lecteur trouvera en annexe A.3 des commentaires supplémentaires sur les stimuli.

Les figures 5.5 et 5.6 montrent des exemples d'images représentant des stimuli avec uniquement une variation de fréquence. Ces images sont générées pour différentes valeur de slant (Figure 5.5) et de tilt (Figure 5.6). L'orientation des masques de Gabor est choisie selon une distribution uniforme afin d'enlever tout indice de variation d'orientation. Il est important de noter que l'utilisation d'orientations aléatoires pour les masques de Gabor correspond à une surface frontoparallèle. En effet une véritable projection induit une modification de la distribution des orientations, centrée autour de la direction du tilt. Nous reviendrons sur ce point à la section 5.3.

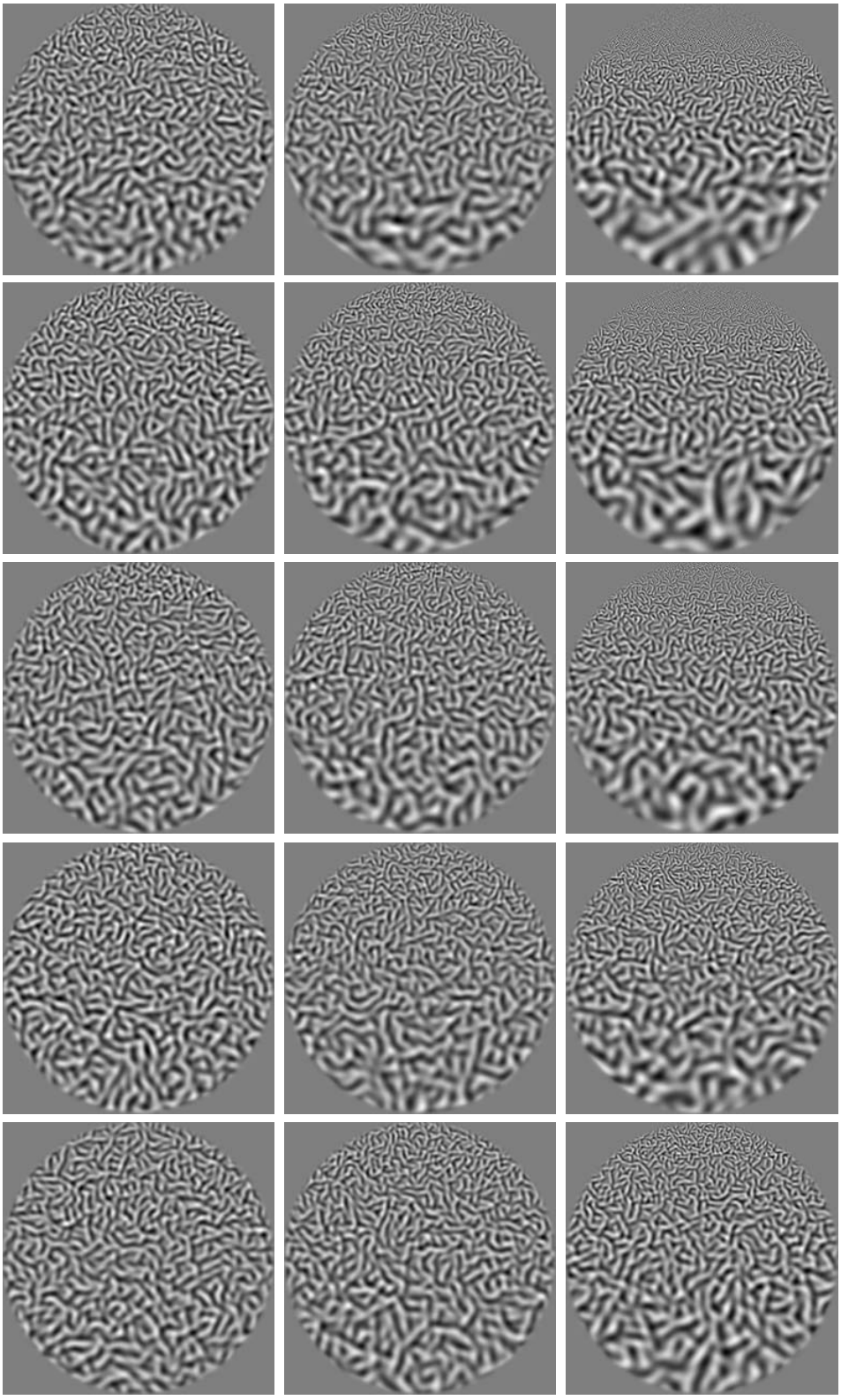


FIG. 5.5 – Exemples de stimuli présentant uniquement une variation de fréquence avec des orientations aléatoires ; tous les stimuli ont un tilt fixé à 90^o et des slant variables ; la colonne de gauche présente des stimuli inclinés à (de bas en haut) 19.5^o, 22.5^o, 27^o, 31.5^o et 34.5^o ; la colonne du milieu présente des stimuli inclinés à 32.5^o, 35.5^o, 40^o, 44.5^o et 47.5^o ; la colonne de droite présente des stimuli inclinés à 45.5^o, 48.5^o, 53^o, 57.5^o et 60.5^o.

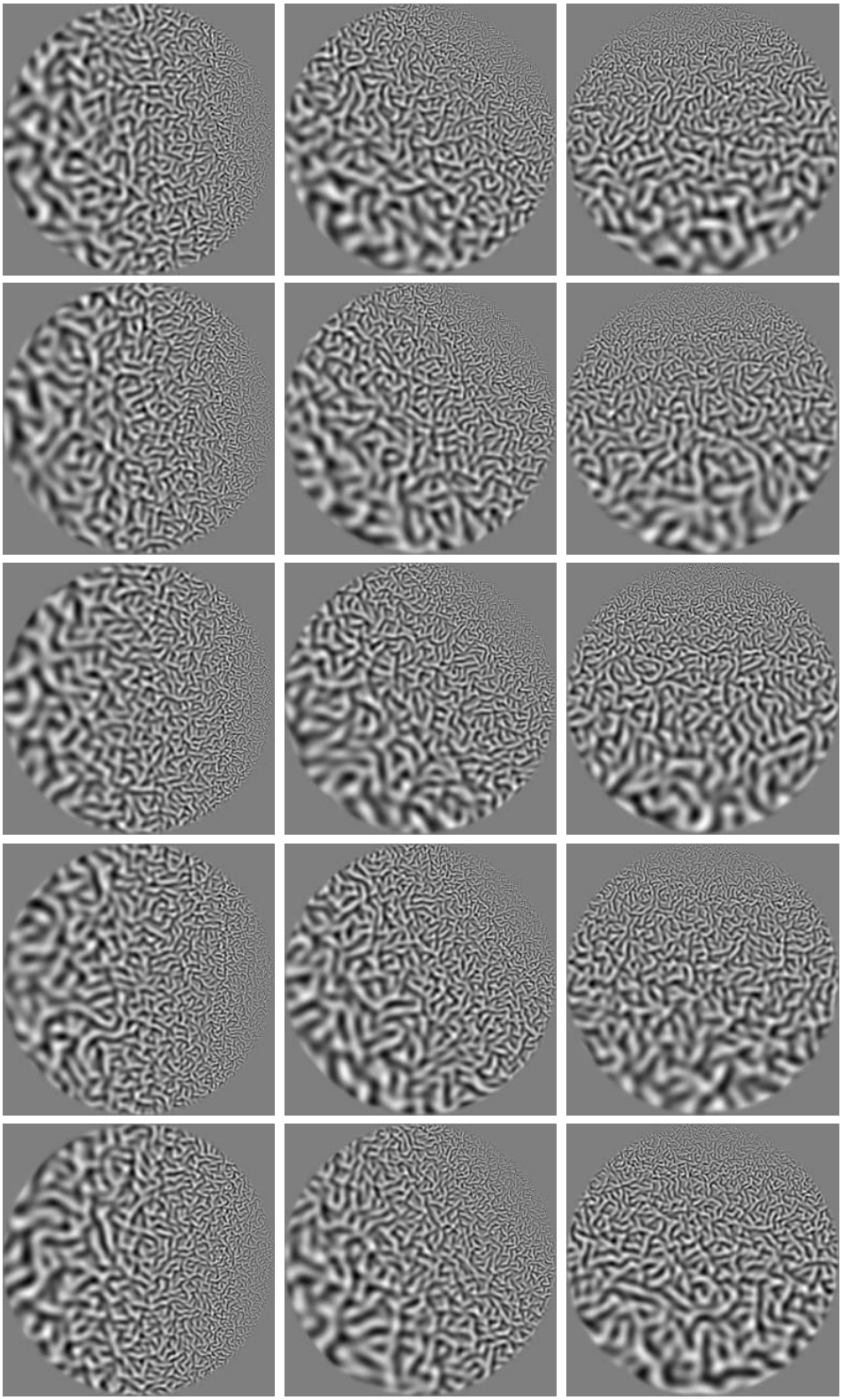


FIG. 5.6 – Exemples de stimuli présentant uniquement une variation de fréquence avec des orientations aléatoires ; tous les stimuli ont un slant fixé à 53^o et des tilt variables ; la colonne de gauche présente des stimuli orientés à (de bas en haut) $-7.5^o, -4.5^o, 0^o, 4.5^o$ et 7.5^o ; la colonne du milieu présente des stimuli orientés à $37.5^o, 40.5^o, 45^o, 49.5^o$ et 52.5^o ; la colonne de droite présente des stimuli orientés à $82.5^o, 85.5^o, 90^o, 94.5^o$ et 97.5^o.

5.1.3 Variation d'orientation

Comme nous l'avons vu à la section 4.4, la projection d'une surface plane induit un gradient d'orientation dans la texture, appelé couramment perspective linéaire et sert d'indice à la perception 3D. Sur les stimuli cela se traduit par une variation de l'orientation des masques de Gabor en fonction de leur position dans l'image.

De la même manière que pour la variation de fréquence, nous calculons ici la variation théorique de l'orientation en fonction du tilt et du slant afin d'obtenir la simulation d'une surface plane inclinée dans l'espace.

5.1.4 Calcul de la variation d'orientation

La modulation d'orientation est simplement donnée par l'orientation d'une droite après sa projection dans l'image (Figure 5.7).

Nous considérons une droite sur la surface 3D initiale orientée d'un angle α par rapport à l'horizontal à une position spatiale (x_s, y_s). La projection de cette droite sur l'image donne une droite à la position spatiale (x_i, y_i) avec une orientation β donnée par :

$$\beta = arctan\left(\frac{S}{C}\right) \tag{5.4}$$

où C et S dépendent de l'orientation initiale α, de l'inclinaison de la surface σ, de son orientation τ et la position de la droite dans la surface initiale (x_s, y_s). Le lecteur trouvera en annexe A.2 le calcul complet de la variation d'orientation spatiale.

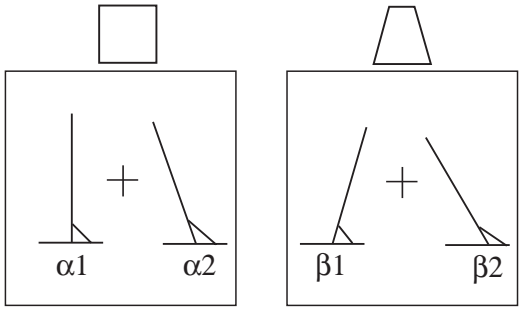


Fig. 5.7 – Exemple de changement d'orientation de deux droites avant projection (à gauche) et après projection (à droite) ; à l'angle α_1 (resp α_2) de la droite dans la surface initiale correspond l'angle β_1 (resp β_2) dans l'image obtenue par projection.

5.1.4.1 Vérification et résultats

De la même manière que pour la variation de fréquence, nous vérifions par simulation la validité de l'équation 5.4 expérimentalement en comparant également le résultat obtenu avec une projection réelle.

La figure 5.8 présente un ensemble de masques de Gabor identiques orientés verticalement. La première rangée de masques présente le résultat de la projection réelle par l'inclinaison de la surface plane formée par l'ensemble des masques. La seconde rangée présente le résultat obtenu par simulation en appliquant l'équation 5.4 pour la même inclinaison et en fonction de leur

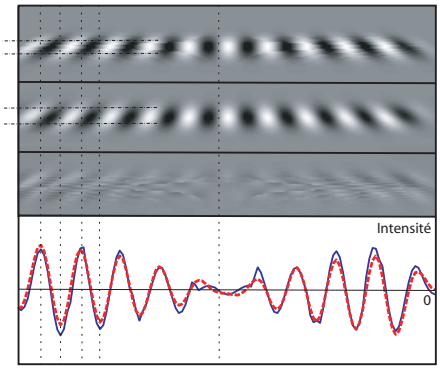


FIG. 5.8 – Vérification de l'équation 5.4 ; de haut en bas : projection réelle ; simulation de l'inclinaison d'un plan à l'aide de l'équation ; différence entre les deux images obtenues ; superposition des variations d'intensité de la projection réelle (courbe bleue en trait plein) et de la simulation (courbe rouge en trait pointillé) obtenues par différence entre les intensités de deux lignes équi-distantes de la ligne passant par le centre des masques.

position respective. La troisième rangée présente la différence entre les deux premières. Celle-ci est quasiment nulle. La dernière ligne présente la superposition des variations d'intensité entre les deux premières lignes. Pour la projection réelle, la variation d'intensité est obtenue en faisant la différence entre les deux profils de luminance situés de manière équidistante de la ligne horizontale passant par le centre des masques. De cette manière la variation d'intensité obtenue prend en compte l'orientation locale des masques. La variation d'intensité pour la seconde rangée de masques obtenus par simulation est calculée de la même manière. Les deux variations d'intensité se superposent très bien. Le lecteur trouvera en annexe A.3 des commentaires supplémentaires sur les stimuli.

Les figure 5.9 et 5.10 montrent des exemples d'images représentant des stimuli avec uniquement une variation d'orientation. Ces images sont générées pour différentes valeur de slant (Figure 5.9) et de tilt (Figure 5.10). La fréquence des masques de Gabor est maintenue constante afin d'enlever tout indice de variation de fréquence. Elle est choisie relativement haute fréquence pour obtenir une perception optimale de l'inclinaison (voir Section 4.3) tout en évitant les problèmes liés à la limitation imposée par la résolution. L'ensemble de ces masques se superpose aux lignes de fuite passant par leur position spatiale.

Afin de renforcer l'impression de surface, des masques de Gabor orientés orthogonalement à la direction du tilt sont ajoutés. Ces nouveaux masques ne rajoutent pas d'indice de variation d'orientation, celle-ci étant nulle dans la direction orthogonale au tilt. De plus la texture ainsi obtenue possède, avant projection, de l'information concentrée principalement autour de deux orientations orthogonales. Bien que cette texture ne soit pas strictement isotropique, une hypothèse d'isotropie (dans un sens faible) peut cependant s'appliquer en chaque position de la surface en rendant l'indice de rupture d'isotropie utilisable. Ceci permet ainsi de s'assurer d'obtenir la meilleure perception de l'inclinaison de la surface du point de vue de toutes les hypothèses applicables (homogénéité et isotropie) (voir Section 4.2 et voir Section 5.3).

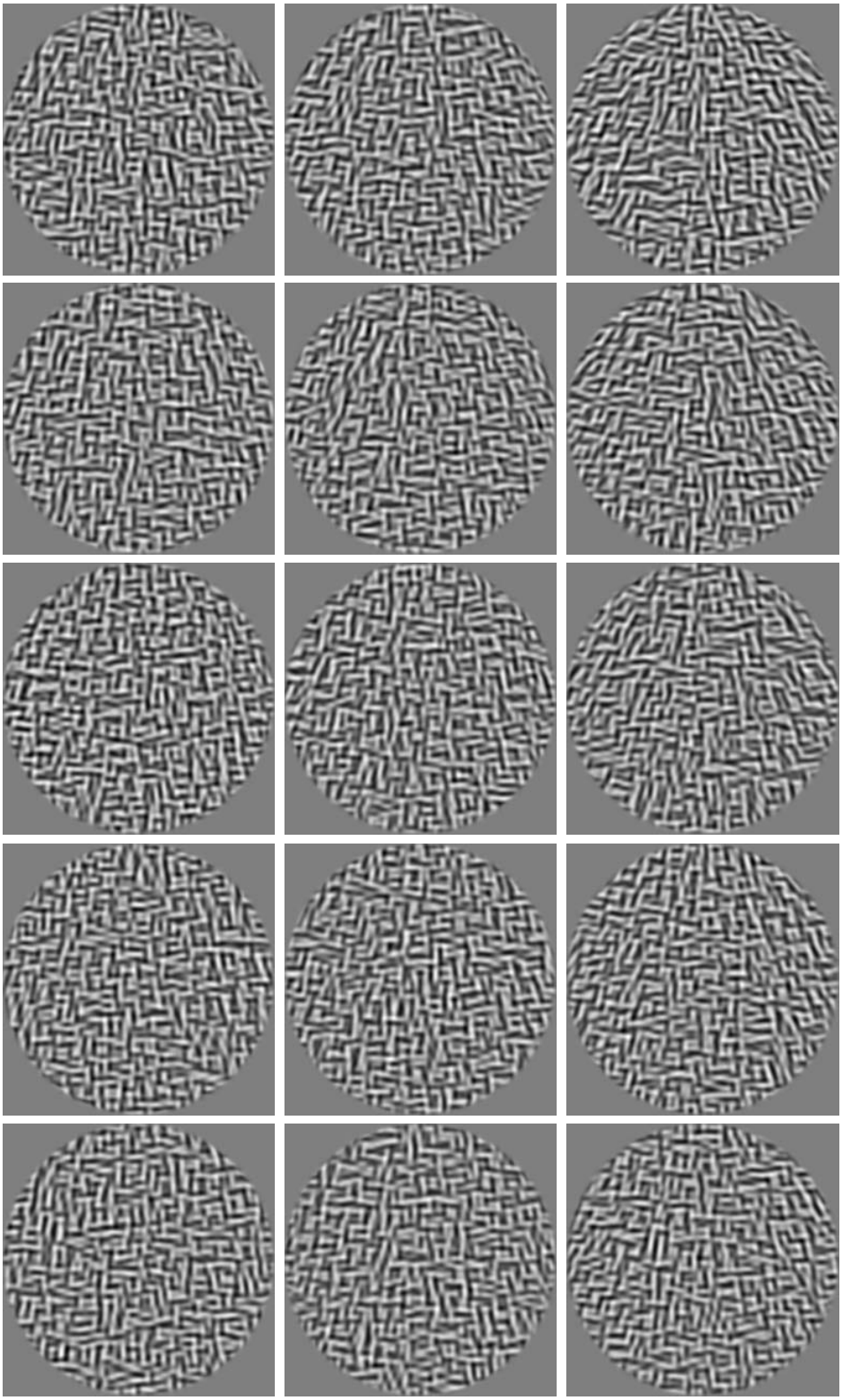


FIG. 5.9 – Exemples de stimuli présentant uniquement une variation d'orientation avec une fréquence constante ; tous les stimuli ont un tilt fixé à 90^o et des slant variables ; la colonne de gauche présente des stimuli inclinés à (de bas en haut) $19.5^o, 22.5^o, 27^o, 31.5^o$ et 34.5^o ; la colonne du milieu présente des stimuli inclinés à $32.5^o, 35.5^o, 40^o, 44.5^o$ et 47.5^o ; la colonne de droite présente des stimuli inclinés à $45.5^o, 48.5^o, 53^o, 57.5^o$ et 60.5^o.

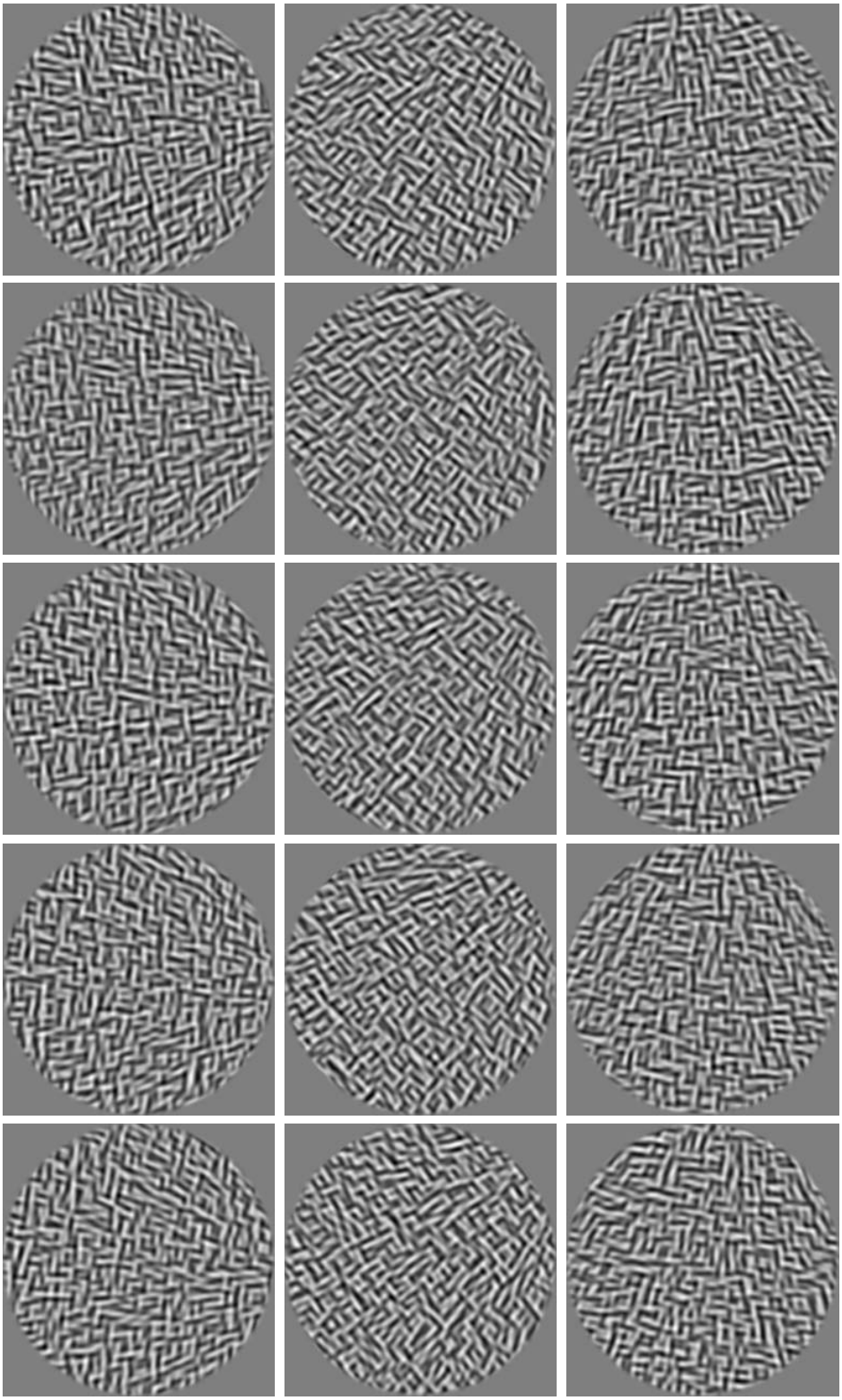


FIG. 5.10 – Exemples de stimuli présentant uniquement une variation d'orientation avec une fréquence constante ; tous les stimuli ont un slant fixé à 53^o et des tilt variables ; la colonne de gauche présente des stimuli orientés à (de bas en haut) $-7.5^o, -4.5^o, 0^o, 4.5^o$ et 7.5^o ; la colonne du milieu présente des stimuli orientés à $37.5^o, 40.5^o, 45^o, 49.5^o$ et 52.5^o ; la colonne de droite présente des stimuli orientés à $82.5^o, 85.5^o, 90^o, 94.5^o$ et 97.5^o.

5.1.5 Combinaison des variations de fréquence et d'orientation

Comme il a été montré aux sections précédentes 5.1.2 et 5.1.3, il est possible de créer des stimuli avec des indices de fréquence et d'orientation indépendants. Il est donc possible d'obtenir des stimuli représentant des surfaces planes inclinées dans l'espace en manipulant soit la variation de fréquence, soit la variation d'orientation ou en appliquant les deux variations simultanément soit en combinaison soit en conflit (i.e. les deux indices indiquent la même information de slant et de tilt ou une information différente sur l'un ou l'autre angle).

5.1.5.1 Stimuli en combinaison

Les figures 5.11 et 5.12 présentent un ensemble de stimuli avec respectivement différentes valeurs de slant et différentes valeurs de tilt. Sur ces exemples les deux indices se combinent pour simuler une surface plane inclinée en profondeur.

5.1.5.2 Stimuli en conflit sur le slant

Les figures 5.13 et 5.14 présentent les stimuli obtenus dans le cas où les informations données par les deux indices sont contradictoires sur l'information de slant. Le tilt reste constant et est fixé à 90^o pour faciliter la perception conformément aux travaux antérieurs (voir Section 4.3).

La figure 5.13 présente des stimuli avec une variation d'orientation fixée à une valeur de slant correspondant à la valeur de référence et une variation de fréquence variable correspondant à des valeurs successives de slant autour de cette référence.

La figure 5.14 présente le conflit inverse du précédent avec des stimuli présentant une variation de fréquence fixée à une valeur de slant correspondant à la valeur de référence et une variation d'orientation variable correspondant à des valeurs successives de slant autour de cette référence.

5.1.5.3 Stimuli en conflit sur le tilt

Les mêmes stimuli peuvent être crées en présentant un conflit sur l'information globale d'orientation (la valeur du tilt). Les figures 5.15 et 5.16 présentent les stimuli obtenus dans le cas où les informations données par les deux indices sont contradictoires sur l'information de tilt. Le slant reste constant et est fixé à 53^o correspondant à une forte inclinaison pour faciliter la perception, également conformément aux travaux antérieurs (voir Section 4.3).

La figure 5.15 présente des stimuli avec une variation d'orientation fixée à une valeur de tilt correspondant à la valeur de référence et une variation de fréquence variable correspondant à des valeurs successives de tilt autour de cette référence.

La figure 5.16 présente des stimuli avec une variation de fréquence fixée à une valeur de tilt correspondant à la valeur de référence et une variation d'orientation variable correspondant à des valeurs successives de tilt autour de cette référence.

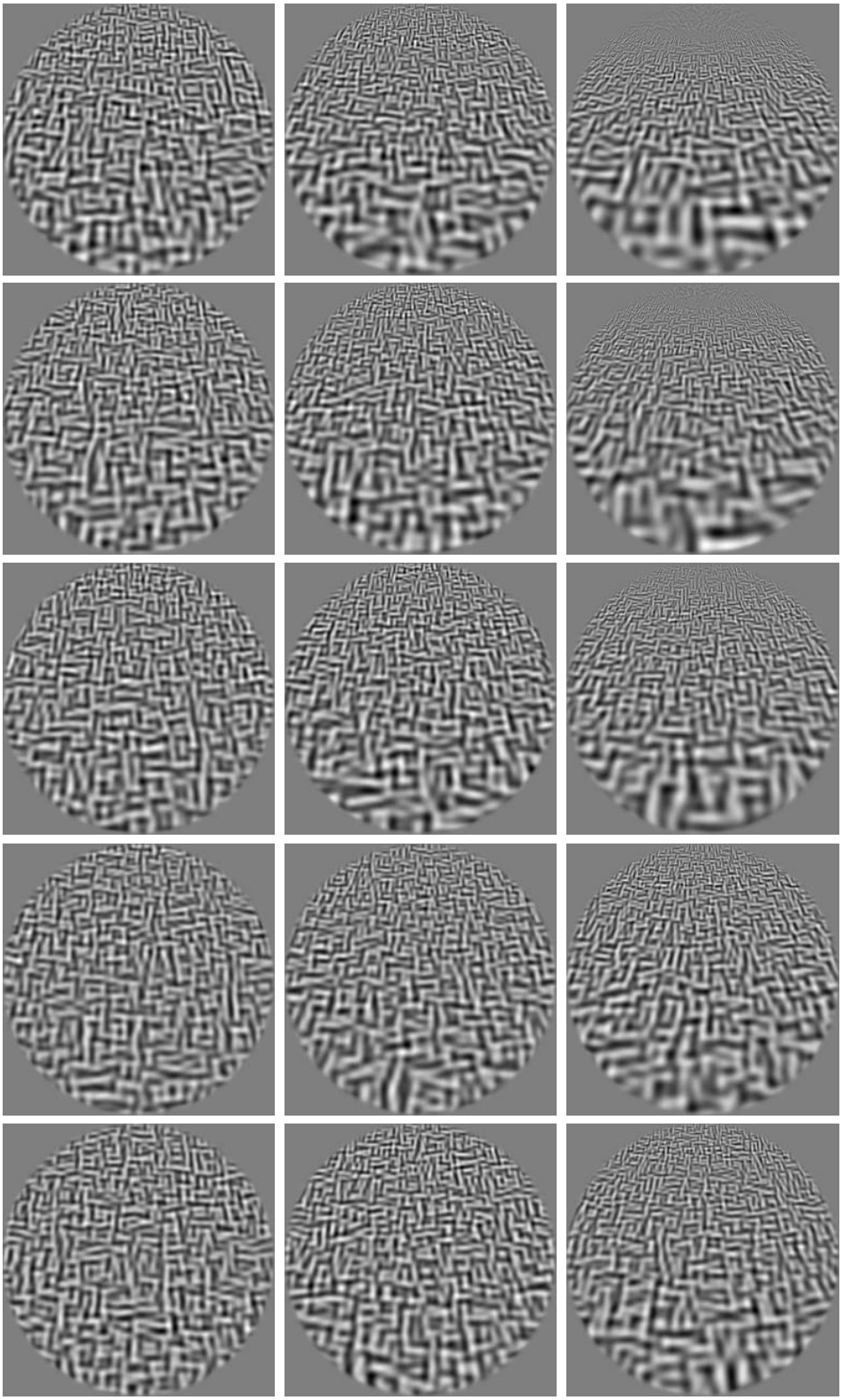


FIG. 5.11 – Exemples de stimuli présentant une variation de fréquence combinée à une variation d'orientation ; tous les stimuli ont un tilt fixé à 90^o et des slant variables ; la colonne de gauche présente des stimuli inclinés à (de bas en haut) $19.5^o, 22.5^o, 27^o, 31.5^o$ et 34.5^o ; la colonne du milieu présente des stimuli inclinés à $32.5^o, 35.5^o, 40^o, 44.5^o$ et 47.5^o ; la colonne de droite présente des stimuli inclinés à $45.5^o, 48.5^o, 53^o, 57.5^o$ et 60.5^o.

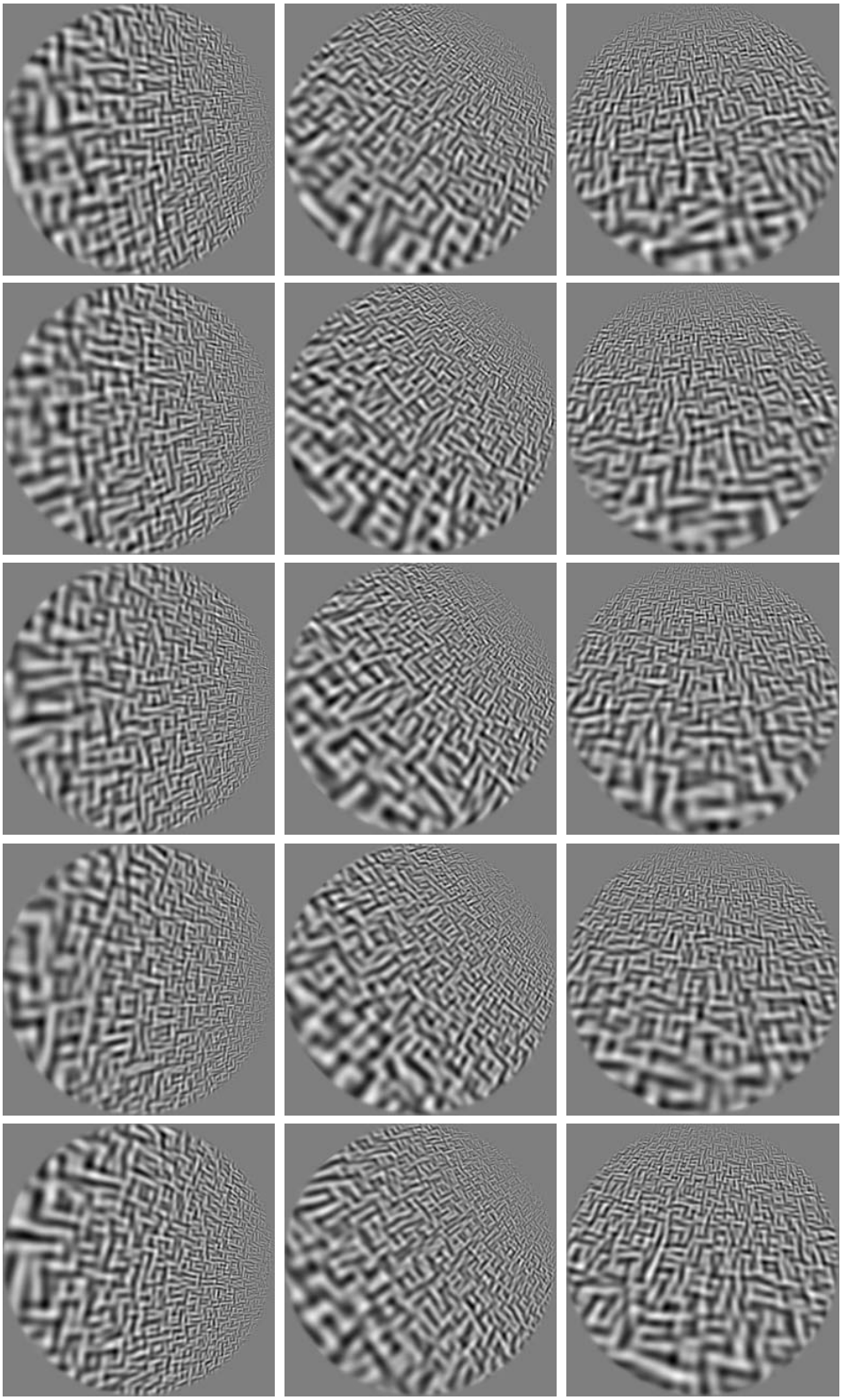


FIG. 5.12 – Exemples de stimuli présentant une variation de fréquence combinée à une variation d'orientation ; tous les stimuli ont un slant fixé à 53^o et des tilt variables ; la colonne de gauche présente des stimuli orientés à (de bas en haut) $-7.5^o, -4.5^o, 0^o, 4.5^o$ et 7.5^o ; la colonne du milieu présente des stimuli orientés à $37.5^o, 40.5^o, 45^o, 49.5^o$ et 52.5^o ; la colonne de droite présente des stimuli orientés à $82.5^o, 85.5^o, 90^o, 94.5^o$ et 97.5^o.

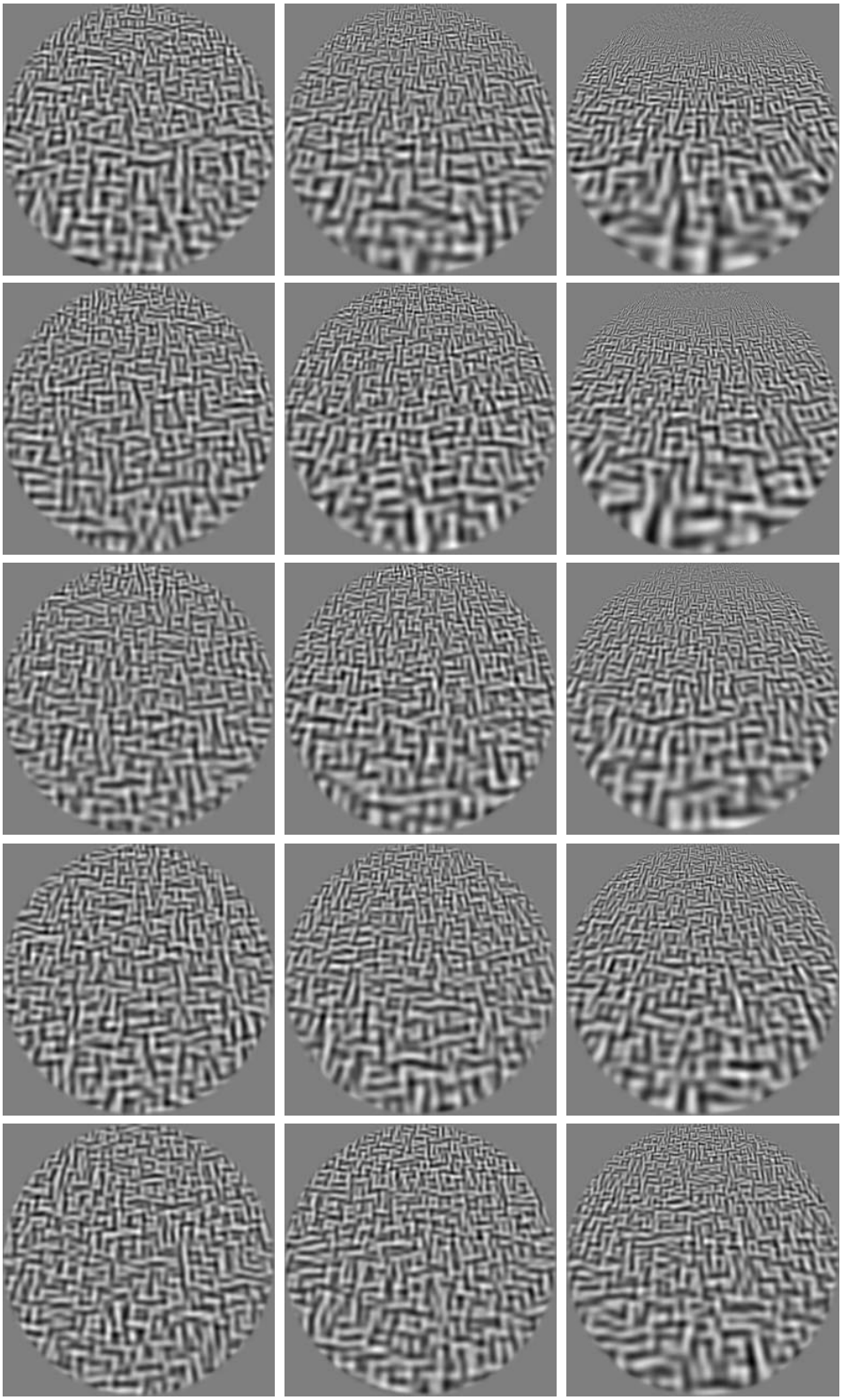


FIG. 5.13 – Exemples de stimuli présentant une variation de fréquence en conflit avec une variation d'orientation fixée aux différentes valeurs de référence de slant (27^o, 40^o et 53^o) ; tous les stimuli ont un tilt fixé à 90^o et des slant variables ; la colonne de gauche présente des stimuli inclinés à (de bas en haut) $19.5^o, 22.5^o, 27^o, 31.5^o$ et 34.5^o ; la colonne du milieu présente des stimuli inclinés à $32.5^o, 35.5^o, 40^o, 44.5^o$ et 47.5^o ; la colonne de droite présente des stimuli inclinés à $45.5^o, 48.5^o, 53^o, 57.5^o$ et 60.5^o.

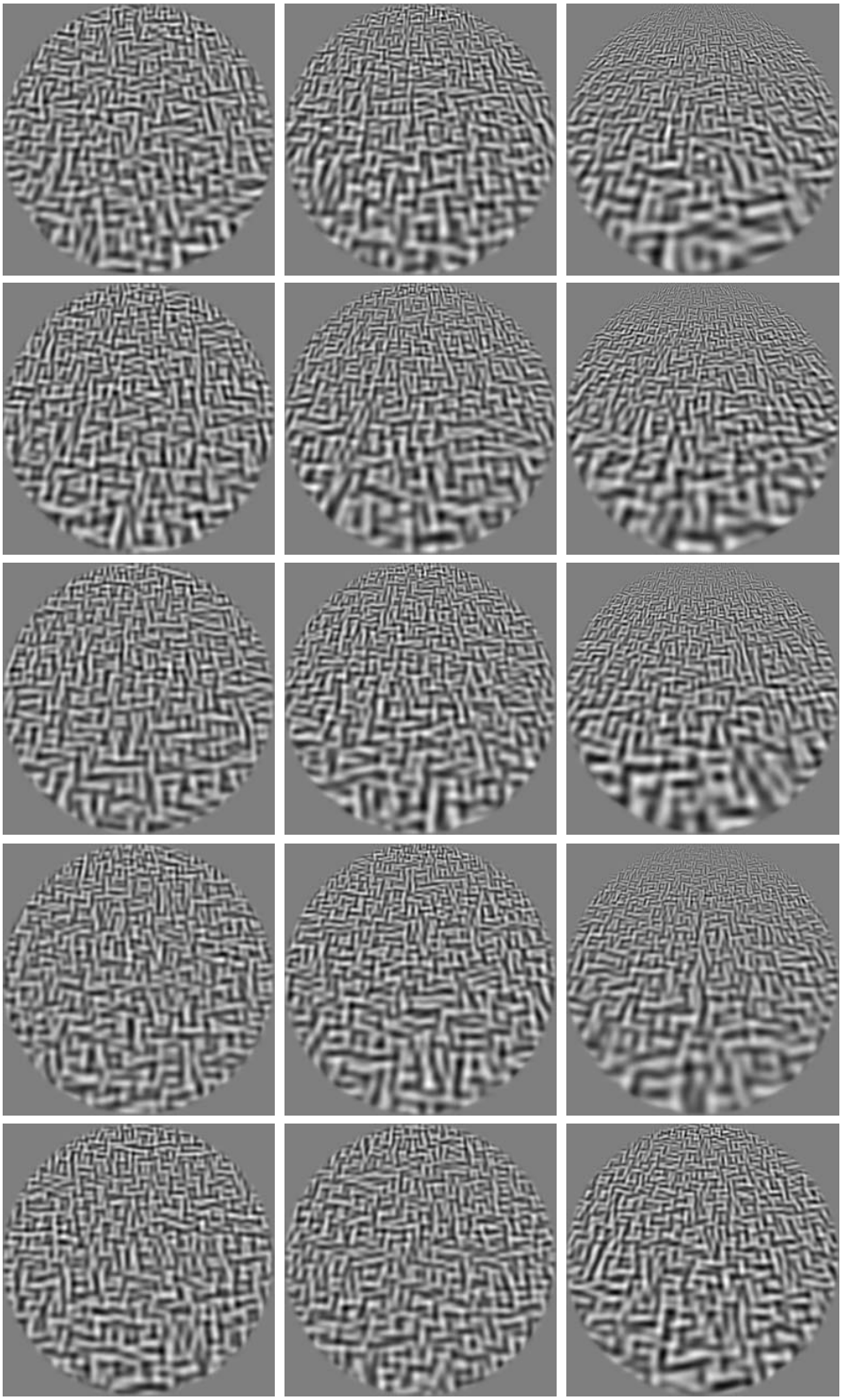


FIG. 5.14 – Exemples de stimuli présentant une variation d'orientation en conflit avec une variation de fréquence fixée aux différentes valeurs de référence de slant (27^o, 40^o et 53^o) ; tous les stimuli ont un tilt fixé à 90^o et des slant variables ; la colonne de gauche présente des stimuli inclinés à (de bas en haut) $19.5^o, 22.5^o, 27^o, 31.5^o$ et 34.5^o ; la colonne du milieu présente des stimuli inclinés à $32.5^o, 35.5^o, 40^o, 44.5^o$ et 47.5^o ; la colonne de droite présente des stimuli inclinés à $45.5^o, 48.5^o, 53^o, 57.5^o$ et 60.5^o.

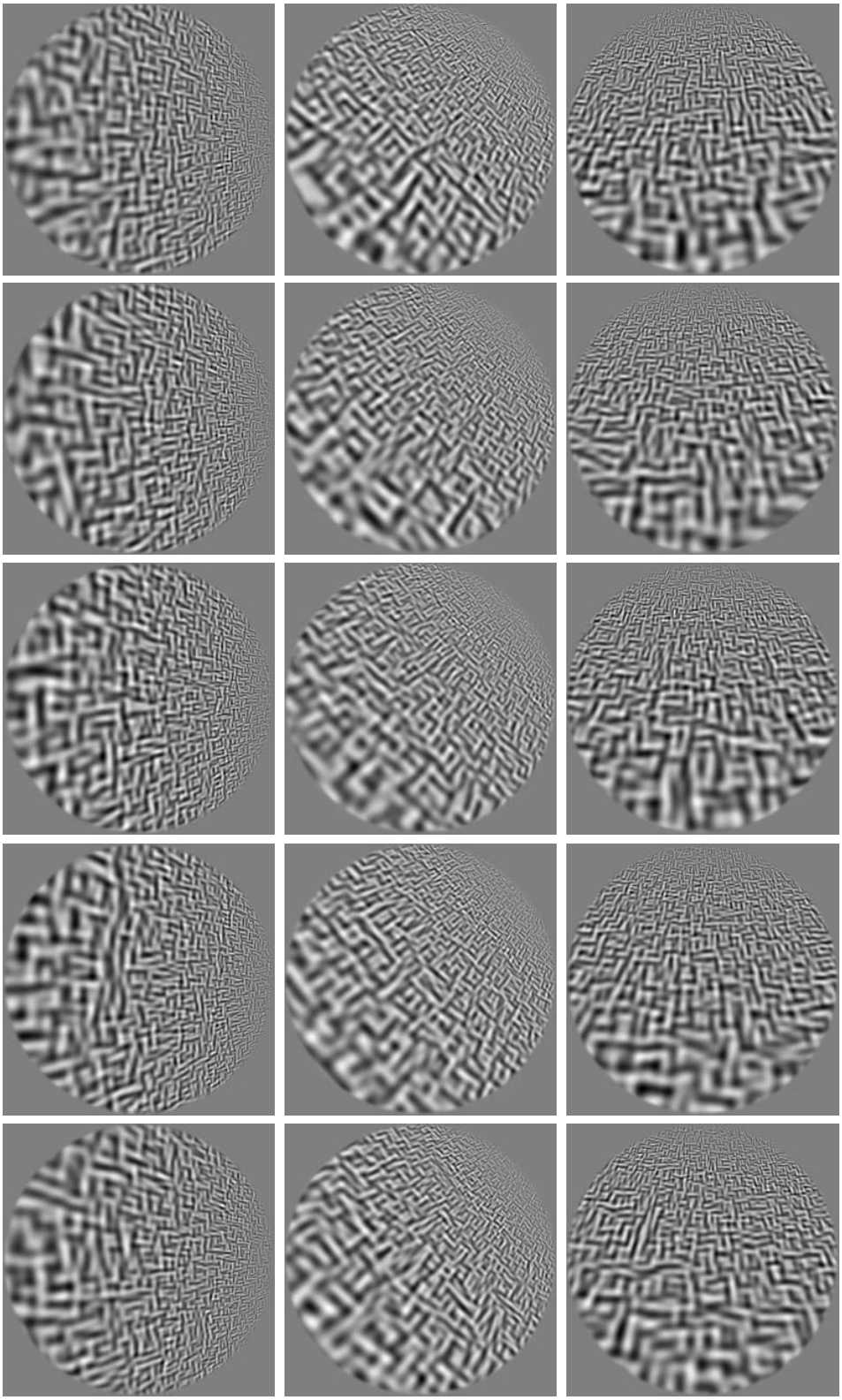


FIG. 5.15 – Exemples de stimuli présentant une variation de fréquence en conflit avec une variation d'orientation fixée aux différentes valeurs de référence de tilt (0^o, 45^o et 90^o) ; tous les stimuli ont un slant fixé à 53^o et des tilt variables ; la colonne de gauche présente des stimuli orientés à (de bas en haut) $-7.5^o, -4.5^o, 0^o, 4.5^o$ et 7.5^o ; la colonne du milieu présente des stimuli orientés à $37.5^o, 40.5^o, 45^o, 49.5^o$ et 52.5^o ; la colonne de droite présente des stimuli orientés à $82.5^o, 85.5^o, 90^o, 94.5^o$ et 97.5^o.

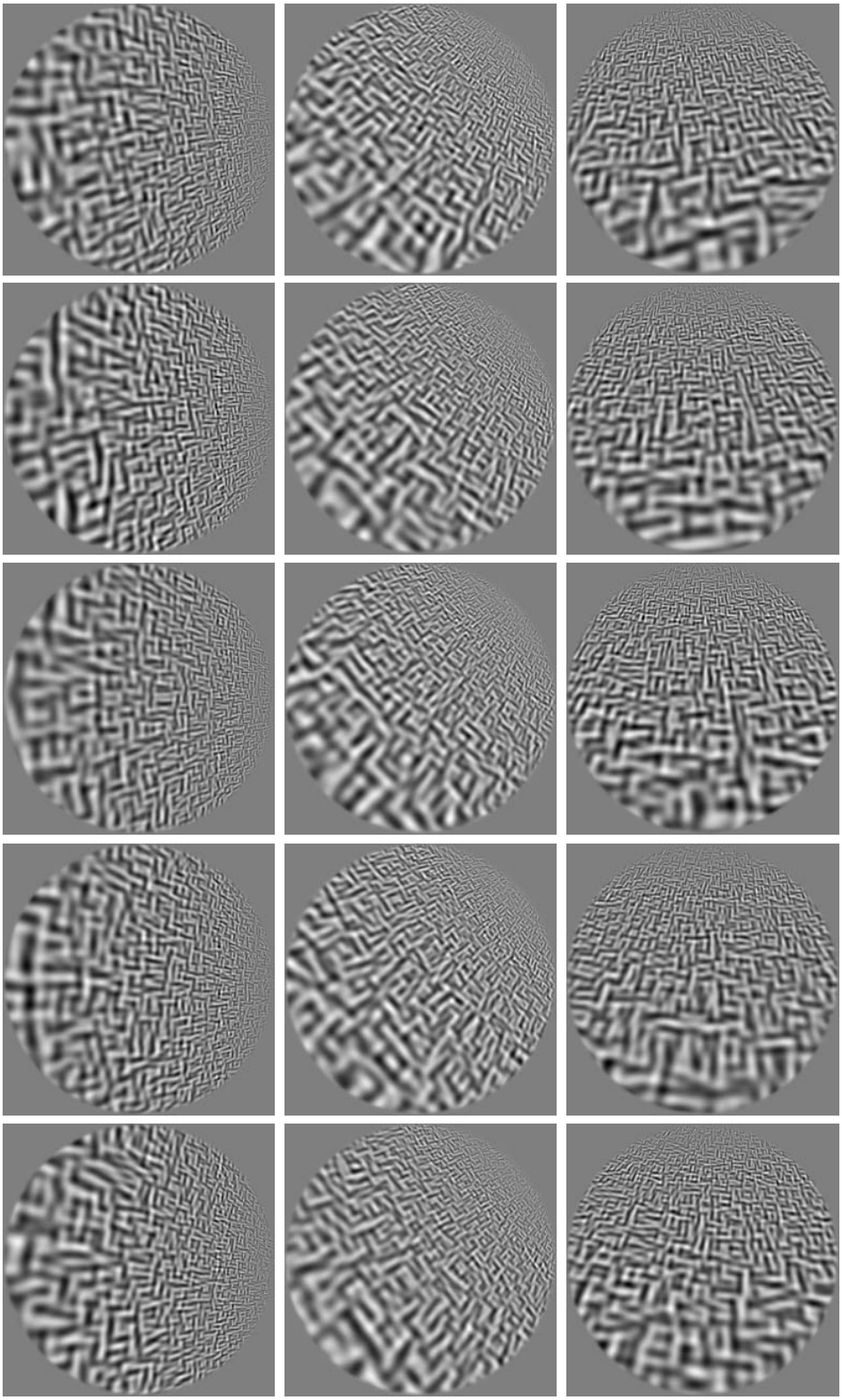


FIG. 5.16 – Exemples de stimuli présentant une variation d'orientation en conflit avec une variation de fréquence fixée aux différentes valeurs de référence de tilt (0^o, 45^o et 90^o) ; tous les stimuli ont un slant fixé à 53^o et des tilt variables ; la colonne de gauche présente des stimuli orientés à (de bas en haut) $-7.5^o, -4.5^o, 0^o, 4.5^o$ et 7.5^o ; la colonne du milieu présente des stimuli orientés à $37.5^o, 40.5^o, 45^o, 49.5^o$ et 52.5^o ; la colonne de droite présente des stimuli orientés à $82.5^o, 85.5^o, 90^o, 94.5^o$ et 97.5^o.

5.2 Expériences

5.2.1 Protocole

Le protocole expérimental a été conçu pour minimiser l'intervention d'indices 3D autres que ceux manipulés dans les stimuli. L'écran est un ViewSonic de 21 pouces G220F (400X350mm), placé à 575mm de l'oeil de l'observateur. Celui-ci a été consciencieusement calibré en luminance (linéarisation de la fonction gamma) et la géométrie a été corrigée. Un repose-menton a été utilisé pour maintenir la tête et gardée constante la distance de vue de l'observateur au cours de l'expérience. Une fois placé sur le repose-menton, l'observateur est entouré d'un morceau de carton cylindrique noir avec uniquement une ouverture circulaire de 85 mm de diamètre centrée sur sa ligne de vision (figure 5.17). Cette configuration permet une ouverture du champ visuel de 14^o tout en empêchant la perception de toute autre information visuelle dans la périphérie (bords de l'écran, objets dans la pièce) une fois l'observateur plongé dans le noir. La vision est maintenue monoculaire à l'aide d'un cache-oeil. L'oeil de l'observateur, l'ouverture et le centre de l'écran sont alignés. L'expérience est conduite à l'aide de la Psychtoolbox [Pel97] [Bra97] sur un Macintosh PowerMac G4 pour contrôler l'affichage, les temps d'exposition et les réponses entrées au clavier.

FIG. 5.17 – Protocole expérimental ; de gauche à droite : équipement ; position de l'observateur ; vision monoculaire à travers l'ouverture circulaire.

5.2.2 Procédure

Il y a deux séries d'expériences correspondant à deux tâches : une tâche de discrimination du slant et une tâche de discrimination du tilt. Chacune est réalisée sur les 5 types de stimuli décrits précédemment.

7 sujets ont pris part à l'ensemble des tests. Les résultats de chaque expérience sont obtenus sur 5 sujets. L'auteur de ce document et 2 sujets naifs par rapport au but de l'expérience prirent part à tous les tests. Les 4 sujets restants n'ont pris part qu'à une seule des deux séries d'expériences (discrimination du slant ou discrimination du tilt uniquement). Les sujets avaient une vue normale ou corrigée pour leur assurer une bonne acuité visuelle.

Le paradigme utilisé était une expérience standard avec deux choix possibles imposés (*2 Alternative Forced Choice*) (Figure 5.19). Pour chaque essai, une croix de fixation apparaît sur un fond gris au début durant 500ms. Le premier stimulus est présenté durant 300ms suivi par un écran gris avec la croix de fixation durant 600ms. Le second stimulus est affiché durant 300ms suivi également par un écran gris et la croix de fixation durant 600ms. L'un des stimuli est un stimulus de référence et l'autre est un stimulus test, l'ordre étant aléatoire (voir ci-

dessous). Finalement afin d'éliminer l'activation résiduelle des photorécepteurs (rémanence rétinienne) présente après le retrait du stimulus du champ visuel et pouvant induire une adaptation à la fréquence et à l'orientation présentées à l'essai précédent, un masque est présenté juste après la disparition du stimulus (Figure 5.18). Il se compose d'un ensemble de masques de Gabor avec des fréquences et des orientations aléatoires indépendamment de leur position spatiale. De cette manière toute adaptation possible à l'une ou l'autre des variations étudiées est éliminée, assurant que la perception du stimulus suivant se fasse exactement dans les mêmes conditions que le précédent.

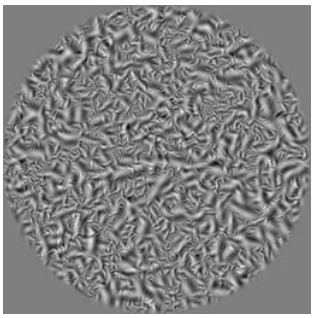


FIG. 5.18 – Exemple de masque utilisé pour réduire l'influence sur le prochain stimulus des fréquences et orientations de l'image présentée à l'essai précédent.

Chaque expérience est associée à un seul type de stimulus (un type de texture). Chaque comparaison est répétée 20 fois conduisant à un total de 360 essais par expérience. Chaque expérience est divisée en 4 blocs de 90 essais et dure environ 25 minutes. Au cours d'un bloc d'essais, l'instruction est donné au sujet de maintenir son attention sur la croix de fixation pour éviter une exploration spatiale éventuelle des stimuli. Il n'a pas de limitation de temps pour donner sa réponse, l'essai suivant commençant dès que celle-ci est enregistrée au clavier.

Pour toutes les expériences, trois angles de référence sont utilisés pour les angles de slant et pour les angles de tilt et pour chacun sont associées six valeurs différentes d'angles de test (Tableau 5.1).

	références	tests					
Slant	27^o	19.5^o	22.5^o	25.5^o	31.5^o	34.5^o	37.5^o
	40^o	32.5^o	35.5^o	38.5^o	44.5^o	47.5^o	50.5^o
	53^o	45.5^o	48.5^o	51.5^o	54.5^o	57.5^o	60.5^o
Tilt	0^o	-7.5^o	-4.5^o	-1.5^o	1.5^o	4.5^o	7.5^o
	45^o	37.5^o	40.5^o	43.5^o	46.6^o	49.5^o	52.5^o
	90^o	82.5^o	85.5^o	88.5^o	91.5^o	94.5^o	97.5^o

TAB. 5.1 – Tableau présentant les valeurs de références des angles de slant et de tilt ainsi que les valeurs des angles de tests associé utilisés pour les expériences de discrimination.

Lors d'une expérience de discrimination du slant, le sujet reçoit l'instruction d'indiquer la surface qu'il a perçue comme étant la plus inclinée en profondeur. Il indique ainsi si sa réponse est la première ou la deuxième surface en appuyant sur deux touches différentes. Lors d'une expérience de discrimination du tilt, le sujet reçoit l'instruction d'indiquer si la

deuxième surface perçue tourne ou non dans le sens horaire par rapport à la première surface. Il appuie ainsi sur la touche correspondant à la flèche pointant vers la droite si la rotation est dans le sens horaire ou sur la flèche pointant vers la gauche si la rotation est dans le sens anti-horaire. Quelques essais préliminaires sont effectués pour s'assurer que le sujet à bien compris la tâche.

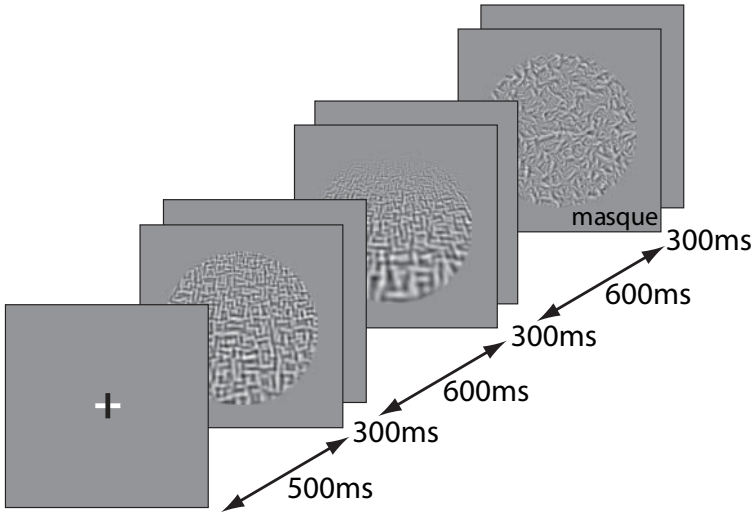


FIG. 5.19 – Schéma temporel de la procédure pour la réalisation d'un essai.

5.2.3 Analyse des résultats

Les données sont analysées séparément pour chaque observateur et chaque type de stimulus. Pour chaque valeur de l'angle de référence, nous calculons la probabilité que le stimulus de test soit perçu avec un angle plus important que l'angle du stimulus de référence. Ces probabilités sont estimées par des fonctions psychométriques, modélisées par des gaussiennes cumulées ([WH1a],[WH1b]).

Pour la tâche de discrimination du slant, la courbe représente la probabilité que l'inclinaison du stimulus test soit plus importante que celle du stimulus de référence.

Pour la tâche de discrimination du tilt, la courbe représente la probabilité que l'orientation du stimulus test soit plus importante que celle du stimulus de référence dans le sens horaire.

Deux paramètres sont extraits de ces fonctions : le point d'égalité subjective (*PSE : Point of Subjective Equality*) et la pente de la fonction en ce point.

La pente correspond en réalité à l'écart-type du modèle de gaussiennes cumulées utilisé pour tracer la courbe psychométrique. Plus la discrimination est facile entre la référence et le test, plus la pente de la courbe psychométrique est importante et proche de la verticale ce qui correspond à une diminution de l'écart-type du modèle. Aussi l'écart-type est utilisé également comme un index reflétant la fiabilité de l'indice étudié pour la réalisation de la tâche considérée.

5.2.4 Expérience de discrimination du slant

5.2.4.1 Résultats en présence des indices seuls et en combinaison

La figure 5.20 présente l'ensemble des courbes psychométriques obtenues sur la tâche de discrimination du slant par les 5 sujets suivant les 3 valeurs du slant de référence (27^o (surfaces frontoparallèles), 40^o et 53^o(surfaces très inclinées)) et les 3 types de stimuli (variation de fréquence seule, variation d'orientation seule et combinaison des deux variations).

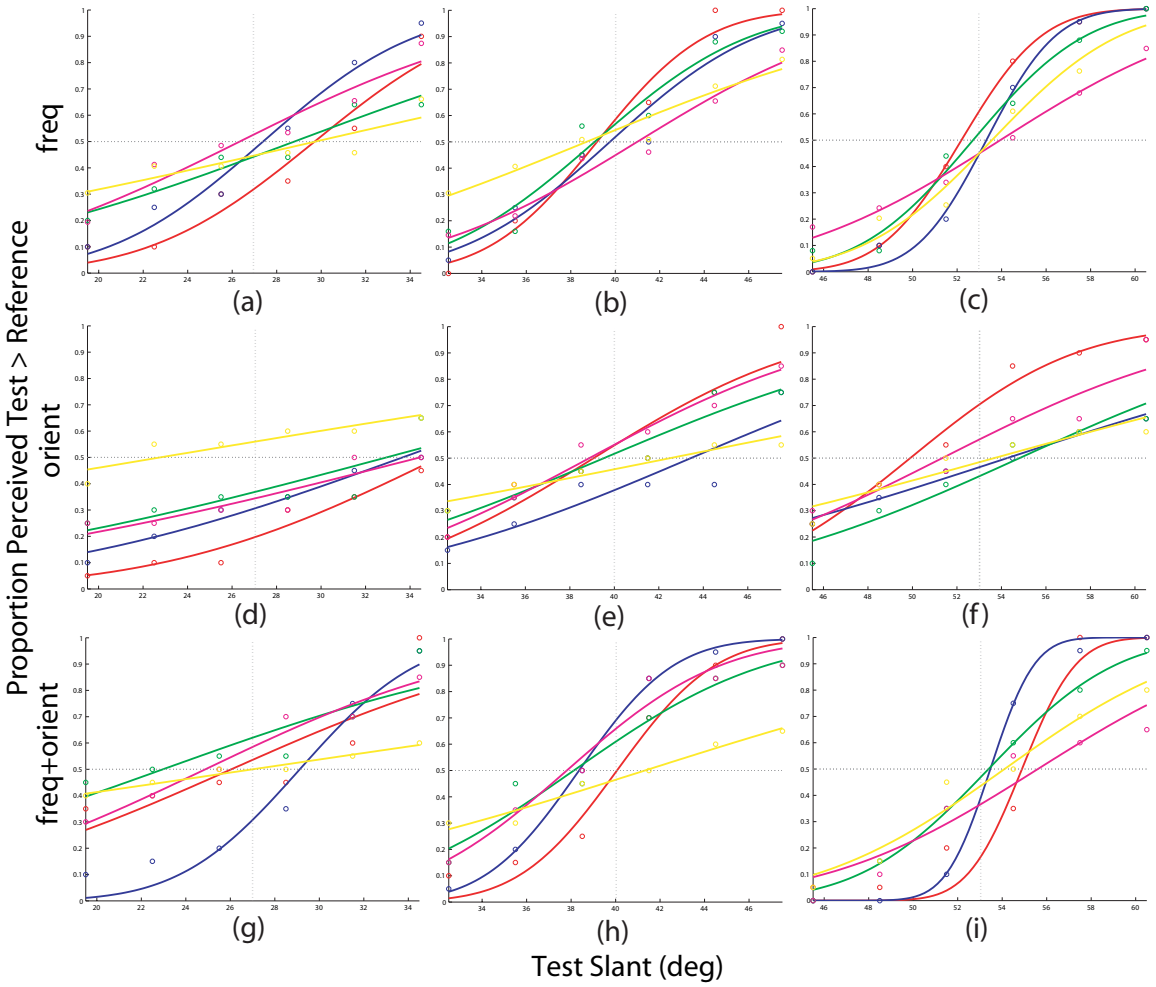


FIG. 5.20 – Résultats de discrimination du slant sur les 5 sujets ; chaque colonne présente les résultats correspondant à chacune des valeurs du slant de référence (27^o,40^o et 53^o) ; chaque courbe psychométrique montre les performances de discrimination obtenues par un sujet en fonction de la valeur du slant de test (voir Tableau 5.1) ; première ligne : résultats sur les textures présentant uniquement une variation de fréquence ; seconde ligne : résultats sur les textures présentant uniquement une variation d'orientation ; dernière ligne : résultats sur les textures présentant les deux types de variation.

Les performances de discrimination obtenues par tous les observateurs sont très similaires pour les différentes conditions.

Pour synthétiser les résultats, la figure 5.21 présente les performances de discrimination du slant moyennées sur les 5 sujets. Pour chacun des 3 types de texture, les 3 courbes psychométriques représentent les résultats obtenus pour chaque valeur du slant de référence.

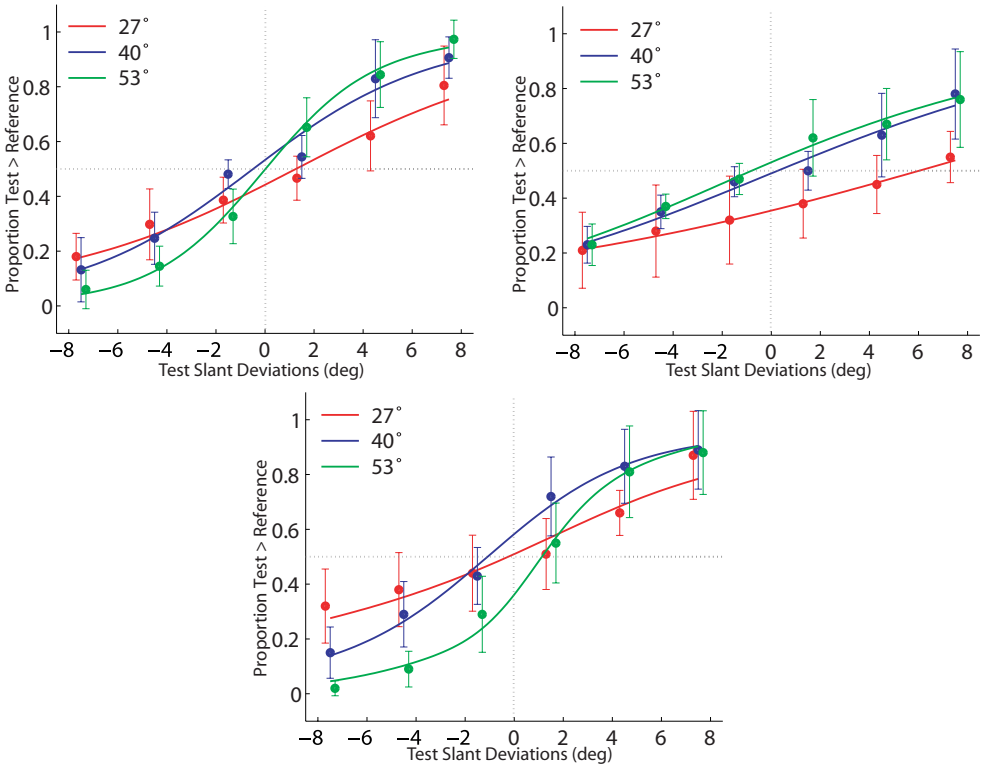


FIG. 5.21 – Résultats de discrimination du slant moyennés sur les 5 sujets ; chaque courbe montre les performances moyennes de discrimination obtenues par les 5 sujets en fonction de la valeur du slant de référence (27^o,40^o et 53^o) (accompagnées de leur erreurs-type) ; première ligne à gauche : résultats sur les textures présentant uniquement une variation de fréquence ; à droite : résultats sur les textures présentant uniquement une variation d'orientation ; deuxième ligne au centre : résultats sur les textures présentant les deux types de variation.

La figure 5.22 montre les biais et les pentes moyennes des courbes psychométriques obtenus pour chaque valeur du slant de référence et pour chaque type de texture.

Pour les 3 types de texture, nous observons une augmentation systématique des performances avec l'augmentation de la valeur du slant de référence. Ce résultat est conforme aux résultats des travaux antérieurs (voir Chapitre 4.3).

Aucun biais n'apparaît sur aucune des textures pour les inclinaisons moyennes et fortes de la surface. Un léger biais apparaît sur les surfaces proches d'une inclinaison frontoparallèle. Il est plus important pour les textures présentant une variation d'orientation (en présence ou non d'une variation de fréquence). Ce biais peut être dû simplement aux mauvaises performances de discrimination du slant obtenues en présence de ce type d'indice.

De bonnes performances sont obtenues sur les textures présentant une variation de fréquence. Par contre de mauvaises performances sont obtenues sur les textures présentant uniquement la variation d'orientation. Des performances équivalentes sont obtenues sur les tex-

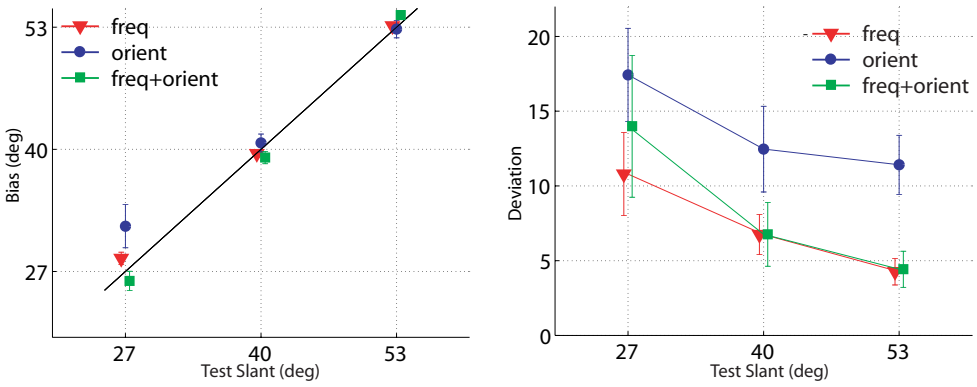


FIG. 5.22 – Analyse des résultats de la discrimination du slant moyennés sur les 5 sujets ; à gauche : biais pour chaque valeur du slant de référence et pour chaque type de texture ; à droite : pente moyenne de chaque courbe psychométrique (au *point d'égalité subjective* (PSE)) en fonction de la valeur du slant de référence et pour chaque type de texture.

tures présentant uniquement une variation de fréquence et celles combinant les deux types de variations pour des valeurs moyennes et grandes du slant. Pour des valeurs petites du slant, un effet de perturbation entre les indices est observé sur les textures présentant la combinaison des deux variations. Pour ce type de texture, les performances obtenues sont meilleurs que celles obtenues avec les textures présentant uniquement une variation d'orientation et elles sont moins bonnes que celles obtenues avec les textures présentant uniquement une variation de fréquence.

En résumé, sur une tâche de discrimination du slant : plus la valeur du slant de la surface est importante, plus les observateurs sont performants pour discriminer entre deux valeurs ; l'indice de variation de fréquence contribue à la bonne perception de surface inclinée ; l'indice de variation d'orientation ne contribue que faiblement à la perception de surface inclinée.

5.2.4.2 Résultats en présence des indices en conflit

Les mêmes expériences sont menées avec des stimuli présentant un conflit entre les informations transmises par les deux types de variation.

La figure 5.23 présente l'ensemble des courbes psychométriques obtenues sur la tâche de discrimination du slant par les 5 sujets suivant les 3 valeurs du slant de référence (27^o (surfaces frontoparallèles), 40^o et 53^o(surfaces très inclinées)) et les 2 types de stimuli présentant un conflit d'indices. La première ligne de la figure 5.23 présente les performance obtenues pour des stimuli présentant une variation de fréquence (fréquence variable) mais en conflit avec une variation d'orientation fixée à la valeur du slant de référence (les stimuli utilisés sont ceux de la figure 5.13). De même la seconde ligne de la figure 5.23 présente les performance obtenues pour des stimuli présentant une variation d'orientation (orientation variable) mais en conflit avec une variation de fréquence fixée à la valeur du slant de référence (les stimuli utilisés sont ceux de la figure 5.14).

Sur l'ensemble des courbes psychométriques, les observateurs obtiennent des performances de discrimination similaires.

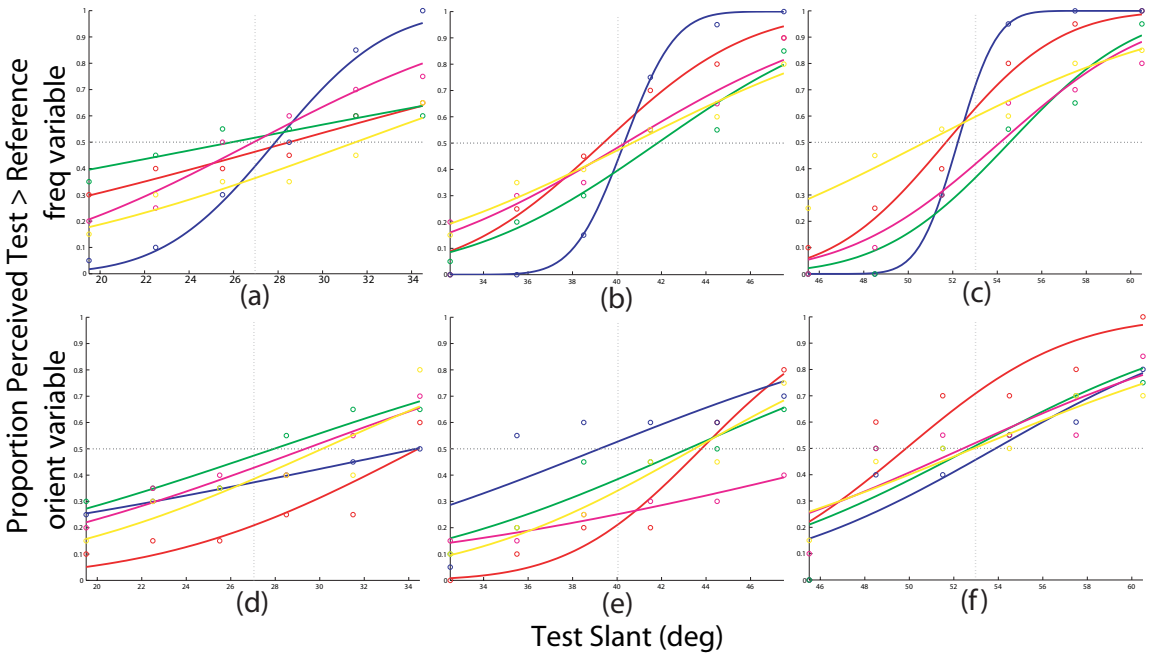


FIG. 5.23 – Résultats de discrimination du slant sur les 5 sujets en cas de conflit d'indices ; chaque colonne présente les résultats correspondant à chacune des valeurs du slant de référence (27^o,40^o et 53^o) ; chaque courbe psychométrique montre les performances de discrimination obtenues par un sujet en fonction de la valeur du slant de test (voir Tableau **??**) ; première ligne : résultats sur les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du slant de référence ; seconde ligne : résultats sur les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du slant de référence.

Pour synthétiser les résultats, la figure 5.24 présente les performances de discrimination du slant moyennées sur les 5 sujets. Pour chacun des 2 types de texture correspondant aux deux cas de conflit, les 3 courbes psychométriques représentent les résultats obtenus pour chaque valeur du slant de référence.

La figure 5.25 montre les biais et les pentes moyennes des courbes psychométriques obtenus pour chaque valeur du slant de référence et pour chaque type de texture. La deuxième ligne de la figure 5.25 montre l'ensemble des pentes moyennes obtenues pour tous les types de texture étudiés.

Nous observons de nouveau une augmentation systématique des performances avec l'augmentation de la valeur du slant de référence.

Aucun biais n'apparaît sur aucune des textures pour les inclinaisons fortes de la surface. Un biais plus important est trouvé pour les inclinaisons faibles et moyennes. Ce biais est plus important pour les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du slant de référence. Il peut être simplement dû aux mauvaises performances de discrimination du slant obtenues en présence de ce type de conflit mettant en jeu l'indice de variation d'orientation.

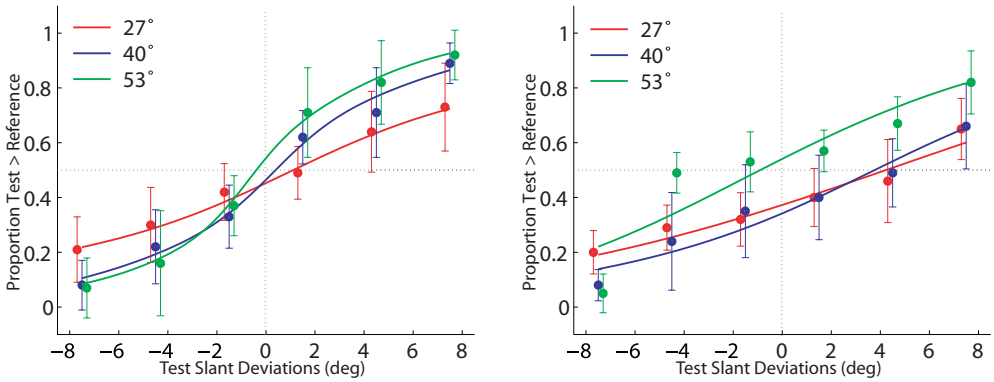


FIG. 5.24 – Résultats de discrimination du slant moyennés sur les 5 sujets en cas de conflit d'indices ; chaque courbe montre les performances moyennes de discrimination obtenues par les 5 sujets en fonction de la valeur du slant de référence (27^o,40^o et 53^o) (accompagnées de leur erreurs-type) ; première ligne à gauche : résultats sur les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du slant de référence ; à droite : résultats sur les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du slant de référence.

De bonnes performances sont obtenues sur les textures présentant une variation de fréquence en accord avec le slant. De mauvaises performances sont obtenues sur les textures présentant une variation d'orientation en accord avec le slant.

La deuxième ligne de la figure 5.25 présente l'ensemble des écart-types obtenus sur les 5 types de textures. Nous observons principalement que les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée induisent de meilleurs performances que les textures présentant uniquement une variation d'orientation. Exactement les mêmes performances sont obtenues entre les textures présentant la combinaison des deux variations et les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée ; les mêmes performances sont obtenues pour des inclinaisons moyennes et fortes de la surface et le même effet de perturbation entre les indices est obtenu pour des inclinaisons faibles. Les meilleurs performances sur toutes la valeur de référence du slant sont obtenues pour les textures présentant uniquement une variation de fréquence.

En résumé, sur une tâche de discrimination du slant l'indice de variation de fréquence apparaît comme étant l'indice contribuant principalement à la perception de surface inclinée. L'indice de variation d'orientation ne contribue que très faiblement. Le slant est donc principalement estimé à partir de l'indice de variation de fréquence.

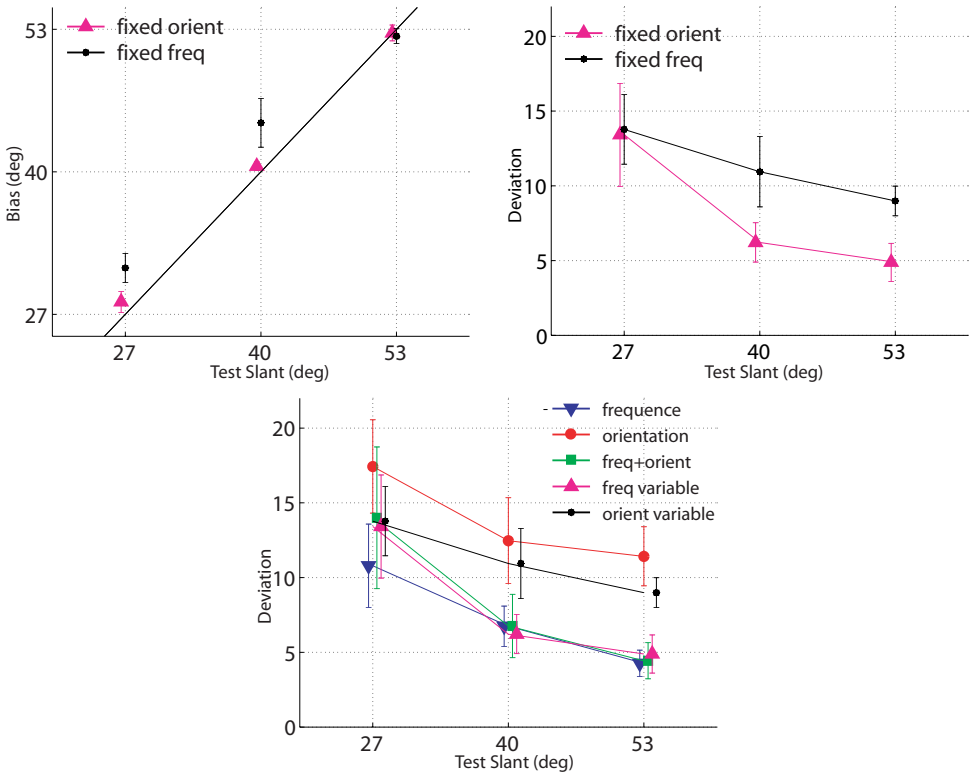


FIG. 5.25 – Analyse des résultats de la discrimination du slant moyennés sur les 5 sujets en cas de conflit d'indices ; première ligne à gauche : biais pour chaque valeur du slant de référence et pour les 2 types de texture présentant un conflit ; à droite : pente moyenne de chaque courbe psychométrique (au *point d'égalité subjective* (PSE)) en fonction de la valeur du slant de référence et pour les 2 types de texture présentant un conflit ; deuxième ligne au centre : ensemble des pentes moyennes obtenues sur les 5 types de texture (variation de fréquence, variation d'orientation, combinaison des deux indices, 2 types de conflit).

5.2.5 Expérience de discrimination du tilt

5.2.5.1 Résultats en présence des indices seuls et en combinaison

La figure 5.26 présente l'ensemble des courbes psychométriques obtenues sur la tâche de discrimination du tilt par les 5 sujets suivant les 3 valeurs du tilt de référence (0^o (surfaces murales), 45^o et 90^o(surfaces de sols)) et les 3 types de stimuli (variation de fréquence seule, variation d'orientation seule et combinaison des deux variations).

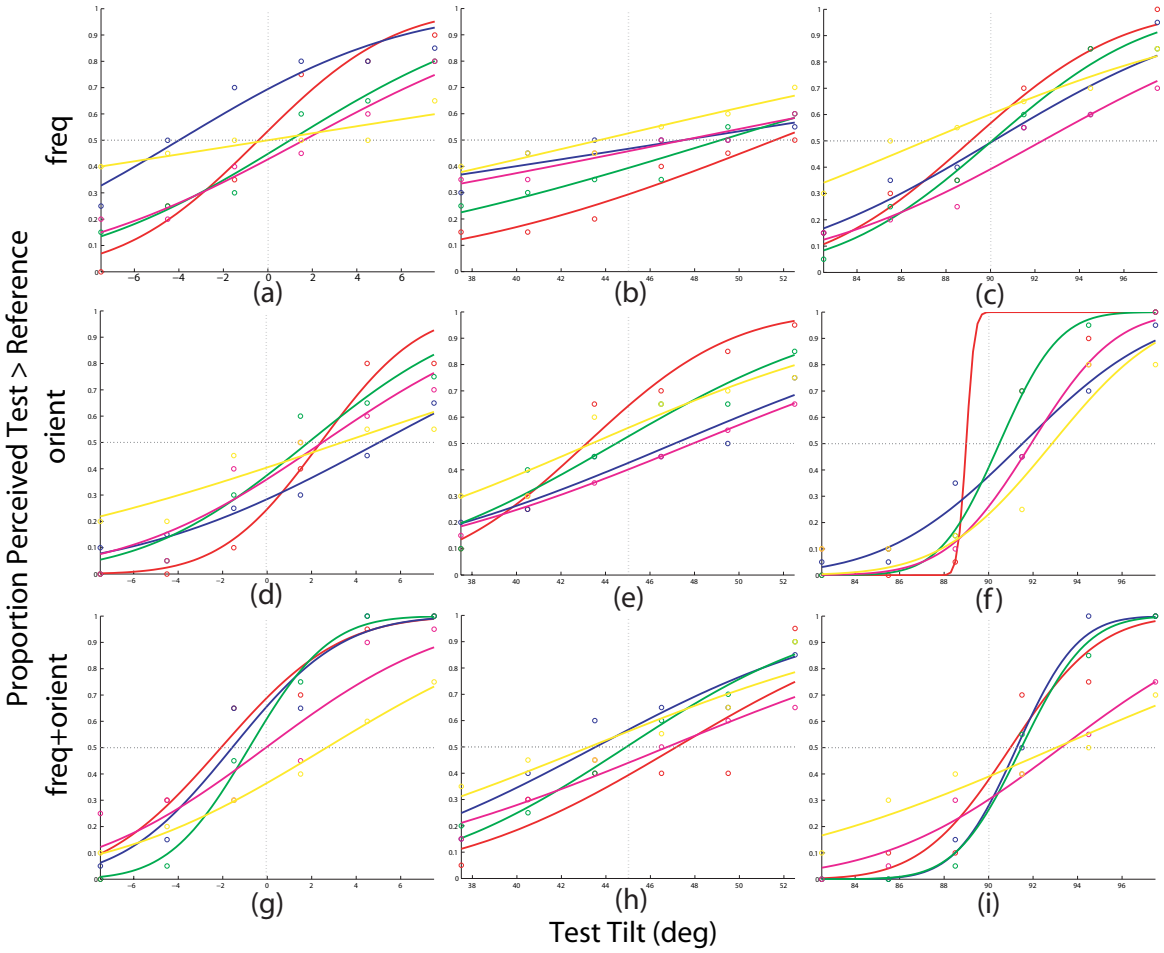


FIG. 5.26 – Résultats de discrimination du tilt sur les 5 sujets ; chaque colonne présente les résultats correspondant à chacune des valeurs du tilt de référence (0^o,45^o et 90^o) ; chaque courbe psychométrique montre les performances de discrimination obtenues par un sujet en fonction de la valeur du tilt de test (voir Tableau 5.1) ; première ligne : résultats sur les textures présentant uniquement une variation de fréquence ; seconde ligne : résultats sur les textures présentant uniquement une variation d'orientation ; dernière ligne : résultats sur les textures présentant les deux types de variation.

Les performances de discrimination obtenues par tous les observateurs sont très similaires pour les différentes conditions.

Pour synthétiser les résultats, la figure 5.27 présente les performances de discrimination du tilt moyennées sur les 5 sujets. Pour chacun des 3 types de texture, les 3 courbes psychométriques représentent les résultats obtenus pour chaque valeur du tilt de référence.

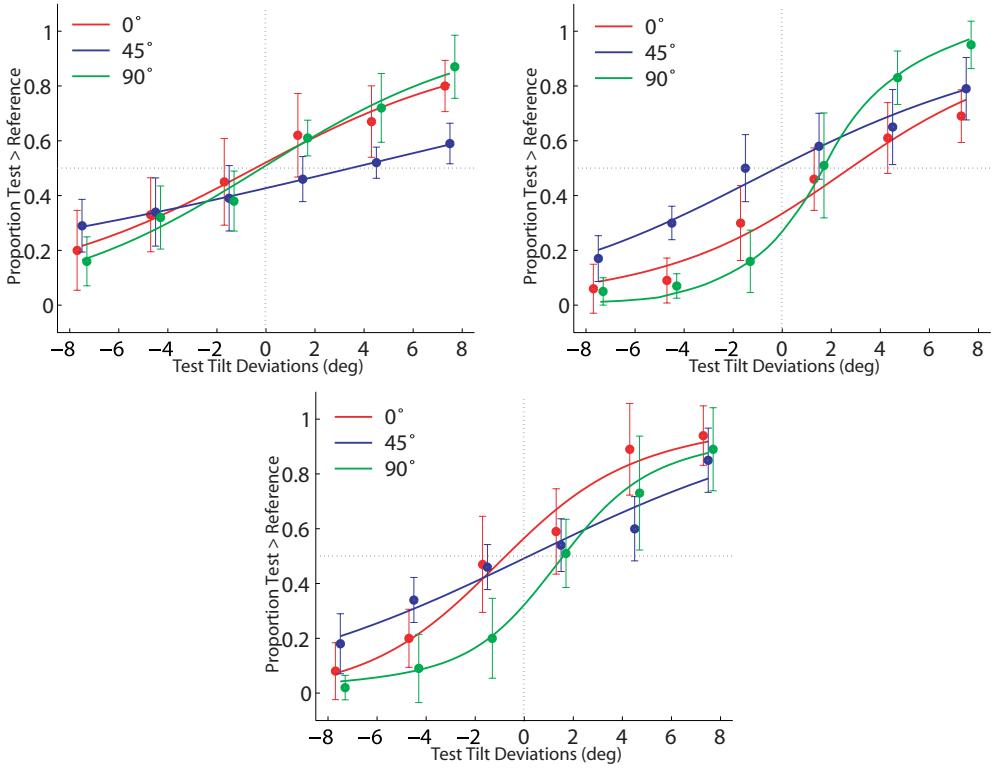


FIG. 5.27 – Résultats de discrimination du tilt moyennés sur les 5 sujets ; chaque courbe montre les performances moyennes de discrimination obtenues par les 5 sujets en fonction de la valeur du tilt de référence (27^o,40^o et 53^o) (accompagnées de leur erreurs-type) ; première ligne à gauche : résultats sur les textures présentant uniquement une variation de fréquence ; à droite : résultats sur les textures présentant uniquement une variation d'orientation ; deuxième ligne au centre : résultats sur les textures présentant les deux types de variation.

La figure 5.28 montre les biais et les pentes moyennes des courbes psychométriques obtenus pour chaque valeur du tilt de référence et pour chaque type de texture.

Sur les 3 textures, nous observons une augmentation systématique des performances pour un tilt à 0^o et 90^o avec les meilleurs résultats à 90^o. Ce résultat est conforme aux résultats des travaux antérieurs (voir Chapitre 4.3).

Aucun biais n'apparait sur aucune des textures pour toutes les valeurs de tilt comme attendu.

De très bonnes performances sont obtenues pour les textures présentant uniquement une variation d'orientation et notamment pour les surfaces orientées similairement à des sols (tilt à 90^o). De moins bonnes performances sont obtenues sur les textures présentant uniquement une variation de fréquence, excepté pour un tilt à 90^o où les performances sont proches des meilleurs obtenues. Les mêmes performances de discrimination sont obtenues à 45^o pour les textures présentant une variation d'orientation avec ou non une variation de fréquence

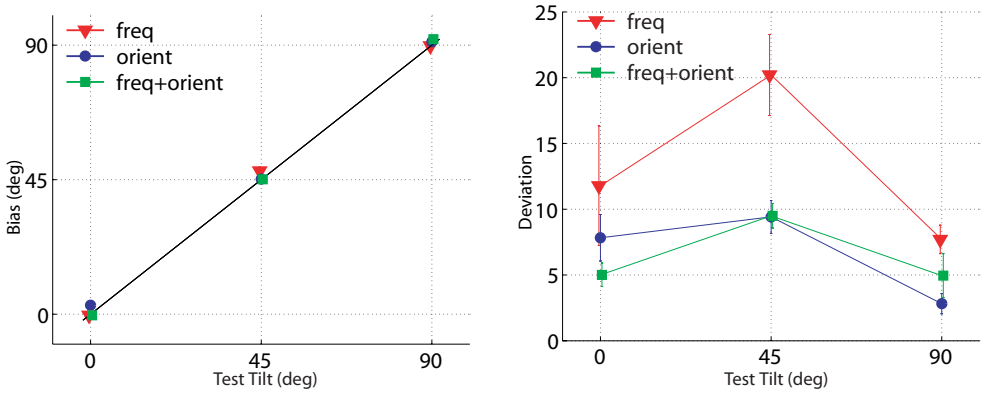


FIG. 5.28 – Analyse des résultats de la discrimination du tilt moyennés sur les 5 sujets ; à gauche : biais pour chaque valeur du tilt de référence et pour chaque type de texture ; à droite : pente moyenne de chaque courbe psychométrique (au *point d'égalité subjective* (PSE)) en fonction de la valeur du tilt de référence et pour chaque type de texture.

associée. Pour les surfaces murales (tilt à 0^o), une diminution des performances est observée pour les textures présentant uniquement une variation d'orientation tandis que sur les textures présentant en plus une variation de fréquence en combinaison, les performances sont meilleurs et restent aussi bonnes que pour les surfaces à 90^o.

5.2.5.2 Résultats avec les indices en conflit

Les mêmes expériences sont menées avec des stimuli présentant un conflit entre les informations transmises par les deux types de variation.

La figure 5.29 présente l'ensemble des courbes psychométriques obtenues sur la tâche de discrimination du tilt par les 5 sujets suivant les 3 valeurs du tilt de référence (0^o (surfaces murales), 45^o et 90^o(surfaces de sols)) et les 2 types de stimuli présentant un conflit d'indices. La première ligne de la figure 5.29 présente les performance obtenues pour des stimuli présentant une variation de fréquence (fréquence variable) mais en conflit avec une variation d'orientation fixée à la valeur du tilt de référence (les stimuli utilisés sont ceux de la figure 5.15). De même la seconde ligne de la figure 5.29 présente les performance obtenues pour des stimuli présentant une variation d'orientation (orientation variable) mais en conflit avec une variation de fréquence fixée à la valeur du tilt de référence (les stimuli utilisés sont ceux de la figure 5.16).

Sur l'ensemble des courbes psychométriques, les observateurs obtiennent des performances de discrimination similaires.

Pour synthétiser les résultats, la figure 5.30 présente les performances de discrimination du tilt moyennées sur les 5 sujets. Pour chacun des 2 types de texture correspondant aux deux cas de conflit, les 3 courbes psychométriques représentent les résultats obtenus pour chaque valeur du tilt de référence.

La figure 5.31 montre les biais et les pentes moyennes des courbes psychométriques obtenus pour chaque valeur du tilt de référence et pour chaque type de texture. La deuxième ligne de la figure 5.31 montre l'ensemble des pentes moyennes obtenues pour tous les types de texture étudiés.

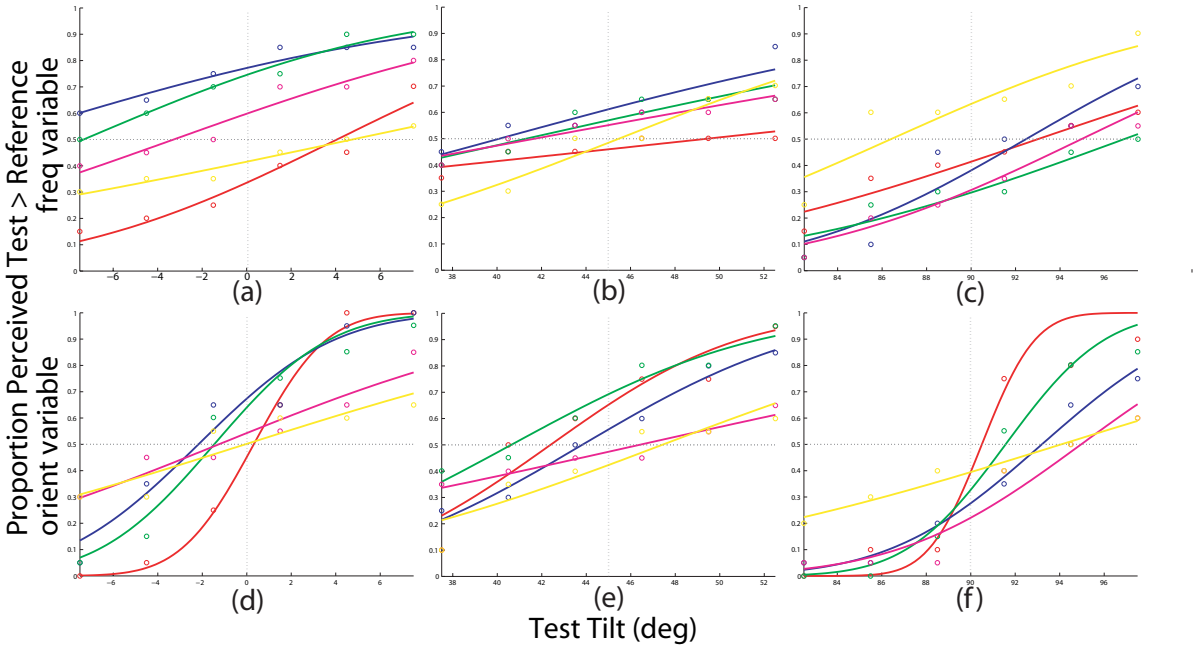


FIG. 5.29 – Résultats de discrimination du tilt sur les 5 sujets en cas de conflit d'indices ; chaque colonne présente les résultats correspondant à chacune des valeurs du tilt de référence (0^o,45^o et 90^o) ; chaque courbe psychométrique montre les performances de discrimination obtenues par un sujet en fonction de la valeur du tilt de test (voir Tableau 5.1) ; première ligne : résultats sur les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du tilt de référence ; seconde ligne : résultats sur les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du tilt de référence.

Nous observons de nouveau une augmentation systématique des performances est obtenue pour un tilt à 0^o et 90^o avec les meilleurs résultats à 90^o.

Aucun biais n'apparaît sur aucune des textures et sur les différentes valeurs du tilt de référence comme attendu.

De bonnes performances sont obtenues sur les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du tilt de référence. Une diminution des performances est obtenue sur les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du tilt de référence.

De très bonnes performances sont obtenues sur les textures présentant une variation d'orientation en accord avec le tilt. De moins bonnes performances sont obtenues sur les textures présentant une variation de fréquence en accord avec le tilt mais sont suffisantes pour faire une estimation du tilt.

La deuxième ligne de la figure 5.31 présente l'ensemble des écart-types obtenus sur les 5 types de textures.

Les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du tilt de référence induisent une diminution équivalente des performances pour chaque valeur de tilt par rapport à celles obtenues avec les textures présentant les deux types de variation en combinaison ; par contre les performances sont les mêmes que sur les textures présentant uniquement une variation d'orientation pour un tilt à 0^o. Les tex-

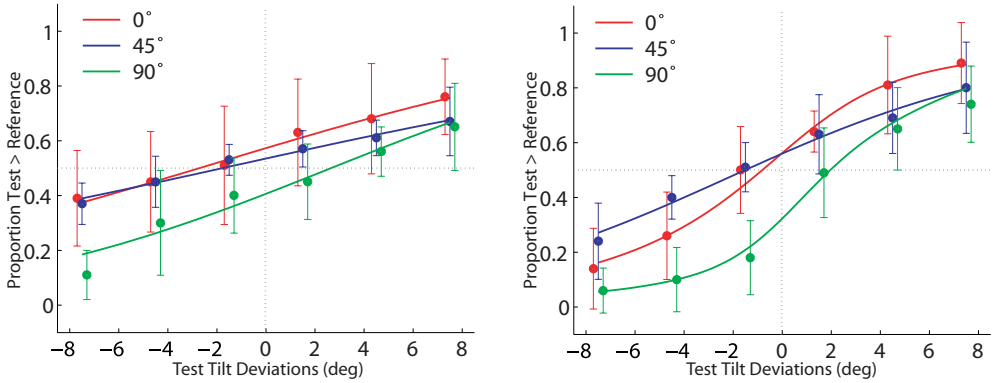


FIG. 5.30 – Résultats de discrimination du tilt moyennés sur les 5 sujets en cas de conflit d'indices ; chaque courbe montre les performances moyennes de discrimination obtenues par les 5 sujets en fonction de la valeur du tilt de référence (0^o,45^o et 90^o) (accompagnées de leur erreurs-type) ; première ligne à gauche : résultats sur les textures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du tilt de référence ; à droite : résultats sur les textures présentant une variation d'orientation en conflit avec une variation de fréquence fixée à la valeur du tilt de référence.

tures présentant une variation de fréquence en conflit avec une variation d'orientation fixée à la valeur du tilt de référence induisent une diminution équivalente des performances pour chaque valeur de tilt par rapport à celles obtenues avec les textures présentant uniquement une variation de fréquence. Les deux types de conflit montrent un effet de perturbation similaire sur l'ensemble des valeurs de tilt testés.

En résumé, sur une tâche de discrimination du tilt l'indice de variation d'orientation apparaît comme étant l'indice contribuant principalement à la perception de surface inclinée. L'indice de variation de fréquence permet cependant une bonne approximation sans atteindre le même niveau de précision. Le tilt peut donc être estimé avec les deux types d'indices.

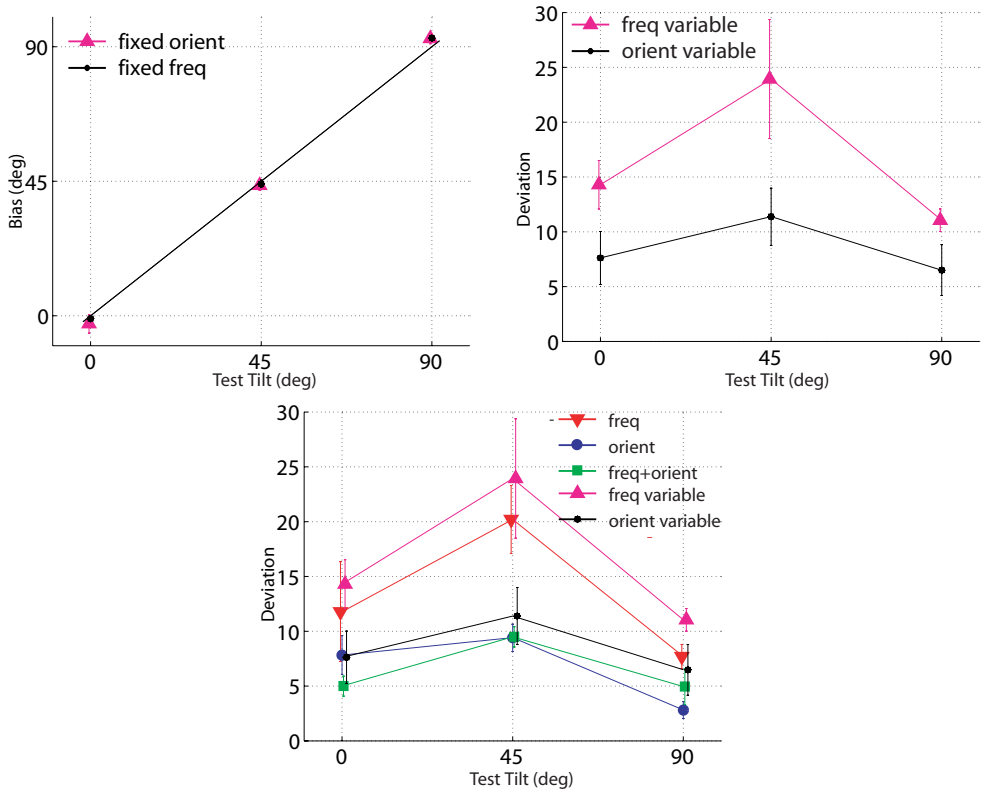


FIG. 5.31 – Analyse des résultats de la discrimination du tilt moyennés sur les 5 sujets en cas de conflit d'indices ; première ligne à gauche : biais pour chaque valeur du tilt de référence et pour les 2 types de texture présentant un conflit ; à droite : pente moyenne de chaque courbe psychométrique (au *point d'égalité subjective* (PSE)) en fonction de la valeur du tilt de référence et pour les 2 types de texture présentant un conflit ; deuxième ligne au centre : ensemble des pentes moyennes obtenues sur les 5 types de texture (variation de fréquence, variation d'orientation, combinaison des deux indices, 2 types de conflit).

5.3 Discussion

Nous présentons ici une analyse des expériences et des résultats que nous avons obtenus. Nous discutons de la validité de nos stimuli pour l'étude des indices de variation de fréquence et de perspective linéaire. Nous argumentons en faveur de l'analyse séparée des deux indices opérée par le système visuel pour la perception 3D en mettant en évidence l'information spécifiquement traitée par chacun des indices. Nous discutons également de la validité de l'hypothèse d'isotropie et de l'hypothèse émise par Li et Zaidi par rapport à nos stimuli. Nous introduisons enfin une perspective vers l'étude séparée des indices de variation de fréquence, de perspective linéaire et de courbure.

5.3.1 Validation des stimuli

La principale contribution de ce travail est la création de stimuli valides pour l'étude de la perception 3D à partir de l'information de texture. Ces stimuli permettent de manipuler séparément deux indices : le gradient de fréquence et la perspective linéaire. Bien que différents travaux aient fait mention de ces deux indices pour la perception 3D [LZ04] [TTD05] [OML03], aucun n'a explicitement étudié la contribution des deux indices indépendamment l'un de l'autre.

Les performances de discrimination obtenues sur les stimuli sont bien porteurs d'indices liés à la perception 3D de surfaces planes. La variation de fréquence est perçue avec suffisamment d'acuité pour obtenir une bonne précision pour la discrimination des inclinaisons et pour détecter un changement d'orientation.

La variation d'orientation, relative à la perspective linéaire, est perçue également avec suffisamment d'acuité pour obtenir une bonne précision pour la discrimination entre différentes orientations de la surface. Cependant nos expériences mettent en évidence que cet indice seul ne participe pas ou très faiblement à la perception de l'inclinaison d'une surface.

La définition des indices de fréquence et d'orientation permet d'obtenir une description paramétrique de l'information de texture. Celle-ci peut être utilisée pour estimer la 3D mais également pour décrire la texture. Il serait alors possible de reprendre l'explication de la classification obtenue par Rosas *et al.* La valeur de la fréquence moyenne pourrait expliquer les mauvaises performances sur les textures de type bruit qui sont très basses fréquences (bruit $1/f$, bruit de Perlin) et les bonnes performances sur les textures présentant des moyennes et hautes fréquences (léopard, points Polka). Cependant pour effectuer cette étude, comme le note Rosas *et al*, l'information de fréquence dépend de la méthode d'analyse du spectre d'amplitude (par exemple par la recherche de pics de fréquence [SF95], par l'analyse de la pente du spectre ou par la combinaison des réponses de filtres (par exemple suivant notre modèle présenté au chapitre 6). Une analyse psychophysique devra alors s'appuyer sur un modèle particulier d'extraction de l'information de fréquence pour rendre compte de l'ordre de facilitation des textures.

5.3.2 Discrimination du slant

Les résultats obtenus sur la tâche de discrimination du slant montrent clairement une amélioration des performances avec l'augmentation de l'inclinaison. Ceci est en accord avec les résultats de la littérature (voir Chapitre 4.3). Cet effet est surtout constaté pour les

textures présentant une variation de fréquence. La présence en combinaison ou en conflit d'une variation d'orientation n'affecte pas les performances pour les valeurs du slant moyennes et importantes (à peu près à partir de 40^o) par rapport à celles obtenues en présence uniquement de la variation de fréquence. En dessous de cette valeur, les variations sont trop faibles et des perturbations peuvent apparaître en cas de présence de la variation d'orientation.

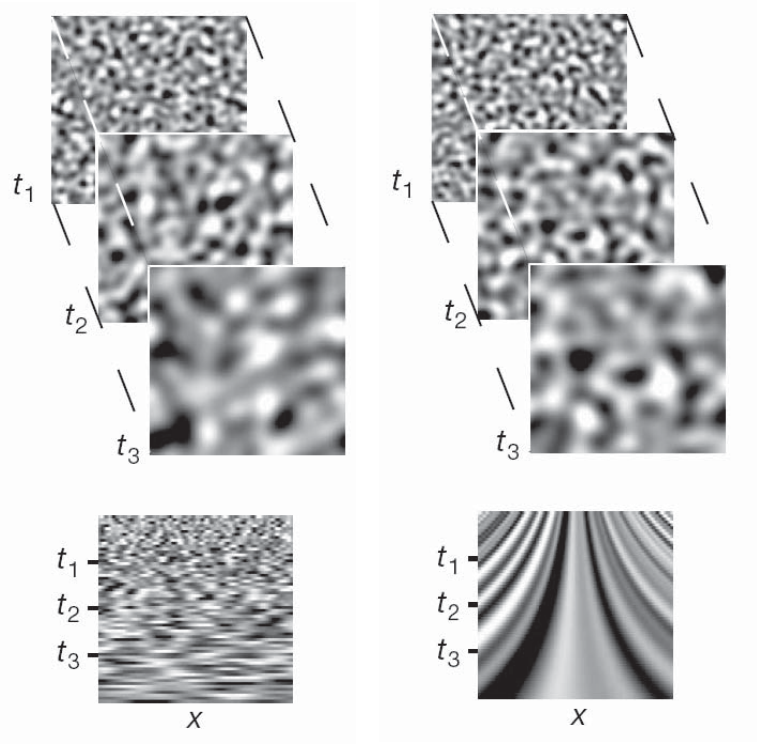


FIG. 5.32 – Exemples de stimuli utilisés par Schrater *et al* (tirés de [SKS01]) ; en haut : exemples de différentes images présentées à différents instants dans une séquence vidéo ; en bas : sections spatiotemporelles ; à gauche : stimuli présentant une texture aléatoire (bruit blanc convolué avec filtres passe-bande successifs) ; la section spatiotemporelle associée illustre l'absence de flux optique (pas de corrélation spatiotemporelle des éléments de la texture, seule le changement d'échelle donne une information de zoom) ; à droite : stimuli présentant une texture aléatoire pour le premier, les stimuli suivants étant des versions zoomées successives ; la section spatiotemporelle associée illustre la présence d'un flux optique (dans ce cas à la fois le flux optique et le changement d'échelle donnent une information de zoom).

Dans nos conditions expérimentales, la perspective linéaire n'a aucun effet sur la précision de l'estimation de l'inclinaison d'une surface plane. En l'absence d'une variation de fréquence, les performances chutent de manière importante. En conflit avec la variation d'orientation, la variation de fréquence est l'indice qui influence le plus les performances obtenues sur cette tâche. Ceci est compatible avec les travaux antérieurs qui ont mis en évidence la relation entre la fréquence locale et l'estimation de la profondeur locale. Nous pouvons notamment citer les travaux de Schrater *et al* [SKS01] où les auteurs ont montré que le système visuel pouvait percevoir des mouvements de zoom par une unique analyse de l'indice de changement de

fréquence (échelle spatiale) dans le temps sans intervention de la corrélation spatiale (i.e sans calcul de flux optique) (Figure 5.32). Leur résultat met en évidence l'utilisation de la fréquence spatiale comme indice de profondeur. Dans nos stimuli cette variation de fréquence temporelle est équivalente à la variation de fréquence spatiale locale. Le système visuel mesurerait donc la variation de profondeur locale sur la surface pour estimer son inclinaison. En l'absence de changement de profondeur, la surface apparaît fronto-parallèle ou très peu inclinée en présence de perspective linéaire.

Il serait intéressant de confronter les performances obtenues avec un observateur idéal utilisant la variation de fréquence. Il serait alors possible de mettre en relation ce résultat avec l'étude effectuée par Knill ([Kni98c] [Kni98a]) et de comparer ainsi la pertinence de l'indice de variation de fréquence face aux indices de compression et de changement de taille pour l'estimation de l'inclinaison d'une surface.

5.3.3 Hypothèse d'isotropie

Les stimuli présentant uniquement une variation de fréquence ont, par construction, une distribution uniforme des orientations en tout point de la surface, indépendamment de l'inclinaison (Figure 5.33 colonne de gauche). Ce type de texture correspont souvent aux textures naturelles présentant peu d'information d'orientation (par exemple un champ de tournesols). Par contre ce n'est pas le cas dans les textures présentant des orientations, mêmes discontinues (par exemple une texture d'osier) où par l'effet de la projection, les orientations tendent à prendre une direction privilégiées ([WM03] [WMeda] [WMedb] [AM04]) (Figure 5.33 colonne de droite).

Les résultats de nos expériences montrent que l'absence où la présence de ce biais dans les stimuli n'affecte pas les performances du système visuel dans une tâche de discrimination du slant (expériences avec les textures présentant uniquement une variation de fréquence ou présentant les deux types de variation). Ce biais dans les orientations semble donc ne pas intervenir dans l'estimation de l'inclinaison d'une surface plane ou requière des conditions particulières de présentation des stimuli pour pouvoir influencer les performances (par exemple avec des orientations plus marquées ou avec champ visuel très large). Cependant, d'après les résultats antérieurs, les conditions choisies pour nos expériences sont supposées permettre une bonne perception de l'inclinaison. Aussi ce biais dans la distribution des orientations semble posséder un poids très faible, dans un modèle de combinaison d'indices.

Une autre manière d'interpréter ces résultats est de considérer l'hypothèse d'isotropie (voir Chapitre 4.2). La variation d'orientation induit par la projection provoque une rupture de l'isotropie présente initialement dans la texture avant projection. Sur nos stimuli, dans le cas où seule la variation de fréquence est présente, l'isotropie est conservée en chaque position de la surface (Figure 5.33 colonne de gauche) et la perception de l'inclinaison est bonne. Lorsque seule la variation d'orientation est présente, c'est-à-dire seul l'indice de rupture de l'isotropie est présent (Figure 5.33 colonne du milieu), les performances diminuent fortement et la discrimination est difficile. Dans le dernier cas où les deux types de variation sont présents, il apparaît également une rupture de l'isotropie mêlée à la variation de fréquence (Figure 5.33 colonne de droite), la perception de l'inclinaison reste cependant la même, sans diminution des performances. De même que l'hypothèse de l'utilisation du bais sur les orientations, l'hypothèse d'isotropie semble donc également intervenir faiblement.

Sans pouvoir conclure définitivement sur l'utilisation de la rupture de l'isotropie comme indice 3D pour estimer l'inclinaison de surfaces planes, nous pouvons observer, d'après les résultats antérieurs et nos propres résultats, que cette hypothèse est souvent surpassée par la présence de gradients (notamment sur la variation de fréquence dans nos expériences ou le gradient de compression dans le cas des expériences de Knill [Kni98a]).

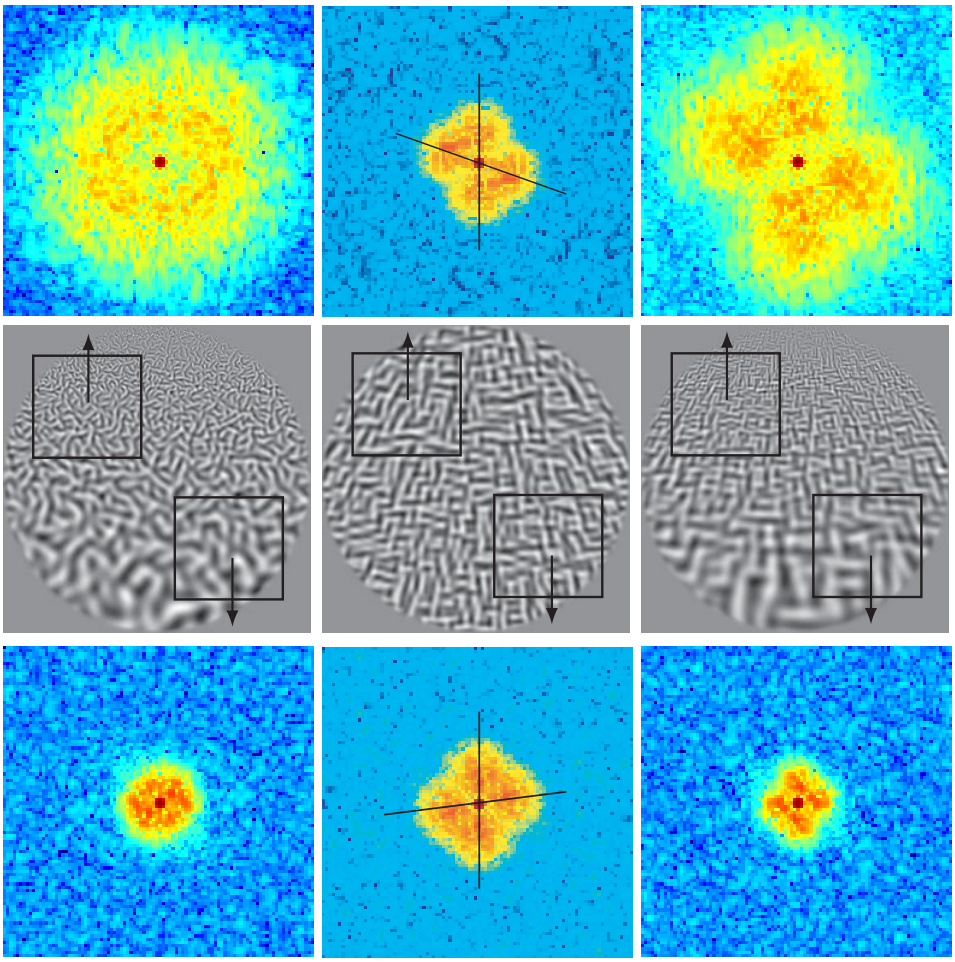


FIG. 5.33 – Spectres locaux moyens pris à différentes positions dans les différents stimuli utilisés dans nos expériences ; la première ligne correspond au spectre d'amplitude de l'imagette située dans la partie supérieure du stimulus présenté à la deuxième ligne ; la troisième ligne correspond au spectre d'amplitude de l'imagette située dans la partie inférieure droite du stimulus ; de gauche à droite : avec uniquement une variation de fréquence, l'isotropie est conservée (énergie du spectre répartie uniformément à toutes les orientations) ; avec uniquement une variation d'orientation, il apparaît une rupture de l'isotropie (les deux orientations principales tendent à prendre la même direction en fonction de la position spatiale analysée) ; avec les deux types de variation, il apparaît également une rupture de l'isotropie mêlée au changement de fréquence.

Cela va dans le sens de l'utilisation d'une hypothèse d'homogénéité. Cependant si l'hypothèse de rupture de l'isotropie intervient bien comme indice 3D, celui-ci n'existe essentielle-

ment qu'en présence d'un gradient de texture associé (dans nos expériences, la présence seule d'une variation d'orientation ne permet d'une faible discrimination entre les inclinaisons). Il est à noter que Knill, pour cette raison, ne fait intervenir l'hypothèse d'isotropie qu'en combinaison avec le gradient de compression. En effet comme il l'indique dans [Kni98c], la rupture d'isotropie est dû à l'effet de compression et de changement de taille induits par la projection perspective de la surface.

Les travaux antérieurs de Rosenholtz et Malik [RM97] ainsi que ceux de Knill [Kni98a] ont cependant montré que la rupture de l'isotropie influence la perception de l'inclinaison. Notamment si l'anisotropie est artificiellement augmentée, la valeur de l'inclinaison perçue augmente et elle diminue dans le cas inverse où l'anisotropie est artificiellement atténuée après projection (voir Chapitre 4.2). Nous proposons de réinterpréter ces données à travers l'utilisation de l'indice de variation de fréquence. En effet une pré-compression de la texture entraîne une augmentation de la fréquence moyenne globale de la texture. De même un étirement initial dans la direction du tilt entraîne un décalage vers les basses fréquences dans cette direction. Des expériences pourraient être conduites afin de déterminer si la fréquence moyenne influe sur les performances de discrimination.

Avec nos stimuli cela consisterait simplement à modifier la fréquence de l'ensemble des masques de Gabor (Figure 5.34). Cette expérience pourrait mettre en évidence non seulement une modification de la perception du slant mais également une modification des performances de discrimination. La variation de la fréquence moyenne pourrait induire une amélioration des performances vers les hautes fréquences et une diminution vers les basses fréquences. Ceci pourrait également expliquer l'amélioration des performances avec l'inclinaison du slant qui correspond également à un décalage vers les hautes fréquences. Si les résultats concordaient avec cette hypothèse, alors il serait possible d'avancer que, contrairement aux conclusions de Rosenholtz et Malik et de Knill, le système visuel n'effectue pas une combinaison complexe entre les gradients de texture et la rupture de l'isotropie mais simplement une analyse des variations de fréquence afin d'estimer l'inclinaison des surfaces planes.

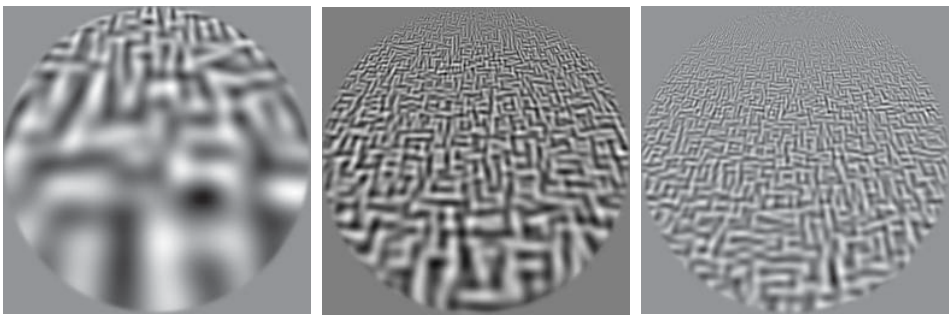


FIG. 5.34 – Exemples de stimuli obtenus en modifiant la valeur de la fréquence moyenne ; de gauche à droite : basse, moyenne et haute fréquence ; la fréquence moyenne modifie-t-elle la perception de l'inclinaison ?

5.3.4 Discrimination du tilt

Pour la discrimination du tilt, nous retrouvons les résultats bien connus d'une bonne estimation pour les orientations à 0^o et 90^o avec les meilleurs résultats pour 90^o et d'une moins bonne estimation pour les orientations autour de 45^o. La perspective linéaire semble

jouer un rôle important pour déterminer la direction en profondeur de la surface inclinée. Cependant la direction de la variation de fréquence influence également l'estimation de la direction et les deux indices peuvent se combiner dans cette tâche.

Les performances obtenues sur la discrimination du tilt sont très bonnes lorsque l'indice d'orientation est présent dans la texture. Il est important de noter que sur les texture présentant une variation d'orientation associée ou non à une variation de fréquence, les sujets ont non seulement bien discriminé l'orientation mais également la direction du tilt (i.e le signe du slant). Ces résultats montrent que le système visuel utilise bien l'information de la convergence des orientations vers un point de fuite. Cela confirme les travaux antérieurs sur l'importance de la perspective linéaire ([TTD05], [ABS98], [OML03], [LZ00], [SBft]) pour la perception 3D. Nos expériences montrent en plus que cette extraction se fait indépendamment de la perception de la profondeur (i.e même en l'absence de variation de fréquence où l'inclinaison de la surface n'est que faiblement perçue). Nos résultats montrent par contre que l'information de perspective linéaire est utilisée pour la perception de l'orientation en profondeur (ou, de manière équivalente, pour l'estimation du signe du slant) mais n'intervient pas dans l'estimation de l'inclinaison de la surface. Cette propriété est en accord avec les résultats obtenus par Saunders et Backus sur des textures présentant un alignement (Figure 4.24) et ceux obtenus par Li et Zaidi sur des surfaces ondulées (Figure 4.21).

Ces résultats montrent également que les masques de Gabor orientés à l'orthogonal de la direction du tilt (ne présentant donc pas de variation d'orientation) n'influencent pas les performances de discrimination. Des pré-tests ont aussi été réalisés sur des stimuli ne présentant que des orientations dans la direction du tilt. La texture obtenue est anisotropique. Aucune différence notable n'a été observée dans les performances de discrimination du slant et du tilt entre ces deux types de texture. Les stimuli utilisés dans les expériences ont été choisis car ils présentent en plus l'avantage de rendre applicable l'hypothèse d'isotropie et de permettre d'évaluer son influence éventuelle (la texture initiale est composée de deux directions principales).

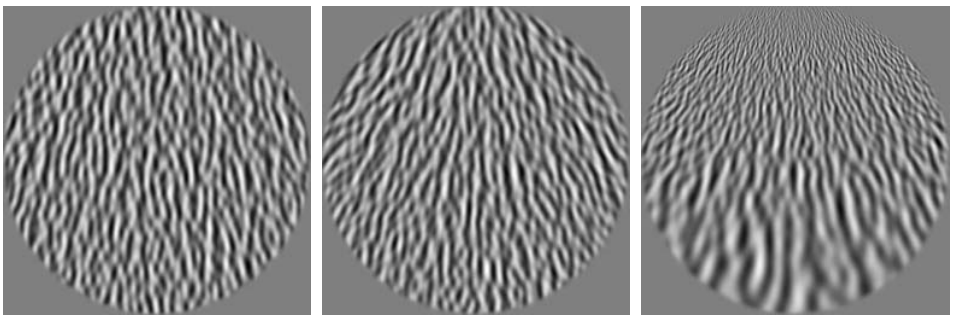


FIG. 5.35 – Exemple de stimuli anisotropiques utilisés en phase de pré-tests ; à gauche et au milieu : stimuli présentant uniquement une variation d'orientation pour un slant à 27^o et à 53^o ; à droite : stimuli présentant une variation d'orientation et une variation de fréquence pour un slant à 53^o.

Il est à noter que ces expériences ont été faites en considérant un angle de roll nul, c'est-à-dire la texture ne présente pas de rotation préalable avant projection (voir Figure 4.19, deuxième ligne). Dans ce cas la direction du point de fuite n'est plus celle du tilt. Si l'on considère un modèle combinant la variation de fréquence et la variation d'orientation

alors l'angle de roll introduit deux mesures différentes pour l'estimation de la direction en profondeur. Il serait intéressant de tester si cela introduit effectivement une difficulté pour estimer le tilt.

Enfin les deux indices donnant une information sur la direction en profondeur de la surface inclinée. Que devient la perception de l'inclinaison lorsqu'ils sont mis en opposition ? La figure 5.36 montre un exemple de stimulus présentant ce type de conflit. Un modèle de combinaison d'indice pourrait donc être également étudié pour connaître l'influence relative de chaque indice, notamment sur l'estimation du tilt.

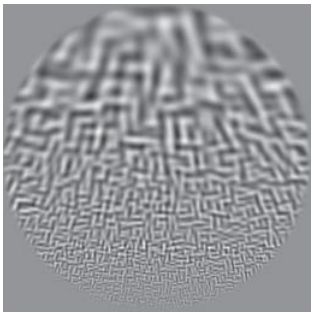


FIG. 5.36 – Exemples de stimulus où la valeur du tilt du gradient de fréquence (-90^o ou un slant négatif) est l'opposée de celle de la perspective linéaire (90^o ou un slant positif) ; quelle est l'inclinaison de la surface ?

5.3.5 Modèle spectral de Li et Zaidi

La figure 5.37 présente les spectres d'amplitude globaux des deux types de textures utilisées dans nos expériences : une texture composée de masques de Gabor orientés aléatoirement et une texture composée de masques de Gabor orientés suivant deux directions orthogonales. Le spectre de la première montre bien que la texture est isotropique. Le spectre de la deuxième texture possède deux composantes d'énergie suivant les deux orientations orthogonales (texture orientée). Cette texture correspond donc à la description de Li et Zaidi et doit permettre un bon rendu de l'information 3D.

Nos résultats montrent que les meilleurs performances de discrimination du tilt et du slant sont obtenues avec la texture orientée en présence des deux indices (variation de fréquence et variation d'orientation). Les sujets interrogés après l'expérience indiquent également mieux percevoir le fait que la surface soit plane. Ceci est en accord avec l'hypothèse de Li et Zaidi ainsi qu'avec les résultats de Saunders et Backus. En plus de ces auteurs nos expériences permettent l'évaluation quantitative de discrimination de l'orientation et de l'inclinaison.

Cependant nos résultats montrent que la texture isotropique transmet également une information suffisante pour extraire l'inclinaison (avec une très bonne précision) et l'orientation (avec une précision moyenne). Donc contrairement à l'hypothèse de Li et Zaidi, il n'est pas nécessaire d'avoir une composante discrète orientée d'énergie. Nous retrouvons ainsi le même résultat que Saunders et Backus [SBft].

L'hypothèse émise par Li et Zaidi est cependant importante car elle met l'accent sur l'utilisation à la fois de l'indice de variation de fréquence et de perspective linéaire. L'analyse du spectre global ne permet pas de décrire avec précision l'information extraire par chaque indice. Nos expériences permettent d'obtenir cette description en mettant l'accent sur l'analyse

séparée du gradient de fréquence et de la perspective linéaire (vue comme un gradient d'orientation). Les deux indices ayant des rôles bien distincts, il est possible d'envisager l'existence de deux mécanismes spécialisés dans l'analyse de ces deux gradients.

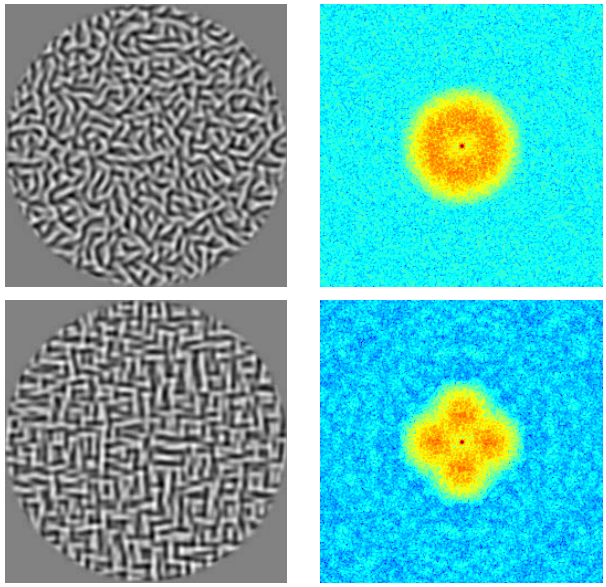


FIG. 5.37 – Spectres globaux des deux types de texture utilisées dans nos expériences ; première ligne, à gauche : texture utilisée pour générer les stimuli présentant uniquement une variation de fréquence (composée d'orientations aléatoires) ; à droite : spectre global isotropique ; deuxième ligne, à gauche : texture utilisée pour générer les stimuli présentant une variation d'orientation (composée de deux orientations orthogonales) ; à droite : spectre global présentant des composantes d'énergie suivant deux orientations.

5.3.6 Effet collatéral de la perspective linéaire

La présence des deux indices peut s'avérée utile pour estimer avec précision l'inclinaison lorsque la texture présente une forte irrégularité spatiale (par exemple une texture composée de point polka répartis avec une densité non-uniforme). Dans ce cas un processus de régularisation du gradient de fréquence est important pour pouvoir effectuer l'estimation car il est possible qu'à certaines positions le gradient soit nul ou très faible. Il serait intéressant d'évaluer alors l'intervention de la perspective linéaire. Celle-ci pourrait notamment introduire une contrainte de forme (par exemple une surface plane) dans le processus de régularisation, ainsi que le note Todd *et al* [TTD05].

Ainsi cet indice n'interviendrait pas directement pour résoudre la tâche mais contribuerait à sa robustesse. Nous pourrions parler alors d'effet collatéral de l'indice de perspective linéaire pour l'estimation de l'inclinaison des surfaces. Ce ne serait pas la combinaison directe des indices qui contribue à l'amélioration de la robustesse et de la précision de l'estimation mais l'un des indices induirait une contrainte supplémentaire permettant de simplifier le traitement de ou des autres indices. Cet effet se retrouve pour d'autres indices tels que la couleur. Celle-ci permet de discriminer facilement différentes régions d'une scènes ou est intégrée dans la représentation en mémoire des objets mais elle contribue également à une augmentation de

l'attention visuelle. Par exemple les travaux de Wichmann *et al* [WSG02] montrent que sur certaines tâches de mémorisation de scènes naturelles, les performances obtenues s'expliquent en partie par le diagnostique de la couleur mais également parce que, dans le cas d'un temps de présentation court, certaines zones particulières de la scène ont eu une saillance perceptive augmentée par leur couleur caractéristique facilitant la mémorisation. Les auteurs distinguent ainsi le niveau des capteurs (l'indice de couleur participe à la saillance et à la segmentation) du niveau de la représentation en mémoire (l'indice de couleur participe à la reconnaissance des objets et des scènes). Ainsi la perspective linéaire participerait à la représentation de la surface en indiquant une surface plane mais participerait également à l'analyse du gradient de fréquence en contraignant la régularisation.

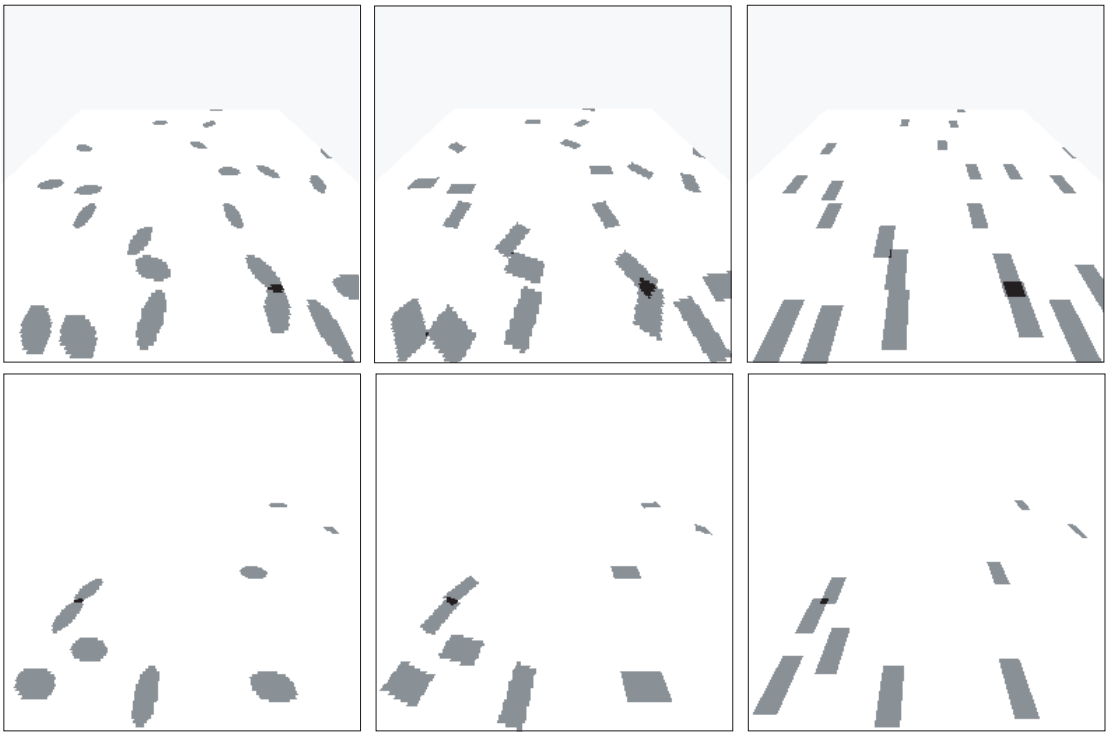


FIG. 5.38 – Exemples d'effet collatéral : stimuli composé de texels (en haut : 22, en bas : 9) positionnés aléatoirement sur une surface plane projetée avec un slant à 40^o et un tilt à 90^o ; de gauche à droite : ellipses avec orientation aléatoires ; rectangles aux même positions et avec les même orientations ; rectangles aux même positions mais tous orientés dans la direction du tilt (présence de la perspective linéaire) ; le manque de texels de la deuxième ligne par rapport à la première ligne nécessite une régularisation plus importante ; la présence de l'indice de perspective linéaire renforce la perception d'une surface plane inclinée.

5.3.7 Vers 3 indices de texture : la fréquence, la convergence et la courbure

Pour aller plus loin, il serait intéressant d'étendre cette étude au cas des surfaces courbes. Les stimuli que nous avons construit peuvent être facilement adaptés à cette étude. Notamment il serait possible d'obtenir des stimuli proches de ceux utlisés par Prins et Kingdom [PK02] (Figure 4.18). Pour l'indice de fréquence, il suffit d'appliquer une variation de fré-

quence correspondant à une surface courbe et non plus à une surface plane. Pour la variation d'orientation il faut dériver la fonction décrivant l'orientation de la courbure locale et l'appliquer à l'orientation des masques de Gabor.

Dans la lignée des travaux de Li et Zaidi [LZ04] sur la séparation des indices de fréquence et de perspective linéaire et dans la lignée de ceux de Knill [Kni01] dans l'analyse également de la courbure, il sera possible d'envisager une approche consistant à distinguer 3 types d'indice de texture :

- la variation de fréquence pour l'analyse de la profondeur locale (inclinaison ou forme) et de l'orientation en profondeur
- la convergence des orientations (perspective linéaire) pour l'analyse de la direction en profondeur et l'estimation du point de fuite
- la courbure des orientations pour l'analyse de la forme

La figure 5.39 présente un cas où ces 3 indices sont présents. Nous proposons ainsi que 3 mécanismes spécialisés soient associés à l'analyse de ces 3 indices de texture pour obtenir une information complète de la 3D.

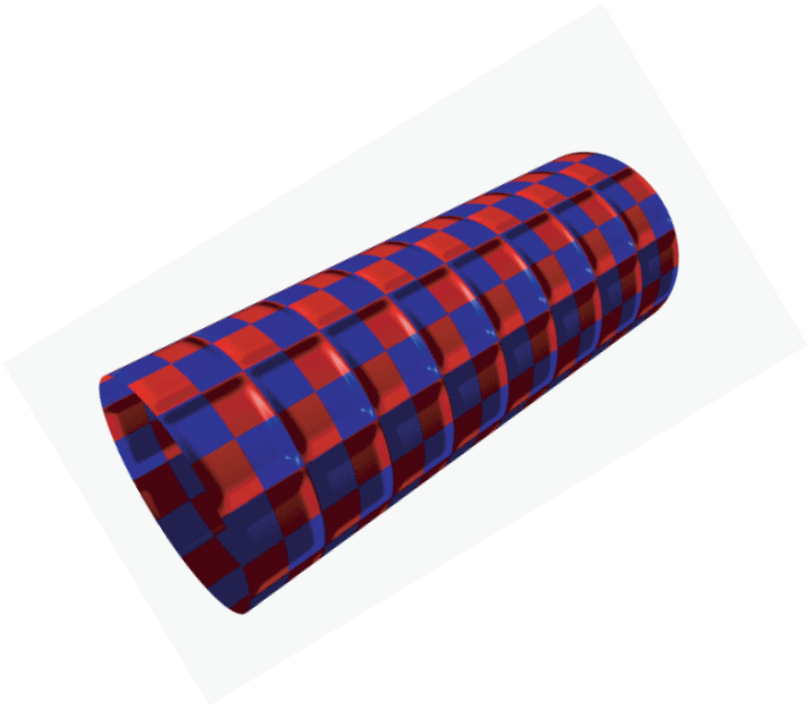


FIG. 5.39 – Cylindre en 3D présentant les 3 types d'indices de texture proposés : les carreaux indiquent la variation de fréquence (forme et direction en profondeur) ; les lignes de fuite correspondent à la perspective linéaire (direction en profondeur) ; les courbes circulaires indiquent la forme cylindrique (forme).

5.4 Conclusion

Nous avons mis en place des expériences psychophysiques afin d'approfondir l'étude des indices de variation de fréquence et de perspective linéaire. Pour cela nous avons crée un nouveau type de stimuli permettant d'analyser ces deux indices séparément. Nous avons ainsi caractériser l'information extraite par chaque indice. Nous avons enfin émis l'hypothèse de l'existence de deux mécanismes spécialisés dédiés à leur traitement spécifique.

Cependant est-il possible de concevoir un modèle du système visuel basé sur l'extraction de ces deux indices ? Ceci est l'objet du chapitre 6. Nous mettrons particulièrement en avant la séparation des indices de fréquence et d'orientation afin de les rendre les plus indépendants possible et de pouvoir leur associer à chacun un modèle d'extraction spécialisé conformément à nos résultats en psychophysique.

Chapitre 6

Modèle de V1 pour la perception 3D

Nous nous intéressons dans ce chapitre à la modélisation du système visuel. A partir d'un modèle des cellules complexes de l'aire V1, nous proposons un modèle biologiquement plausible pour l'analyse de la fréquence dans une image. Nous appliquons ce modèle au problème de l'extraction de la forme à partir de la texture. Nous nous basons à la fois sur la description des premières étapes du système visuel faite au chapitre 3 et sur notre étude psychophysique des indices de fréquence et de perspective linéaire faite au chapitre 5.

D'abord nous présentons les filtres *log-normaux*, développés en remplacement des filtres de Gabor classiques à cause de leur plausibilité biologique et de leur propriétés théoriques. Nous décrivons ensuite une méthode d'extraction de la fréquence moyenne locale dans une image, conçue comme une série de combinaisons des réponses des filtres et permettant de séparer les informations de fréquence et d'orientation. Une dernière étape permet d'extraire les paramètres 3D (les angles de tilt et de slant) de surfaces planes et incurvées. Un modèle complet d'analyse de la texture pour en extraire la forme est ainsi obtenu. La méthode est évaluée sur différentes bases de texture et de scènes naturelles et elle est comparée à des algorithmes connus.

Ce travail a donné lieu à plusieurs communications [6] [5] [4] [3] et à deux articles en cours de révision [2] [1].

6.1 Modèles des cellules complexes

Nous présentons dans cette section notre modèle des cellules complexes permettant de décomposer l'information visuelle en fréquence et en orientation. Pour cela nous avons développé des filtres spatio-fréquentiels appelés filtres *log-normaux* dont nous allons décrire les caractéristiques et les avantages sur les filtres utilisés plus classiquement.

6.1.1 Comment choisir les filtres appropriés ?

La technique de filtrage spatio-fréquentielle est abondamment utilisée pour résoudre les problèmes d'analyse en vision par ordinateur (par exemple la reconnaissance d'objets, la reconnaissance de visage, l'analyse de texture). Un grand nombre de filtres spatio-fréquentiels ont ainsi été développés (par exemple les filtres de Gabor, les filtres log-Gabor et les différences

de gaussiennes (DoG) (voir [Wal01] et [BNB04] pour des études comparatives)). Nous devons choisir le plus approprié pour notre application spécifique. Pour cela nous nous inspirons du fonctionnement du cortex visuel primaire.

Comme nous l'avons décrit au chapitre 3, le système visuel effectue un découpage du champ visuel (ici nous considérons qu'il se réduit au plan de l'image traitée) en un ensemble de sous-régions locales (les champs récepteurs) en interactions les unes par rapport aux autres. Le cortex visuel primaire (V1) est organisé en colonnes de fréquences et d'orientations et réalise ainsi un échantillonnage du spectre local correspondant à la région spatiale couverte par le champ récepteur. Cette fonction est en particulier réalisées par les cellules complexes, insensibles à la phase locale de l'image.

Afin de reproduire cette décomposition de l'image en régions locales, le modèle de cellules complexes choisi doit posséder la propriété des filtres de Gabor d'être localisés à la fois dans l'espace et le domaine fréquentiel (i.e dans une région spatiale définie). Ces filtres sont appliqués sur le spectre d'énergie de la région locale (le spectre d'amplitude élevé au carré). Cela permet d'être insensible aux translations spatiales locales (i.e à la phase locale) tout en conservant les statistiques du second ordre de l'image (ou de la texture). Cela permet également de réaliser un lissage local de ces statistiques conduisant à une estimation plus robuste.

Les réponses des filtres spatio-fréquentiels appliqués sur le spectre d'énergie correspond à un modèle général des réponses des cellules complexes (voir Chapitre 3). Les filtres de Gabor sont des filtres particulièrement utilisés. Ils peuvent être facilement paramétrés en fréquence et en orientation conduisant à un échantillonnage polaire du spectre d'énergie. Cependant, comme le montre Wallis [Wal01], ces filtres ne sont pas des modèles complètement satisfaisants des cellules complexes. De plus ils présentent l'inconvénient de ne pas être à variables séparables (en fréquence et en orientation). Enfin dans le cas de l'étude des scènes naturelles et de la texture, deux transformations géométriques indépendantes doivent être prises en compte : le zoom et la rotation. Un changement de facteur de zoom se traduit par une variation de fréquence et une rotation se traduit par une rotation équivalente dans le spectre de la région analysée. Il apparaît donc nécessaire de pouvoir analyser ces deux variations de manière indépendantes. Comme nous l'avons décrit au chapitre 5, une analyse séparée de la variation de fréquence et de la variation d'orientation (pour la perspective linéaire) semble être un modèle plausible du fonctionnement de la perception 3D à partir de la texture. Ceci impose également de pouvoir séparer ces deux indices afin de leur associer un mécanisme d'analyse spécialisé. Aussi en remplacement des filtres de Gabor, des filtres à variables séparables ont été développés (par exemple les filtres log-Gabor de Field [Fie87] ou les filtres dits *log-normaux* de Knutsson *et al* [KWG94]) et ont été employés dans différentes techniques de vision par ordinateur.

Dans ce travail nous proposons un nouveau type de filtres, appelés filtres *log-normaux*, reposant sur la définition de la fonction log-normale connue en statistique et qui présentent l'avantage d'être à variables séparables mais également d'être bien adaptés aux transformations géométriques de zoom et de rotation.

6.1.2 Filtres Log-normaux

Les filtres log-normaux sont obtenus à partir de la distribution log-normale, classiquement utilisée en statistique [CS88] et qui permet de caractériser beaucoup de données biologiques [LSA01]. Les filtres sont appliqués sur le spectre d'énergie de régions locales de l'image à

analysée. Ils sont donc définis en 2 dimensions spatiales. La largeur de bande en orientation est définie par une enveloppe de type gaussienne. L'énergie d'un filtre log-normal est défini par :

$$|G_{i,j}(f,\theta)|^2 = |G_i(f).G_j(\theta)|^2 = A\frac{1}{f^2}exp\left(-\frac{1}{2}\left(\frac{ln(f/f_i)}{\sigma_r}\right)^2\right).cos^{2n}(\frac{\theta-\theta_j}{2}) \tag{6.1}$$

Avec $G_{i,j}$, la fonction de transfert du filtre, $G_i(f)$ et $G_j(\theta)$ représentant respectivement la composante fréquentielle (radiale) et en orientation (tangentielle) du filtre. Leur enveloppe est gaussienne sur une échelle log-fréquence et en orientation. f_i est la fréquence centrale, θ_j, l'orientation centrale, σ_r, la largeur de bande en fréquence. n controle la largeur de bande en orientation de manière à se rapprocher d'une enveloppe gaussienne. A est un facteur de normalisation tel que $||G_{i,j}(f,\theta)||^2 = 1$ (A=(= $\frac{2^{2n}}{2\pi C_{2n}^n}\frac{1}{\sigma_r\sqrt{2\pi}}$)).

Les filtres log-normaux partagent les mêmes propriétés que les filtres de Gabor (i.e la localisation spatiale et dans le domaine de Fourier ; la paramétrisation en fréquence et en orientation). Par définition ces filtres sont à variables séparables en fréquence et en orientation (Equation 6.1). 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 2π-periodique en orientation et un support fini (évitant d'effectuer une troncature aux limites du filtre). Cette fonction se rapproche d'une enveloppe gaussienne avec une précision supérieure à 0.1%. Finalement le gain du filtre est nul en $f = 0$ quelque soit la largeur de bande utilisée (l'équation 6.1 impose une composante continue nulle). Ceci permet d'obtenir des filtres toujours bien définis dans le quadrant du plan de Fourier échantilloné.

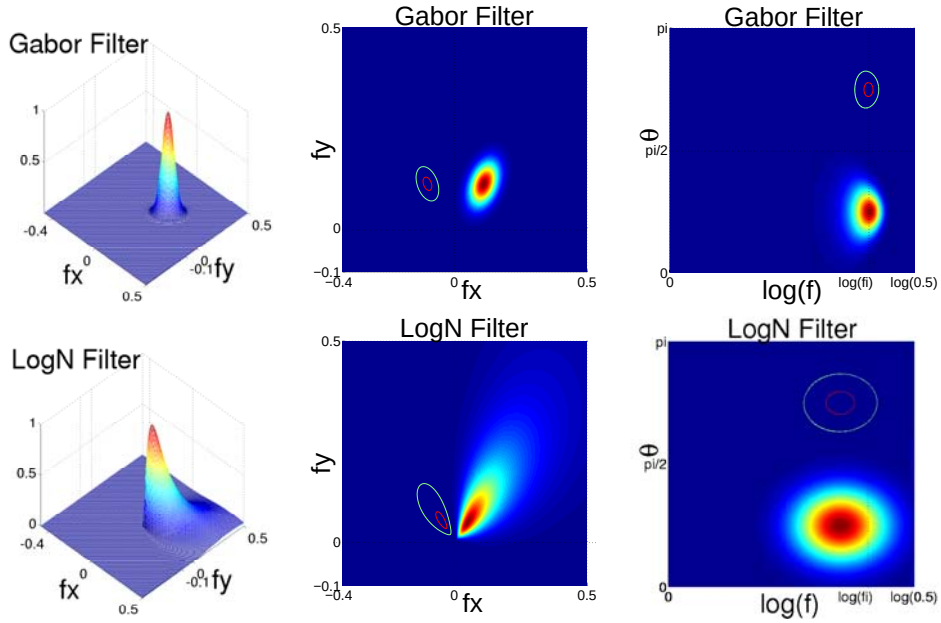


FIG. 6.1 – Comparaison entre les filtres de Gabor (première ligne) et les filtres log-normaux (deuxième ligne) ; à gauche : représentation 3D ; au milieu : le filtre et ses contours à 50% et 90% du maximum d'énergie en coordonnées cartésiennes ; à droite : même filtre mais représenté en coordonnées log-polaires où nous pouvons observer que les filtres log-normaux deviennent symétriques de manière similaire au profil des cellules corticales [Wal01].

La figure 6.1 présente une comparaison entre les filtres de Gabor et les filtres log-normaux. Les filtres sont représentés en 3D, en coordonnées cartésiennes et en coordonnées log-polaires. La figure 6.2 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.

En coordonnées cartésiennes, le filtre log-normal présente une dissymétrie qui augmente avec les basses fréquences. Cette propriété, associée au fait que le gain du filtre est nul en $f = 0$ quelque soit la largeur de bande utilisée , permet au filtre log-normal d'être bien défini à toutes les fréquences, notamment aux très basses fréquences. Ceci n'est pas le cas avec des filtres de Gabor qui peuvent donner des réponses bruitées en très basses fréquences ne permettant pas une analyse robuste des composantes très basses fréquences (si la largeur de bande du filtre de Gabor est augmentée pour pouvoir couvrir plus d'échantillons du spectre d'énergie, des fréquences négatives peuvent être alors prises en compte).

Il est également possible d'observer qu'à la fois en coordonnées cartésiennes et en coordonnées logpolaires, les filtres sont bien adaptés à un échantillonnage du spectre. En coordonnées logpolaires, les filtres log-normaux deviennent symétriques (l'enveloppe devient gaussienne) ce qui conduit à une meilleur couverture du spectre que les filtres de Gabor qui présentent une dissymétrie forte dans ce cas.

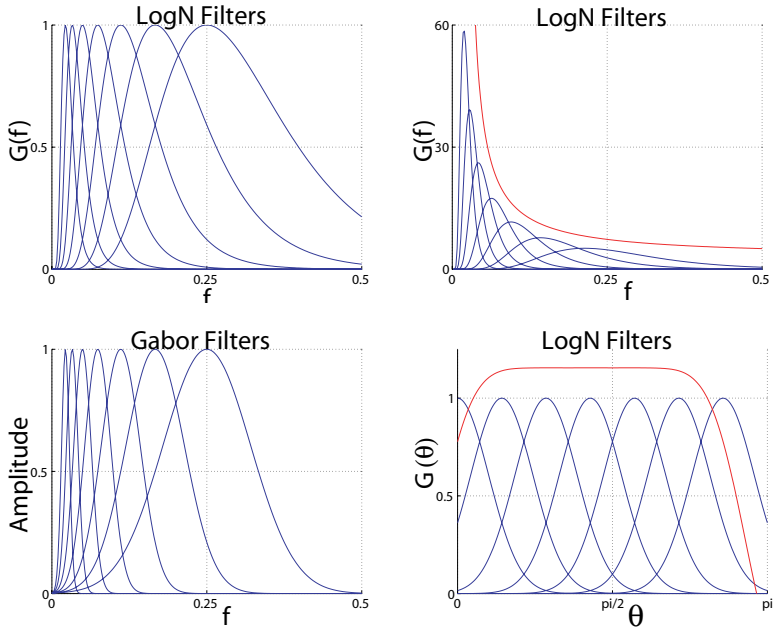


FIG. 6.2 – Comparaison entre les profils des filtres de Gabor et ceux des filtres log-normaux ; première ligne, à gauche : profils en fréquence des filtres log-normaux ; à droite : profils en fréquence des filtres log-normaux avec le coefficient $1/f$ (décroissance des maximum des filtres) ; deuxième ligne, à gauche : profils en fréquence des filtres de Gabor ; à droite : profils en orientation des filtres log-normaux ; la couverture des filtres log-normaux est obtenue en effectuant leur somme : elle suit une loi en $1/f$ suivant les fréquences et reste constante suivant les orientations.

La figure 6.2 présente les différents profils d'un banc de filtres de Gabor et d'un banc de filtres log-normaux. Pour ce dernier les profils ont été reportés avec et sans le facteur en amplitude $1/f$ afin de d'observer la dissymétrie des filtres, notamment en basses fréquences. La figure 6.2 deuxième ligne à droite montre également la bonne couverture des orientations où la somme des profils des filtres est constante.

Finalement les filtres log-normaux présentent d'autres propriétés algébriques qui les rendent attractifs pour l'analyse d'image : la fonction de transfert est C^∞ ; les parties réelles et imaginaires sont en quadrature ; la fonction est auto-similaire et peut ainsi servir d'ondelette mère.

D'après [Wal01], toutes les caractéristiques présentées sur les filtres log-normaux correspondent aux caractéristiques principales des cellules complexes. Les filtres log-normaux peuvent ainsi être considérés comme une bonne approximation de leur réponse spatiale.

6.1.3 Banc de filtres log-normaux

Pour échantillonner la moitié supérieure du spectre d'amplitude, nous définissons un banc de filtres log-normaux avec des largeurs de bande constantes en fréquence relative et en orientation (voir Chapitre 6.2.2 une version avec largeur de bande non constante). Sur la figure 6.3 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'un banc équivalent composé de filtres de Gabor avec une largeur de bande similaire.

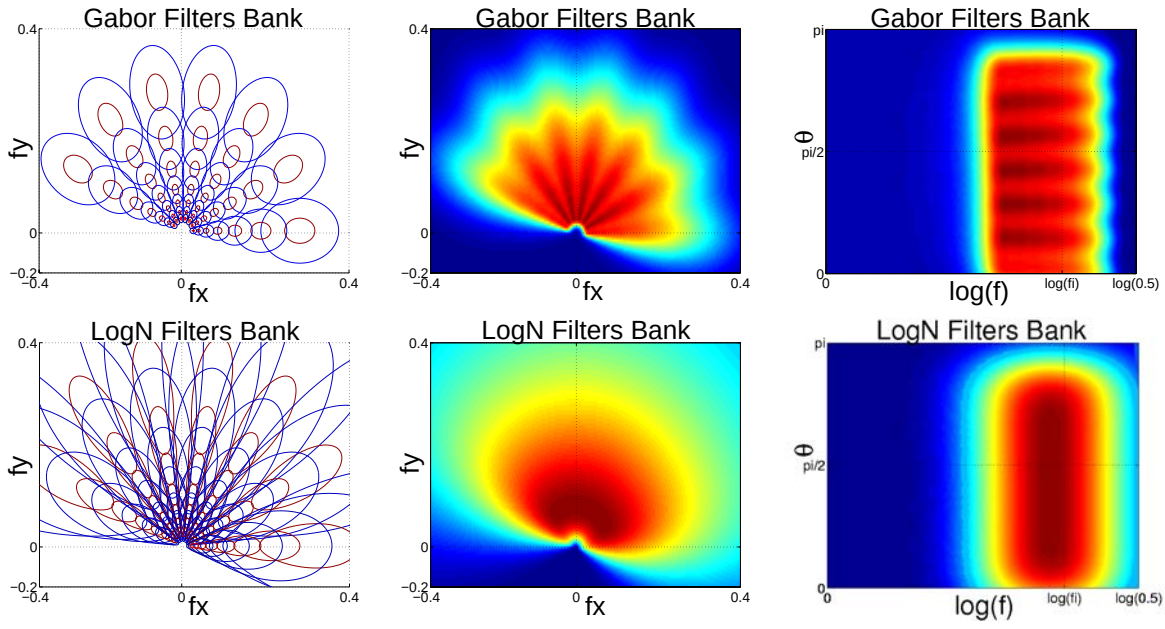


FIG. 6.3 – Comparaison entre les filtres de Gabor (première ligne) et les filtres log-normaux (deuxième ligne) ; à gauche : contours du filtre à 50% et à 90% du maximum d'énergie ; au milieu : banc de filtres dans l'espace de fréquences en coordonnées cartésiennes ; à droite : banc de filtres à 50% et à 90% du maximum d'énergie en coordonnées logpolaires.

6.1.4 Zoom et rotation

Les réponses des filtres log-normaux présentent deux propriétés de migration relatives aux transformations en zoom et en rotation de l'image. Par migration nous entendons que la réponse du filtre *migre* en fonction de la transformation géométrique de l'image sans dépendre de la valeur de la fréquence ou de l'orientation autour desquelles s'effectue la transformation. Ainsi un zoom ou une rotation de l'image peuvent être directement suivis par la réponse des filtres log-normaux.

Migration de la réponse du filtre avec le zoom
Une variation de fréquence peut être produite par un zoom de l'image et peut s'écrire :

$$i(x,y) \rightarrow i(\alpha x, \alpha y)$$

Ceci donne dans le domaine de Fourier :

$$S(f,\theta) \rightarrow \frac{1}{\alpha^4} S(\frac{f}{\alpha},\theta) \tag{6.2}$$

où α est le facteur de zoom et S la densité spectrale de puissance (spectre d'énergie) de l'image.

Si nous exprimons la réponse du filtre $C_{i,j}$ sur une échelle logpolaire en posant $v = ln(f)$ nous obtenons :

$$C_{ij} = A \int_{\theta=0}^{2\pi} G_\theta^2(\theta - \theta_j) \int_{v=-\infty}^{+\infty} \frac{1}{\alpha^4} S(e^v,\theta) exp\left(-\frac{1}{2}\left(\frac{v - v_i)}{\sigma_r}\right)^2\right) dv d\theta \tag{6.3}$$

Il est possible d'observer que le coefficient $1/f^2$ disparaît avec le changement de variable.
En appliquant l'équation 6.3 sur $S(\frac{f}{\alpha},\theta)$ en posant $v_m = v + ln(\alpha)$ nous obtenons :

$$C_{ij}(\alpha) = A \int_{\theta=0}^{2\pi} G_\theta^2(\theta - \theta_j) \int_{v_m=-\infty}^{+\infty} \frac{1}{\alpha^4} S(e^{v_m},\theta) exp\left(-\frac{1}{2}\left(\frac{v_m - (v_i - ln(\alpha))}{\sigma_r}\right)^2\right) dv_m d\theta \tag{6.4}$$

Ainsi un zoom de l'image peut être parfaitement suivi par le changement dans la réponse du filtre grâce au terme $1/f^2$ et à l'expression de la fréquence en échelle logarithmique. Dans le cas du filtre de Gabor, la réponse à une variation de fréquence est toujours dépendante d'un facteur de fréquence supplémentaire et donc ne donne pas un accès direct au zoom. Cette propriété n'est pas non plus vérifiée avec des filtres Log-gabor [Fie87] ou la version des filtres log-normaux développés par Knutsson *et al* [KWG94].

Migration de la réponse du filtre avec la rotation
De la même manière que pour la composante fréquentielle, l'expression de la rotation de l'image conduit à un parfait suivi par migration des réponses des filtres log-normaux. Une rotation d'angle β induit uniquement une modification sur la composante en orientation du filtre :

$$C_{ij}(\alpha,\beta) = A \int_{\theta=0}^{2\pi} G_\theta^2(\theta - (\theta_j - \beta)) \int_{v'=-\infty}^{+\infty} \frac{1}{\alpha^4} S(e^{v'},\theta) exp\left(-\frac{1}{2}\left(\frac{v' - (v_i - ln(\alpha))}{\sigma_r}\right)^2\right) dv' d\theta \tag{6.5}$$

6.2 Modèle de V1 pour l'analyse des fréquences

La modélisation des cellules complexes présentes dans l'aire corticale V1 constitue l'élément de base pour l'analyse de l'information visuelle effectuée par le système visuel. Elle permet de récupérer les composantes de fréquence et d'orientation sur l'ensemble des positions spatiales du champ visuel. Cependant la réponse de ces filtres est relativement pauvre même si elle reflète les statistiques du signal étudié. Ainsi notamment pour le problème de l'analyse de la 3D dans les images, des informations à la fois plus pertinentes et plus robustes doivent être extraites. Par exemple au lieu d'indiquer simplement la présence ou l'absence d'une composante fréquentielle dans l'image (voir le modèle de Oliva et Torralba décrit au chapitre 2.1), les mesures réelles de la fréquence moyenne et des orientations permettent de décrire de manière plus précise l'information contenue dans l'image.

D'autre part le système visuel analyse des signaux naturels qui présentent de nombreuses sources de bruit ou de perturbations. Aussi, de manière similaire aux réponses des photo-récepteurs de la rétine, la modélisation des connexions locales entre cellules complexes doit permettre de rehausser l'information à extraire tout en réduisant ces perturbations.

Nous présentons un modèle pour l'analyse des fréquences basé sur la structure de l'aire corticale V1. Cette modélisation est adaptée à l'extraction de l'indice de fréquence nécessaire au problème de la perception 3D. Nous décrivons d'abord une méthode d'extraction de la fréquence moyenne locale que nous intégrons ensuite à une analyse d'image basée sur une décomposition en sous-régions locales (par imagettes) et sur une normalisation corticale permettant d'obtenir une estimation robuste de la fréquence moyenne sur l'ensemble de la surface.

6.2.1 Modèle d'extraction de la fréquence moyenne locale

Dans cette section, la propriété de séparabilité des filtres en fréquence et en orientation est utilisée pour estimer la fréquence moyenne locale d'une image (i.e si la texture possède plusieurs composantes fréquentielles, celles-ci ne sont pas considérées individuellement). Prenons la réponse fréquentielle du $i$ème filtre log-normal :

$$G_i^2(f) = \frac{1}{f^2} exp\left(-\frac{1}{2}\left(\frac{ln(f/f_i)}{\sigma_r}\right)^2\right) \tag{6.6}$$

De manière similaire à Knutsson *et al* [KWG94] [GK95], le rapport des réponses de deux filtres adjacents peut s'exprimer par :

$$\begin{aligned}\frac{G_{i+1}^2(f)}{G_i^2(f)} &= exp\left(-\frac{1}{2\sigma_r^2}[(ln(f/f_{i+1})^2 - (ln(f/f_i))^2]\right) \\ &= \left(f/\sqrt{f_i f_{i+1}}\right)^{\frac{ln(f_{i+1}/f_i)}{\sigma_r^2}}\end{aligned} \tag{6.7}$$

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

$$G_{i+1}^2(f) = \frac{f}{\sqrt{f_i f_{i+1}}} G_i^2(f) \tag{6.8}$$

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 $S(f, \theta)$:

$$C_i = \int_f G_i^2(f) \int_\theta S(f,\theta) \sum_j G_j^2(\theta) f df d\theta \tag{6.9}$$

Le rapport des réponses de deux filtres par bande de fréquence adjacentes C_{i+1} et C_i donne :

$$\frac{C_{i+1}}{C_i} = \frac{1}{\sqrt{f_i f_{i+1}}} < f >_i \tag{6.10}$$

où $< f >_i$ représente la fréquence moyenne locale estimée à la $i$ème bande de fréquence. L'équation 6.10 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 $< f >_i$ 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 $< f >$:

$$< f >= \sum_i \frac{C_i}{\sum_k C_k} < f >_i \tag{6.11}$$

La réponse C_i dans l'équation 6.9 peut être interprétée comme la mesure du poids de la caractéristique fréquentielle locale extraite de l'image parmis l'ensemble des caractéristiques extraites. En effet si la réponse C_i est forte alors cela signifie que la fréquence moyenne $< f >_i$, 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 $< f >$ de l'équation 6.11 est basée sur une estimation globale sommée sur l'ensemble de toutes les fréquences. Contrairement à [GL96] et à [CM02], dans cette approche il n'y a pas besoin de se restreindre à une bande de fréquence particulière. Il n'est pas non plus nécessaire d'effectuer au préalable l'estimation d'une fréquence diagnostique qui serait adaptée aux caractéristiques de la texture. Enfin cette technique permet d'obtenir une estimation plus robuste au bruit qui serait présent à une fréquence particulière.

Cette méthode présente également l'avantage de réaliser la séparation entre l'information de fréquence et d'orientation, ainsi que le suggèrent les travaux de Li et Zaidi [LZ00]. La fréquence locale est estimée sans aucune hypothèse sur les statistiques des orientations, c'est-à-dire notamment sans hypothèse d'isotropie, comme il a été suggéré à de multiples reprises (voir Chapitre 4.2). Seule une hypothèse d'homogénéité locale sur les statistiques des composantes des fréquences est utilisée.

6.2.2 Banc de filtres à largeur de bande relative décroissante

Afin d'évaluer la précision de la méthode d'extraction, celle-ci est appliquée sur un réseau dont la fréquence est parfaitement connue (figure 6.4 gauche). Ce réseau contient une augmentation linéaire de la fréquence et l'ensemble des réseaux crées couvrent les très basses fréquences ($f = 0$) jusqu'aux très hautes fréquences ($f = 0.33$).

Les courbes d'estimation finales (Figure 6.4 droite) sont obtenues après approximation de la fonction (méthode du simplex) de l'ensemble des estimations obtenues à différentes

positions spatiales du réseau considéré. L'estimation de la fréquence moyenne locale, suivant la méthode décrite précédemment, est obtenue sur une fenêtre glissante (imagette 2D) avec une taille définie (96X96 pixels)(voir Section 6.2.3 pour la description de l'analyse par imagettes). La courbe en pointillés représente l'estimation obtenue avec le banc de filtres log-normaux décrit au chapitre 6.1.2. L'erreur d'estimation apparaît faible dans les basses fréquences mais augmente rapidement vers une valeur asymptotique vers les hautes fréquences.

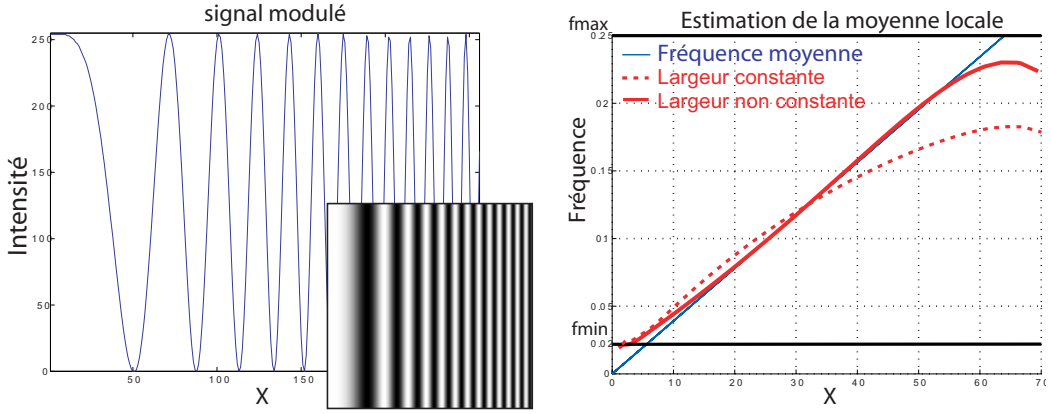


FIG. 6.4 – Estimation de la fréquence locale sur un réseau ; à gauche : projection du réseau avec le signal 1D représentant la modulation correspondante du signal ; à droite : fréquence moyenne réelle du réseau (ligne pleine, rectiligne) ; estimation de la fréquence moyenne avec une largeur de bande relative constante des filtres log-normaux (ligne en pointillée) ; estimation de la fréquence moyenne avec une largeur de bande relative non-constante (ligne pleine).

Une des caractéristiques des filtres log-normaux est que leur largeur de bande relative en fréquence est indépendante 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 σ_r.
Ici, l'équation 6.8 impose une valeur de σ_r égale à $ln(f_{i+1}/f_i)$ afin de conserver leur rapport constant et égal à 1. Les filtres du premier banc décrit au chapitre 6.1.2 sont ainsi 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 σ_r de manière à ce que le rapport décrit précédemment deviennent légèrement supérieur à 1, l'équation 6.10 reste une approximation valide de la fréquence moyenne locale. Par contre l'analyse de l'équation 6.2.1 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 σ_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 6.4 à droite) 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 6.5 à droite présente différents résultats de l'estimation de la méthode pour différentes tailles d'imagettes (avec un coefficient identique égale à 2), depuis des tailles très petites (48X48 pixels) jusqu'à des tailles très importantes (128X128 pixels). Il est possible d'observer que la précision de la méthode n'est que faiblement influencée par la taille de l'imagette, qui pourra donc être choisie librement.

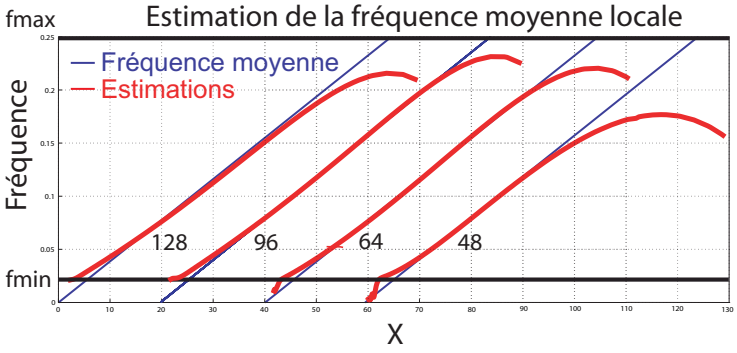


FIG. 6.5 – Estimation de la fréquence moyenne locale pour différentes tailles d'imagettes, respectivement 128X128, 96X96, 64X64, 48X48 pixels.

Cette largeur de bande relative non-constante en fréquence a également été observée sur les cellules corticales. Les données collectées sur le cortex chez l'Homme et chez le macaque [DAT82] 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.

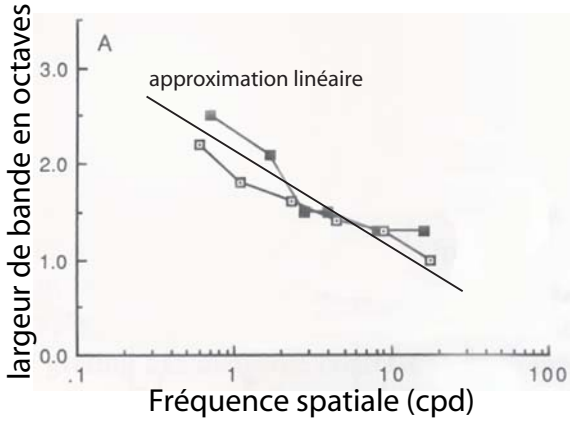


FIG. 6.6 – Données physiologiques collectées à partir du cortex de l'être humain et du macaque (tirées de [DAT82]) ; chaque point représente la largeur de bande des cellules corticales en fonction de leur fréquence centrale ; il est possible d'observer une décroissance linéaire de la largeur de bande (approximativement d'un facteur 2).

6.2.3 Décomposition et normalisation

6.2.3.1 Analyse locale de l'information d'amplitude

Afin d'estimer la fréquence locale sur toute la surface de l'image, celle-ci est décomposée en imagettes (Figure 6.7). La définition automatique d'une taille optimale des imagettes afin d'étudier les propriétés locales d'une texture est un problème encore non résolu. Dans notre cas, des imagettes de taille $96X96$ pixels sont jugées appropriées pour capturer les propriétés

statistiques des images de taille $256X256$ pixels. Une fenêtre de Hamming est appliquée sur chaque imagette afin d'éviter les effet 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 imagette de 96 pixels de côté perçue à une distance de 1m correspond à un angle visuel de 1^o ce qui est la taille moyenne des champs récepteurs de V1. La précision spatiale peut être adaptée en faisant varier le décalage entre les imagettes. Généralement un décalage de 8 ou 4 pixels est choisi, ce qui correspond à une décomposition de l'image en $21X21$ ou $42X42$ imagettes. La figure 6.7 présente les différentes étapes de prétraitement appliquées sur chaque imagette avant l'application des filtres corticaux.

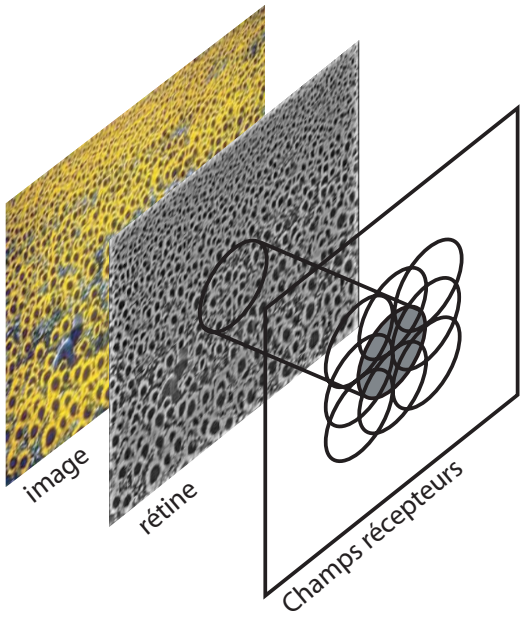


FIG. 6.7 – Les différentes étapes de prétraitement appliquées à chaque imagette ; de gauche à droite : image initiale d'un champ de tournesols ; préfiltrage par la rétine ; décomposition locale en imagettes de l'ensemble de l'image de manière similaire aux champs récepteurs des cellules corticales.

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 introduit en physiologie de champ récepteur associé à chaque cellules (voir Chapitre 3). La méthode d'extraction de la fréquence moyenne locale décrite à la section précédente s'applique sur le spectre d'amplitude de chaque imagette locale. Ainsi à travers cette méthode de décomposition locale de l'image, il est possible de conserver l'information de position spatiale sans l'utilisation de l'information de phase, comme nous l'avons décrit au chapitre 2.1.

6.2.3.2 Normalisation corticale

Les réponses par bande de fréquence, correspondant aux coefficients C_i dans les équations 6.10 et 6.11, 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 $G^2_{i,j}$, notée $G^2_{i,j,norm}$, peut alors se réécrire de la manière suivante :

$$G^2_{i,j,norm}(f,\theta) = \frac{G^2_{i,j}(f,\theta)}{\sum_k G^2_{k,j}(f,\theta) + \epsilon} = \frac{G^2_i(f)}{\sum_k G^2_k(f) + \frac{\varepsilon}{G^2_j(\theta)}} \tag{6.12}$$

la constante ε permet d'éviter des rehaussements des réponses des filtres dans le cas où l'énergie globale est trop faible dans la bande d'orientation considérée ($\varepsilon = 0.1$). En introduisant $G^2_{i,j,norm}$ au lieu de $G^2_{i,j}$ dans l'équation (6.9), l'équation (6.10) restent inchangées. La combinaison finale est obtenue en appliquant l'équation (6.11).

Ce processus est comparable à la normalisation divisive de Heeger [Hee93] bien qu'ici les réponses soient renforcées uniquement selon l'orientation. Il est également important de remarquer que ce calcul de la réponse par bande 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 de Li et Zaidi décrits au chapitre 5 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.

6.3 Représentation de la surface locale et récupération de la forme finale

Cette section décrit comment récupérer les paramètres géométriques (les angles de tilt et slant) à partir de la variation de fréquence. Différentes techniques ont été proposées. Par exemple Super et Bovik dans [SB95a] effectuent une recherche exhaustive du tilt et du slant afin de minimiser la variance des paramètres dans une version retro-projetée de la surface. Cette méthode dépend du pas d'échantillonnage des angles dans l'étape de recherche et est coûteuse en calcul. Les méthodes développées par Hwang *et al* dans [HLC98] et Lelandais *et al* dans [LBP05] sont basées sur la mise en correspondance des échelles locales (inverse de la fréquence locale) avec une courbe parabolique d'où le tilt et le slant peuvent être directement extraits. Cependant cette méthode n'est pas adaptée à l'estimation des surfaces courbes.

Ici nous présentons une autre méthode pour retrouver les paramètres géométriques des surfaces. Nous établissons d'abord la relation entre la fréquence locale de la surface (perçue) et la fréquence locale de l'image en projection perspective. Nous établissons ensuite les expressions permettant d'obtenir le tilt et le slant. Nous décrivons enfin la méthode d'estimation de l'orientation de sous-régions locales supposées planes et composées de plusieurs imagettes. Cette méthode permet de résoudre le problème de l'estimation de la forme par la texture sans calcul intensif.

6.3.1 Relations géométriques

La figure 6.8 présente le système de coordonnées d'une projection perspective associé à une surface plane (voir Section 2.3). (x_w, y_w, z_w) représentent les coordonnées du monde, (x_s, y_s), les coordonnées de la surface vue et (x_i, y_i), les coordonnées de l'image projetée. 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. $zw0$) est la coordonnée de l'image

(resp. de la surface) sur l'axe z_w. On note $dzw0$ la distance entre l'image et la surface . τ 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. σ est l'angle de slant qui est l'angle entre l'axe z_w et la normale à la surface en $zw0$ et sa valeur est comprise entre 0 et $\pi/2$.

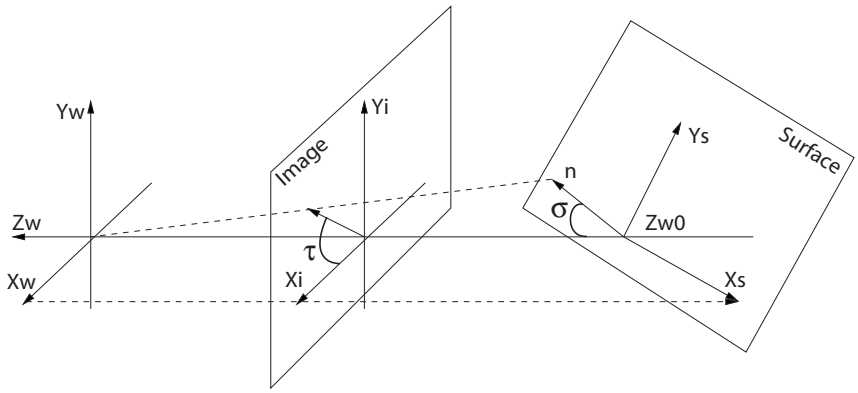


FIG. 6.8 – Modèle de projection perspective

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 [SB95a]) :

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{\begin{bmatrix} (cos(\sigma) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (cos(\tau) & sin(\tau) \\ -sin(\tau) & cos(\tau) \end{bmatrix}}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{A}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{6.13}$$

avec $a_i = \dfrac{-sin(\sigma)sin(\tau)x_i + cos(\tau)sin(\sigma)y_i + dcos(\sigma)}{d + dzw0}$ 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_{Li} s'exprime par :

$$I_{Li}(f_i, x_i) = \int_u i_i(u) w_i(u - x_i) e^{-j2\pi(u-x_i)^t f_i} du \tag{6.14}$$

où i_i est l'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 :

$$i_s(v) w_s(v - x_s) = i_i(u) w_i(u - x_i) \tag{6.15}$$

En appliquant la transformée de Fourier inverse de $i_s(v)w_s(v-x_s)$ dans 6.14, on obtient :

$$I_{Li}(f_i, x_i) = \int_u \left(\int_{f_s} I_{Ls}(f_s) e^{j2\pi(v-x_s)^t f_s} \right) e^{-j2\pi(u-x_i)^t f_i} du df_s \tag{6.16}$$

$$= \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi((v-x_s)^t f_s - (u-x_i)^t f_i)} du df_s \tag{6.17}$$

D'après 6.13, l'approximation au premier ordre de $(v - x_s)^t$ donne :

$$(v - x_s)^t = \frac{1}{a_i}\left(I - \frac{\nabla a_i x_i^t}{a_i}\right) A^t (u - x_i)^t = R^t (u - x_i)^t \tag{6.18}$$

En remplaçant 6.18 dans 6.17, on obtient la relation entre I_{Li} et I_{Ls} :

$$I_{Li}(f_i, x_i) = \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi(u-x_i)^t (R^t(x_i) f_s - f_i)} du df_s \tag{6.19}$$

$$= \int_{f_s} I_{Ls}(f_s)\delta(R^t(x_i) f_s - f_i) df_s = \frac{1}{|det(R)|} I_{Ls}(R^{-t}(x_i) f_i) \tag{6.20}$$

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

$$f_i = R^t(x_i) f_s \approx \frac{1}{a_i}\left(I - \frac{\nabla a_i x_i^t}{a_i}\right) A^t f_s \tag{6.21}$$

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 6.21, l'expression de la variation locale de fréquence de l'image est :

$$df_i = -\frac{1}{ai}\left[\nabla^t a_i dx_i + \nabla a_i dx_i^t\right] f_i \tag{6.22}$$

La fréquence de l'image f_i peut s'exprimer en coordonnées polaires par :

$$f_i = v_i [cos(\varphi_i)\ sin(\varphi_i)]^t \tag{6.23}$$

L'équation 6.22 devient :

$$\begin{aligned} df_i &= dv_i [cos(\varphi_i)\ sin(\varphi_i)]^t + v_i [-sin(\varphi_i)\ cos(\varphi_i)]^t d\varphi_i \\ &= -\frac{\nabla^t a_i dx_i}{a_i} v_i [cos(\varphi_i)\ sin(\varphi_i)]^t - \frac{1}{a_i}\nabla a_i dx_i^t v_i [cos(\varphi_i)\ sin(\varphi_i)]^t \end{aligned} \tag{6.24}$$

Si l'on considère le gradient df_i dans la direction φ_i, en multipliant par $[cos(\varphi_i)\ sin(\varphi_i)]$ l'équation 6.24 devient :

$$dv_i = -\frac{\nabla^t a_i dx_i}{a_i} v_i - \frac{1}{a_i}[cos(\varphi_i)\ sin(\varphi_i)]\nabla a_i dx_i^t [cos(\varphi_i)\ sin(\varphi_i)]^t v_i \tag{6.25}$$

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

$$\frac{dv_i}{v_i} = -\frac{3}{2}\frac{\nabla a_i^t dx_i}{a_i} \tag{6.26}$$

soit

$$dln(v_i) = -\frac{3}{2}\frac{sin(\sigma)\,[-sin(\tau)\ cos(\tau)]^t\,[dx_i dy_i]}{a_i} \tag{6.27}$$

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 :

$$tan(\sigma) = \frac{|dln(v_i)|f}{\frac{3}{2}|dx| - |dln(v_i)|\,(-sin(\tau)x_i + cos(\tau)y_i)} \tag{6.28}$$

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

$$\sigma \propto |dln(v_i)| \tag{6.29}$$

6.3.2 Extraction de la forme

Les relations géométriques obtenues permettent de calculer les angles de tilt et de slant à chaque position spatiale après avoir calculé le gradient de fréquence en cette position. La méthode d'estimation est basée sur l'extraction locale de la moyenne des gradients de fréquence sur des sous-régions de la surface composée de l'ensemble des estimations locales de fréquence (i.e un moyennage sur un ensemble d'imagettes) (Figure 6.9). Nous supposons que la surface couverte par la sous-région est plane c'est-à-dire que les angles de tilt et slant sont supposés constants sur toute cette sous-région. Les angles sont calculés en utilisant l'équation 6.29. Ce processus est répété sur toutes les sous-régions sur l'ensemble de 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. Dans le cas d'une surface plane, le tilt et le slant moyens sont calculés à partir de l'ensemble des estimations locales (par sous-régions). Dans le cas d'une surface courbe, les estimations locales définissent la normale à la surface en chaque position (i.e en chaque centre de sous-région).

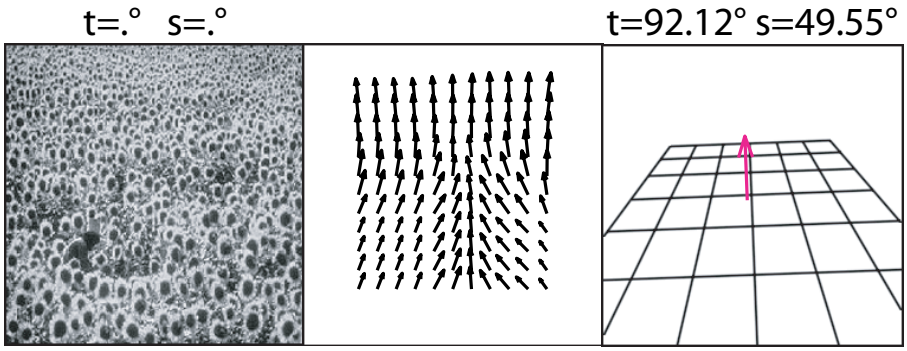


FIG. 6.9 – Estimations sur l'ensemble d'une image ; à gauche : image d'un champ de tournesols ; au milieu : estimations locales du tilt (orientation des flèches) et du slant (longueur des flèches) localisées au centre des imagettes ; à droite : estimation finale de la surface supposée plane et représentée par un quadrillage projeté de manière équivalente (en utilisant les angles d'orientation estimés).

L'information de forme est obtenue sans aucune hypothèse sur les statistiques des orientations, c'est-à-dire sans hypothèse d'isotropie par exemple. Cette méthode ne requière qu'une

hypothèse d'homogénéité, ce qui correspond à supposer une stationnarité locale (ou faible stationnarité) dans les statistiques spatiales des composantes fréquentielles. Cette méthode ne repose pas non plus sur une technique d'optimisation. La forme est extraite de manière complètement *feedforward*. Enfin cette approche peut être reliée à la structure connue des cellules corticales dédiées à l'analyse des gradients tels que les gradients de texture [TSNT02] ou le flux optique [GH03] dans les aires supérieures telle que V3 ou MT (voir Chapitre 3.5).

6.4 Résultats

L'algorithme proposé pour extraire la forme à partir de la texture basé sur l'estimation de la fréquence moyenne locale peut être vu comme une combinaison successives de réponses des filtres corticaux et un moyennage local. La complexité est linéaire pour l'étape de filtrage rétinien, l'étape de combinaison des filtres et l'éatpe finale d'extraction de la forme de la surface. L'étape la plus coûteuse est le calcul de réponse des filtres qui nécessite une transformation de Fourier sur toutes les imagettes. Ceci dépend du nombre total d'imagettes locales (ici $21X21 = 441$ correspondant à une taille de $96X96$ pixels avec un décalage de 8 pixels). Un niveau asymptotique de la précision est atteint pour $7X7 = 49$ filtres (fréquence et orientation) ce qui correspond au nombre de filtres de V1 [SW90]. La taille des sous-régions pour le calcul des orientations locales a été fixé à $10X10$ imagettes. Pour tous les résultats et les tests présentés, tous ces paramètres sont fixés à ces valeurs. Avec un processeur à 2GHz, l'estimation de la forme d'une texture de taille $256X256$ pixels prend approximativement 1 minute avec une implantation basique en Matlab.

6.4.1 Evaluation sur la base de Super et Bovik

Nous présentons d'abord une évaluation de la précision de la méthode proposée avec deux autres techniques développées par Super et Bovik [SB95a] et Hwang *et al* [HLC98].

La figure 6.10 présente les différents résultats obtenus sur la base de textures de Super et Bovik. Dans [HLC98] une comparaison de la précision des deux méthodes est présentée sur cette base : l'erreur moyenne obtenue par Super et Bovik sur l'estimation du tilt et du slant est respectivement de 3.70^o et de 2.84^o ; Hwang *et al* obtiennent une erreur moyenne respectivement de 1.75^o et de 2.18^o. L'algorithme proposé ici atteint une erreur moyenne de 2.41^o 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^o 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 $dzw0$ de la figure 6.8).

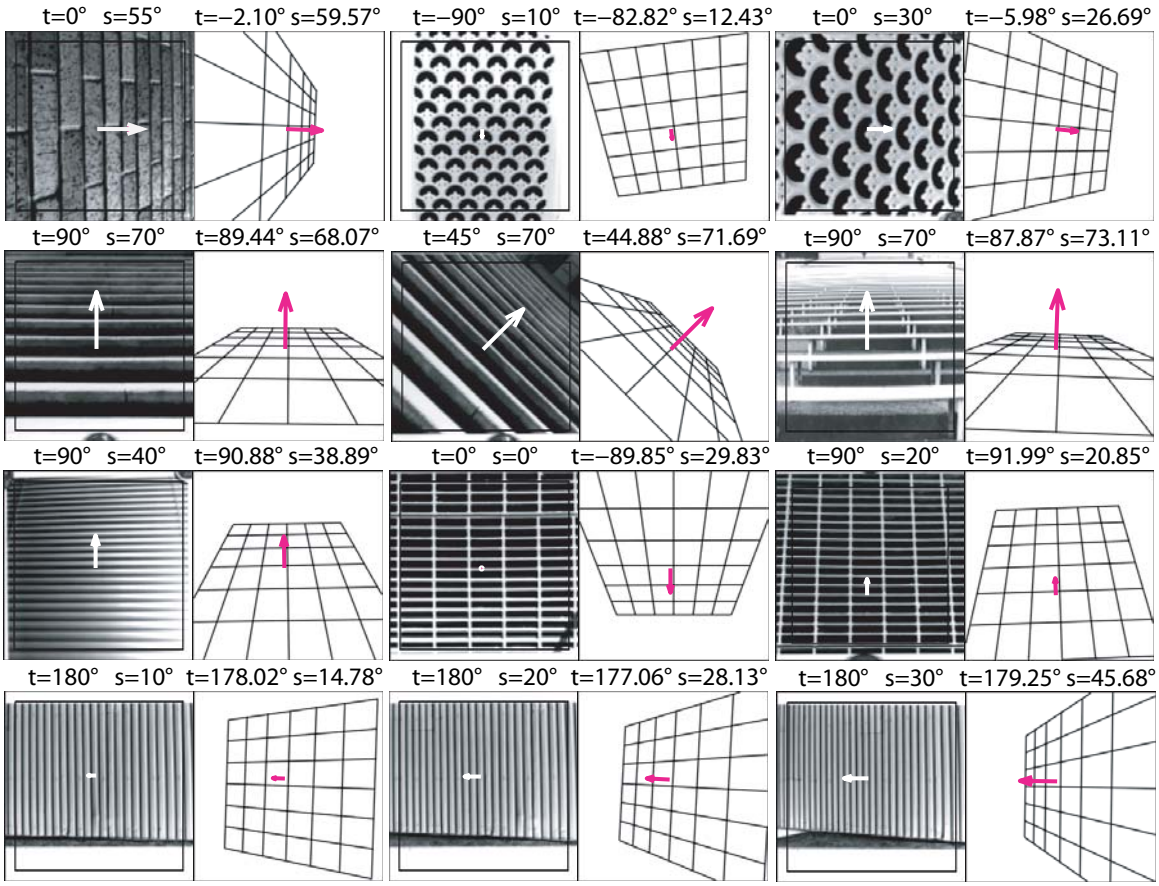


FIG. 6.10 – Résultats obtenus sur la base de texture de Super et Bovik [SB95a].

6.4.2 Evaluation sur une grande base de textures

Pour évaluer la robustesse de notre méthode sur différentes caractéristiques de texture, nous avons crée une base composée de 208 exemples de textures (la plupart sont tirées de la collection de textures de Brodatz, certaines sont des textures artificiellement crées, d'autres enfin sont tirées d'images naturelles) [1].

Chaque exemple est d'abord projeté avec un tilt à 0^o suivant 3 valeurs de slant : 30^o, 45^o et 60^o. De même chaque exemple est projeté pour un slant fixé à 45^o suivant 3 valeurs de tilt : 0^o, 45^o et 90^o. L'erreur d'estimation est moyennée sur l'ensemble des exemples de texture et sur toutes les projections. Nous obtenons une précision moyenne de 18.15^o (variance 25.61^o) sur l'estimation du tilt et de 12.35^o (variance 8.1^o) 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 6.11 présente des résultats d'estimation de l'orientation des surfaces texturées provenant de la base utilisée. Des exemples présentant différents types d'irrégularités ont été choisis.

La table 6.12 à gauche présente le détail des résultats pour les différentes configurations d'orientation. Sur la figure 6.12 à droite, les exemples de textures présentés correspondent aux cas où l'estimation de l'orientation est la plus difficile pour la méthode développée. De mauvaises estimations sont obtenues sur les textures présentant uniquement de très hautes fréquences qui, une fois inclinées en profondeur, induisent des problèmes d'aliasing dus à la résolution et se reportant sur l'analyse spectrale (Figure 6.12 à droite première ligne). L'algorithme atteint également ses limites lorsqu'il est appliqué sur des textures contenant uniquement de très basses fréquences spatiales Figure 6.12 à droite deuxième ligne). Comme les imagettes ont une taille constante et prédéfinie, les variations spatiales peuvent ne pas être englobées par elles. Finalement il est sensible à de fortes irrégularités de position et de taille (i.e de fortes non-stationnarités locales) créant des *manques* d'information de fréquence (Figure 6.12 à droite troisième ligne à gauche) et à la violation de l'hypothèse de fréquence moyenne constante (Figure 6.12 à droite troisième ligne à droite). Pour résoudre ces deux dernières limitations, deux mécanismes peuvent être combinés à l'estimation de la fréquence : une adaptation automatique de la taille de l'imagette en fonction de la fréquence moyenne ; une méthode de régularisation permettant un lissage robuste de la surface formée par l'ensemble des estimations locales de la fréquence avant l'étape d'extraction de la forme.

[1] cette base est accessible à l'adresse suivante : http ://www.lis.inpg.fr/massot

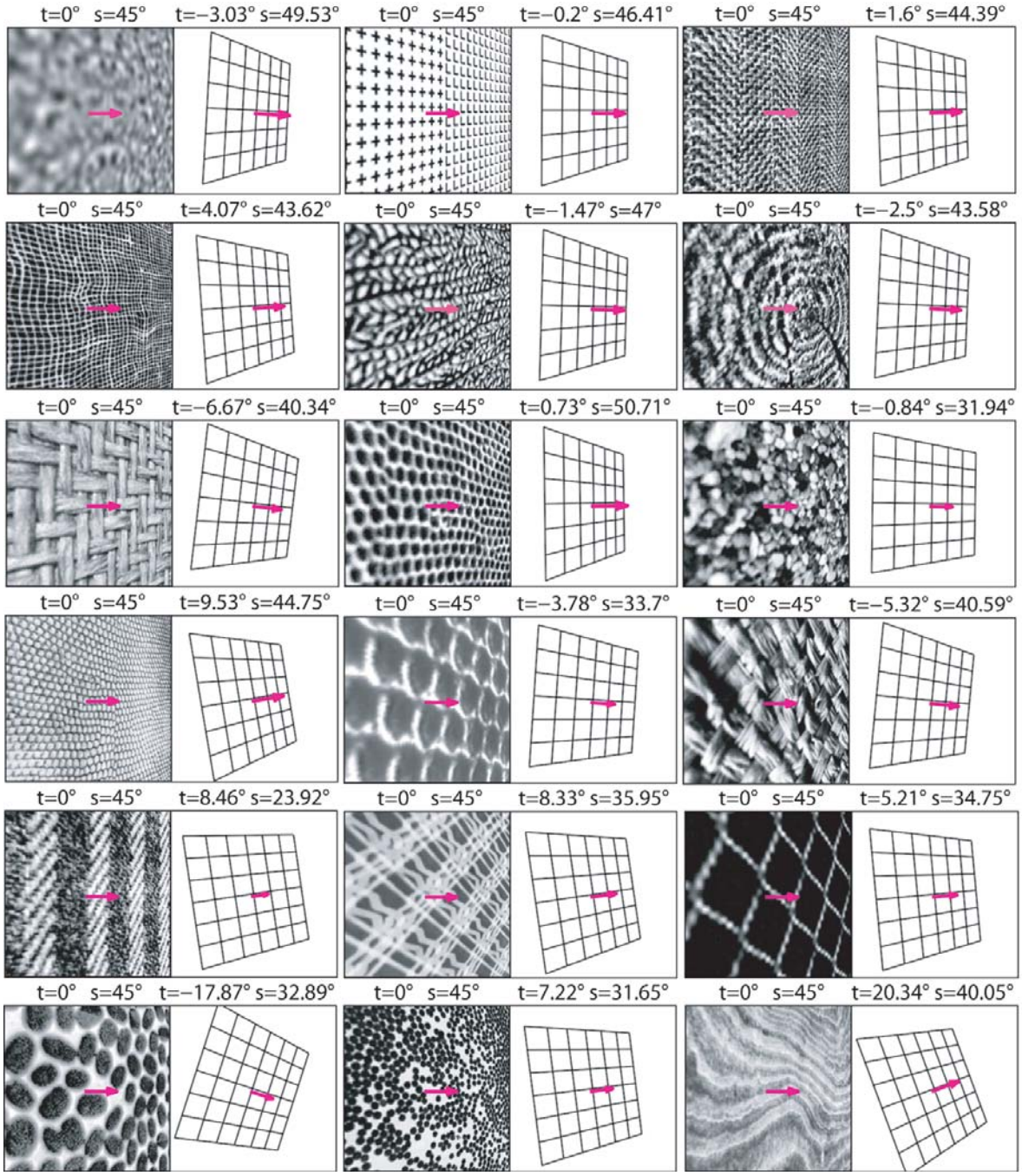


FIG. 6.11 – Exemples de résultats obtenus sur la base de textures (composée de 208 exemples, chacun projeté suivant 5 configurations de tilt et slant différentes conduisant à un total de 1040 textures) ; les exemples sélectionnés ont été choisis afin de donner un aperçu des différentes irrégularités présentes dans la base ; les résultats sont organisés de haut en bas suivant une précision décroissante.

σ	$\tau = 0^o$	
	tilt	slant
30^o	26.65 (34.62)	7.21 (5.25)
45^o	17.31 (27.85)	11.13 (7.90)
60^o	15.21 (23.12)	18.83 (11.23)

τ	$\sigma = 45^o$	
	tilt	slant
45^o	16.97 (22.26)	12.66 (8.27)
90^o	14.61 (20.19)	11.96 (7.77)

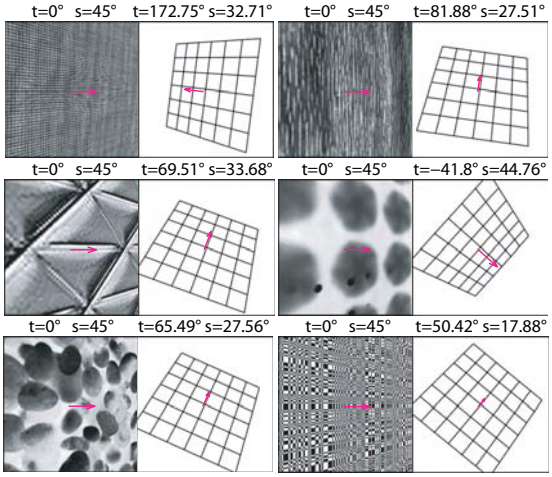


FIG. 6.12 – Résultats détaillés ; à gauche : erreurs moyennes sur les estimations obtenues sur les 5 configurations d'orientation différentes sur l'ensemble des textures de la base utilisée (la variance est indiquée entre parenthèses) ; à droite : exemples de textures où une bonne estimation est difficile à obtenir.

Pour évaluer la qualité de la précision obtenue, nous comparons nos résultats avec ceux obtenus sur une centaine de textures par la technique développée par Hwang *et al* [HLC98] et par celle développée par Lelandais *et al* [LBP05]. 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 développés 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 au faible inclinaison 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^o 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 respectivement une précision de 15.9^o (variance 14.9^o) et de 31.4^o (variance 23.3^o). La précision obtenue par notre méthode est donc comparable avec, en plus, moins de dispersion (variance plus faible). Ces résultats permettent d'affirmer que les performance de la méthode proposée 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 comme la méthode n'incorpore pas de techniques d'optimisation, son coût calculatoire est inférieur aux deux autres techniques présentées.

6.4.3 Evaluation sur une base de scènes naturelles

La figure 6.13 présente les résultats obtenus sur des scènes naturelles.

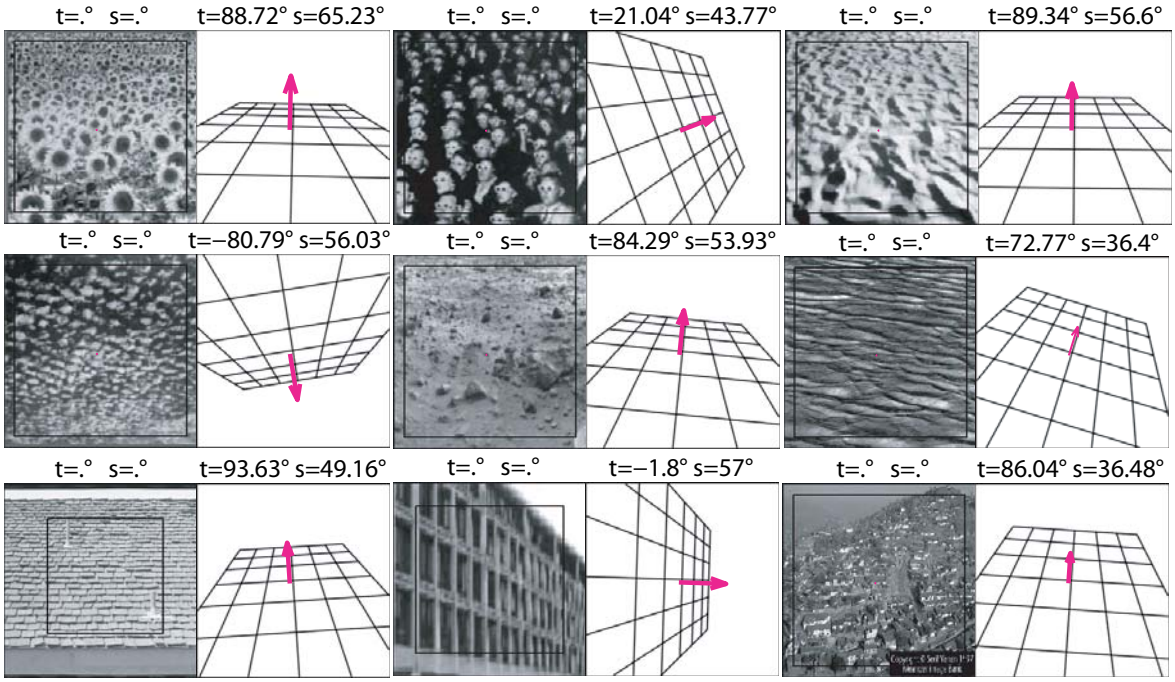


FIG. 6.13 – Résultats obtenus sur des scènes natuelles tirées de [GL96] et d'images non référencées ; aucune estimation théorique n'accompagne ces images ; les résultats obtenus sur les 4 premières images se comparent favorablement aux résultats obtenus par Lindeberg et Garding dans ([GL96]).

Les 4 premières images sont tirées de [GL96] sur lesquelles la technique proposée obtient des résultats très proches. Les images restantes sont des exemples non référencés.

La figure 6.14 présente aussi des estimations sur différentes régions de scènes multitexturées.

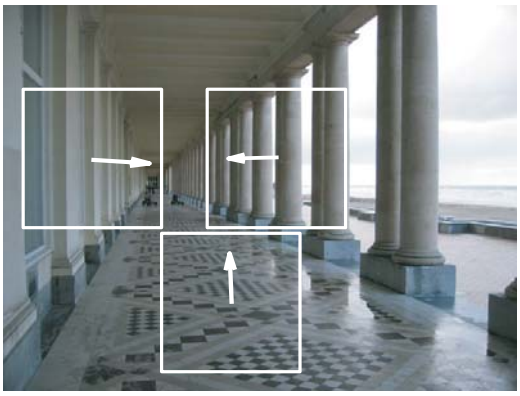


FIG. 6.14 – Résultats obtenus sur une scène multitexturée

Il est important de noter que les résultats ont été obtenus avec exactement les même paramètres que dans les tests précédents (sur la base de textures) excepté la taille de la zone analysée qui est adaptée manuellement afin de respecter l'hypothèse d'homogénéité. Tous ces résultats montrent la capacité de la méthode à traiter des textures directement extraites de scènes naturelles afin de récupérer une information de perspective.

6.4.4 Evaluation sur une base de surfaces courbes

La figure 6.15 présente des résultats obtenus sur des surfaces courbes.

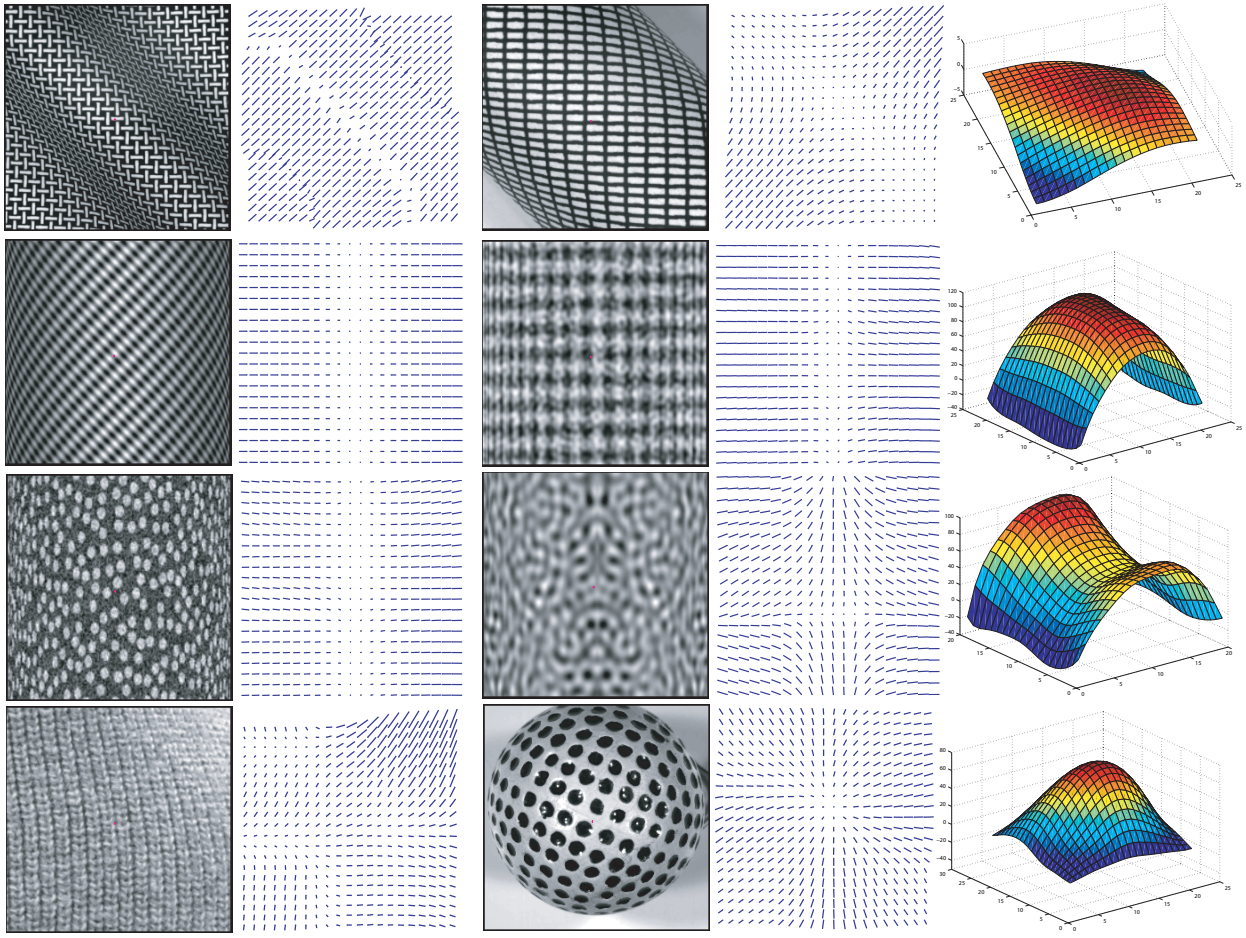


FIG. 6.15 – Résultats obtenus sur des surfaces courbes ; chaque exemple présente la texture originale, la carte-aiguille (chaque trait est dans la direction du tilt local et la longeur est proportionnelle au slant estimé) et pour les exemples de la dernière colonne, la reconstruction 3D de la forme à partir des informations de tilt et de slant est présentée (les codes permettant d'obtenir la carte-aiguille et la reconstruction à base de *shapelets* sont distribués par Peter Kovesi [Kovns]).

Les deux premiers exemples sont tirés de la base de Super et Bovik. Les 4 exemples suivant présentent des surfaces cylindriques avec une augmentation progressive de l'irrégularité. Le

premier et le quatrième exemple proviennent des travaux de Sakai et Finkel [SF95] et ont été spécialement crées pour mettre en défaut les méthodes basées sur l'estimation de moments (par exemple la méthode de Super et Bovik [SB95b] ou celle de Loh et Kovesi [LK05]). Notre algorithme est capable de retrouver une forme cylindrique dans chaque cas. Les deux derniers exemples sont tirés des travaux de Clerc et Mallat [CM02] et la forme obtenue est très proche de leur estimations.

6.4.5 Evaluation sur des stimuli psychophysiques

La figure 6.16 présente les résultats obtenus sur des exemples créés pour des expériences en psychophysique.

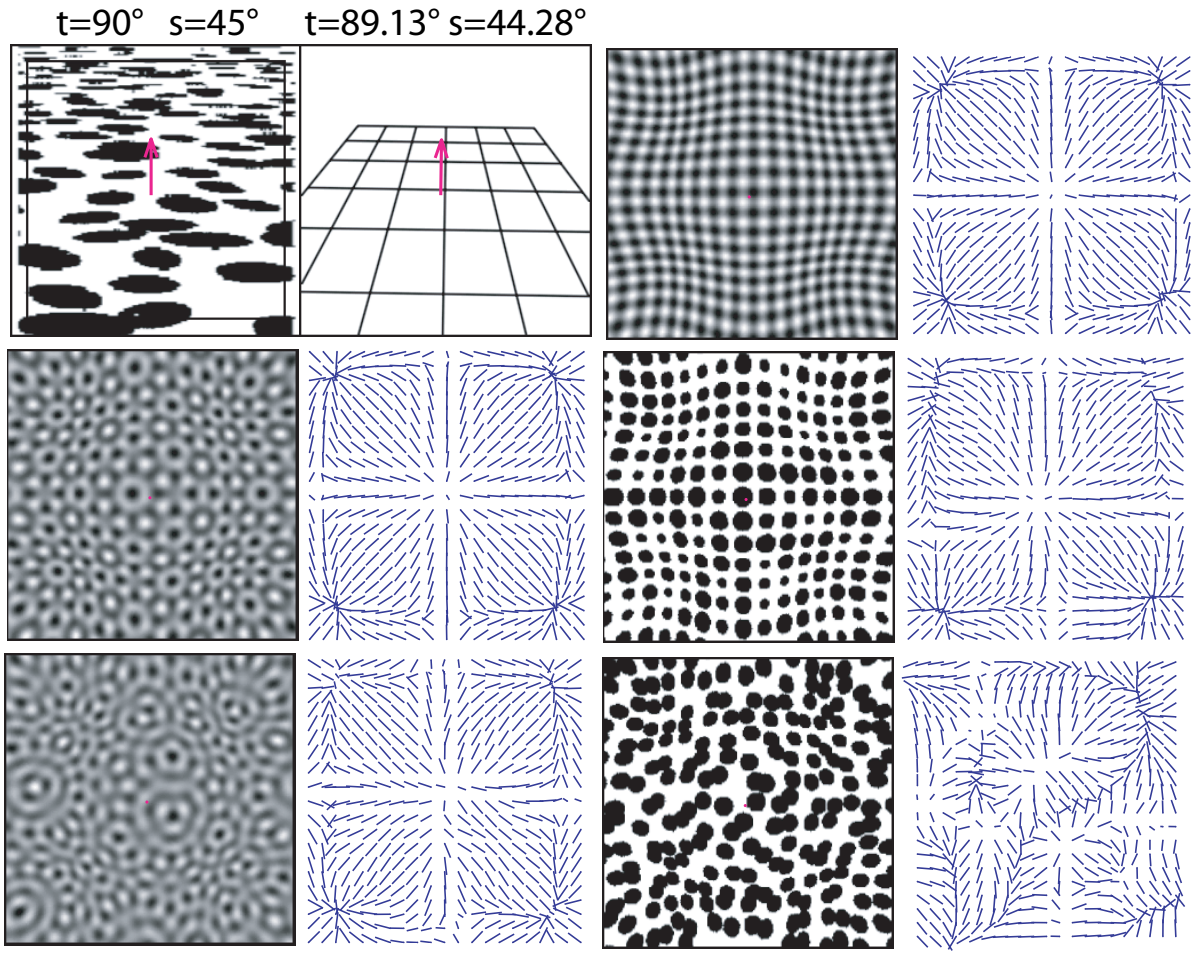


FIG. 6.16 – Résultats obtenus sur des surfaces planes et courbes provenant de stimuli utilisés dans travaux en psychophysiques (voir [Kni98b] [LZ04]).

La première colonne présente un exemple tiré des travaux de Knill [Kni98b] et représente une surface plane inclinée (le tilt est égale à 90^o et le slant, à 45^o) couverte par des ellipses avec une taille et une position aléatoire. Les exemples suivants sont tirés des travaux de Li et Zaidi [LZ04]. Toutes les formes sont identiques, seules les caractéristiques de la texture sont

modifiées. Pour le dernier exemple l'algorithme est incapable de retrouver la forme mais de manière similaire à notre perception. Nous observons également que l'algorithme peut parfaitement retrouver la forme du troisième exemple, correspondant à une texture ne présentant qu'une variation de fréquence avec l'ensemble des indices d'orientation retirés (voir la section 4.4). Cependant dans ce cas le système visuel est incapable de retrouver la forme. Cela suggère qu'une pondération importante est accordée à l'indice de courbure pour l'estimation d'une surface courbe dans un processus de fusion d'indices. Dans le cas du stimulus considéré, le système visuel peut être fortement biaisé par le manque d'information en orientation en indiquant une surface plane frontoparallèle, ce qui ne permettrait pas d'utiliser l'information de fréquence indiquant, elle, le changement local de profondeur.

6.5 Conclusion

Nous avons présenté un modèle d'extraction de la fréquence locale sur des images naturelles et artificielles. Celui-ci peut être vu comme un modèle biologiquement plausible du cortex primaire (aire V1). En effet il est basé sur une analyse du spectre d'amplitude sans l'utilisation de l'information de phase. Il est robuste aux translations locales et réalise un lissage des statistiques de la texture. L'échantillonnage du spectre est réalisé à l'aide de filtres *log-normaux* qui sont à la fois bien adaptés au zoom et à la rotation et qui sont une bonne approximation des réponses des cellules corticales. Basé sur ces filtres corticaux, nous avons développé un modèle cortical dédié à l'extraction de la fréquence moyenne locale. Ce modèle impose une largeur de bande relative décroissante au banc de filtres *log-normaux*, de manière similaire à l'organisation des cellules du cortex visuel humain. Ceci suggère une explication empirique à cette caractéristique particulière. Ce modèle cortical réalise également la séparation entre les informations de fréquence et d'orientation, conformément à nos résultats en psychophysique. Nous appliquons ce modèle au problème d'extraction de la forme par la texture et, sous une hypothèse d'homogénéité, montrons qu'il est capable de récupérer une information de forme précise même dans le cas de textures irrégulières. Les performances sont comparables aux algorithmes spécialement développés en vision par ordinateur. Enfin il reproduit en partie la perception sur les stimuli de Li et Zaidi et suggère l'importance de l'indice de variation d'orientation associé à la courbure dans le cas de surfaces courbes.

La figure 6.17 présente la modèle cortical d'analyse des fréquences. Celui-ci peut être vu comme une simple combinaison des cellules dans une architecture uniquement *feedforward*. Il est divisé en 4 étapes :

1. Le modèle de V1 :
   Dans chaque imagette les réponses du banc de filtres sont calculées. Cela modélise la décomposition de l'ensemble de l'image en fréquence et en orientation réalisée par le cortex V1. Figure 6.17.1 montre la réponse de chaque filtre défini à une fréquence centrale et une orientation spécifique, associées au centre de l'imagette étudiée. Il est alors possible d'observer l'évolution spatiale de la réponse en énergie pour chaque fréquence et chaque orientation.
2. Le calcul des réponses par bandes de fréquence :
   La figure 6.17.2 présente le modèle de calcul des bandes de fréquence. Il s'agit d'appliquer la normalisation donnée par la formule 6.2.3.2 sur l'ensemble des orientations à une fréquence donnée. La réponse par bande de fréquence correspond à la somme des réponses obtenues à chaque fréquence centrale. Cela permet de rehausser les infor-

mations contenues autour de chaque fréquence centrale lorsque celle-ci est significative. Cela contribue en même temps à la séparation (l'indépendance) entre les informations de fréquence et d'orientation.

3. La combinaison des filtres :
   La réponse de chaque bande de fréquence est ensuite combinée avec une bande voisine (ici la précédente dans l'ordre des fréquences centrales) en utilisant l'équation 6.10. Cela donne une estimation de la fréquence moyenne locale localisée au centre de l'imagette étudiée et autour des deux fréquences centrales étudiées. La figure 6.17.3 montre la combinaison de bande de fréquence se limitant à une combinaison simple de filtres.
4. L'estimation finale sur l'ensemble de la surface :
   La combinaison finale est obtenue grâce à l'équation 6.11. La figure 6.17.4 monte que cela correspond à une somme pondérée des estimations locales de la fréquence moyenne. Le poids est directement la réponse de la bande de fréquence associée (ici en considérant la bande avec la fréquence centrale inférieure). Cela conduit à un modèle de combinaison de cellules très simple pour l'extraction de la fréquence moyenne. Cette méthode s'appuie sur une estimation à large bande en tirant partie de l'ensemble des estimations obtenues aux différentes fréquences centrales conduisant à une estimation finale robuste sans perte d'information.

Une étape de régularisation robuste associée à l'estimation des gradients de fréquence sur l'ensemble de la surface pourra permettre d'améliorer la robustesse du modèle face aux irrégularités de la texture. Une seconde amélioration est l'adaptation du modèle pour prendre en compte les variations de d'orientation, notamment pour estimer la courbure des surfaces. Pour cela il suffira de considérer les bandes d'orientation au lieu des bandes de fréquence et d'adapter la normalisation corticale en conséquence. Les deux modèles obtenus pourront donc être très similaires et rendra aisée la combinaison des estimations obtenues à partir des deux indices. Ainsi deux mécanismes indépendants pourront être développés, chacun dédié à l'analyse d'un type de gradient conformément à l'hypothèse de Li et Zaidi [LZ04].

Le lecteur remarquera que ce modèle d'analyse correspond également au modèle d'analyse des scènes naturelles, décrit au chapitre 2.1. L'architecture que nous présentons est donc générale et peut s'adapter aussi bien aux problèmes de catégorisation et d'estimation de l'organisation spatiale des scènes naturelles qu'à l'analyse de texture et à le perception 3D.

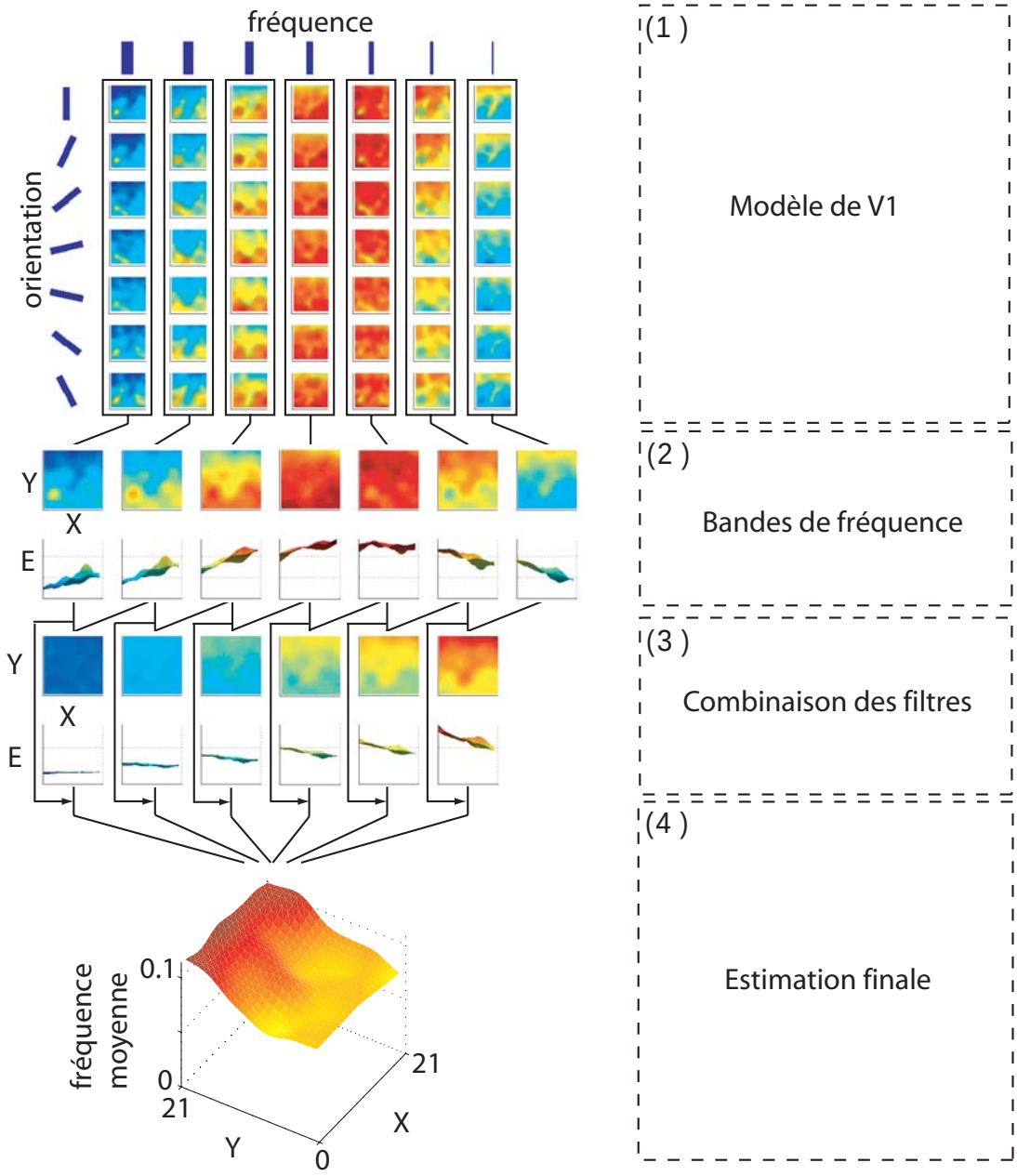


FIG. 6.17 – Modèle cortical d'analyse des fréquences (appliqué sur l'image du champ de tournesols après les étapes de prétraitement présentés à la figure 6.7) : (1) décomposition en fréquence et en orientation de l'information ; chaque figure représente la réponse individuelle d'un filtre (pour une fréquence centrale et une orientation donnée) appliqué sur chaque imagette recouvrant l'ensemble de l'image ; (2) calcul des bandes de fréquence après avoir effectué la normalisation par orientation ; (3) combinaison des bandes de fréquence en utilisant l'équation 6.10 ; (4) estimation finale de la fréquence sur l'ensemble de l'image en utilisant l'équation 6.11 ; grâce à la normalisation par orientation le poids attribué à chaque combinaison de bande pour l'estimation finale est directement la réponse la valeur de la réponse de la bande de fréquence correspondante ce qui conduit à un schéma de connection entre cellules extrêmement simple.

Chapitre 7

Conclusions et perspectives

Dans ce travail nous nous intéressons au fonctionnement du système visuel et à sa capacité à extraire l'information 3D dans les scènes naturelles. Les modèles actuels de catégorisation rapide reposent sur une décomposition de l'image en spectres d'amplitudes locaux. Ce modèle permet d'analyser les statistiques locales de fréquence et d'orientation tout en s'affranchissant de l'information de phase. Dans notre approche, nous faisons en plus l'hypothèse qu'une segmentation en régions de l'image est effectuée au préalable afin de ne considérer que des surfaces recouvertes d'une texture homogène. Lors du passage du monde 3D au plan 2D de l'image, la texture subit des déformations dues à la projection perspective. Afin de retrouver la configuration spatiale de la surface en 3D, nous cherchons à extraire les angles d'inclinaison (slant) et d'orientation indiquant la direction en profondeur (tilt) en différentes positions de la surface. Il s'agit ainsi de récupérer la forme de la surface à partir de la texture.

Notre travail est divisé en deux parties complémentaires : des expérimentations psychophysiques afin d'étudier les indices de gradient de fréquence et de perspective linéaire et le développement d'un modèle biologiquement plausible basé sur les filtres *log-normaux* servant de modèle des cellules complexes.

Afin d'étudier précisément l'influence relative du gradient de fréquence et de la perspective linéaire, nous avons crée des stimuli spécifiques. Ceux-ci représentent une texture construite à partir de la juxtaposition de masques de Gabor. Chaque masque peut être paramétré indépendamment en fréquence et en orientation ce qui permet d'effectuer une séparation totale entre les deux indices. Deux tâches ont été utilisées : la discrimination du slant et la discrimination du tilt. Nos résultats montrent l'importance du gradient de fréquence pour l'estimation du slant ce qui est conforme au fait que la fréquence permet d'estimer la profondeur. La perspective linéaire n'intervient que faiblement et nous émettons l'hypothèse qu'elle interviendrait indirectement en renforçant l'impression de surface plane. L'estimation de la direction en profondeur (tilt) repose sur une combinaison des deux indices. Ces résultats montrent que le système visuel utilise bien ces deux indices pour la perception 3D, confirmant ainsi l'hypothèse de Li et Zaidi. Enfin en perspectives nous proposons de considérer le gradient de fréquence, la perspective linéaire et la courbure comme trois indices séparés de la perception 3D. Une modification simple de nos stimuli pourra permettre de tester facilement la contribution de l'indice de courbure en combinaison avec l'indice de fréquence.

En parallèle nous avons développé un modèle biologiquement plausible d'extraction de la forme à partir de la texture. Celui-ci se base sur l'utilisation de filtres spatio-fréquentiels, les

filtres log-normaux, dont nous montrons les avantages théoriques (filtre à variables séparables ; gain nul ; la réponse migre proportionnellement en présence d'un zoom ou d'une rotation) et leur validité en tant que modèle des cellules complexes. Nous avons développé une méthode permettant d'estimer la fréquence moyenne locale dans l'image et d'extraire finalement le tilt et le slant de la surface. Chaque étape s'appuie sur des données physiologiques et psychophysiques, depuis la rétine jusqu'au cortex primaire V1. Elle peut être considérée comme un modèle plausible d'extraction d'indices bas-niveau, complètement *feed-forward*. Pour estimer la précision et démontrer la robustesse de la méthode, nous avons présenté de nombreux résultats sur différentes bases de données comportant des textures artificielles et naturelles, des scènes naturelles, des surfaces courbes et des stimuli psychophysiques. Nous avons montré la capacité de l'algorithme à prendre en compte différents types d'irrégularités. 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 calculatoire. 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. Une perspective intéressante est la possibilité d'adapter le modèle à l'analyse des variations d'orientation. Le nouveau modèle pourrait extraire l'information liée à la perspective linéaire. Cela permettrait d'obtenir deux mécanismes indépendants spécialisés dans le traitement de chaque indice et de pouvoir combiner facilement leurs estimations, conformément à l'hypothèse de Li et Zaidi.

A l'issu de ce travail, nous avons d'abord mis en évidence l'importance des indices séparés de gradient de fréquence et de perspective linéaire pour la perception 3D. En se basant sur cette approche, nous avons développé un modèle biologiquement plausible d'extraction de la forme à partir de l'indice de fréquence et nous avons montré que les première étapes du système visuel pouvaient réaliser cette fonction. Ceci constitue de premiers éléments de réponse pour découvrir comment le système visuel interprète la 3D présente dans les scènes naturelles.

Une perspective à long terme envisageable à l'issu de ce travail est l'extension de notre étude à des modèles réalisant l'extraction d'autres indices 3D, notamment la disparité binoculaire, la parallaxe de mouvement et la variation d'illumination. En effet pour chacun de ces indices, des modèles basés également sur des filtres spatio-fréquentiels ou spatio-temporels ont été développés par différents auteurs. Ces travaux peuvent être mis en relation avec notre propre modèle. Cela nous permet d'envisager le développement d'un modèle commun d'extraction et d'analyse de gradients spatio-temporels de texture pour la perception 3D. Chaque modèle associé spécifiquement à l'analyse d'un gradient est susceptible de réalisé un codage de l'information sous une forme semblable aux autres modèles. Un processus de combinaison d'indices pourra alors être facilement obtenu. Celui-ci pourra être validé à l'aide des données existantes obtenues en psychophysique. Ce modèle pourra suggérer à son tour des expérimentations psychophysiques et des études en neurophysiologie. Ce modèle représenterait une étape importante dans la compréhension des mécanismes mis en jeu dans le système visuel pour interpréter l'environnement naturel en 3D. C'est ce travail que je me propose de poursuivre en postdoctorat.

Annexe A

Annexe

A.1 Calcul de la variation de fréquence

Afin d'obtenir la modulation de fréquence en fonction l'inclinaison de la surface 3D (initiale), nous établissons la relation entre la fréquence sur cette surface et la fréquence sur l'image (résultat de la projection).

Nous considérons la transformée de Fourier I_{Li} dans une région de l'image $im_i(x)$. Cette région est notée L_i et correspond à une fenêtre spatiale notée $w_i(x)$. Autour de la position spatiale x_i, I_{Li} s'exprime par :

$$I_{Li}(f_i, x_i) = \int_u im_i(u) w_i(u - x_i) e^{-j2\pi(u-x_i)^t f_i} du \tag{A.1}$$

En utilisant la relation $im_i(u)w_i(u - x_i) = im_s(v)w_s(v - x_s)$, avec u la projection de v, im_s, la surface initiale et $w_s(x)$, une fenêtre spatiale prise sur cette surface, la transformée de Fourier inverse donne :

$$im_s(v)w_s(v - x_s) = im_i(a_i A^{-1} v) w_i(a_i A^{-1}(v - x_s)) \tag{A.2}$$

$$= \int_{f_s} I_{Ls}(f_s) e^{j2\pi(v-x_s)^t f_s} df_s \tag{A.3}$$

La tansformée de Fourier dans le plan image autour de x_i donne :

$$I_{Li}(f_i, x_i) = \int_u \left(\int_{f_s} I_{Ls}(f_s) e^{j2\pi(v-x_s)^t f_s} \right) e^{-j2\pi(u-x_i)^t f_i} du df_s \tag{A.4}$$

$$= \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi((v-x_s)^t f_s - (u-x_i)^t f_i)} du df_s \tag{A.5}$$

Exprimons $(v - x_s)^t$ en fonction de $(u - x_i)^t$.
Le développement au premier ordre de dx_s donne :

$$dx_s^t \approx dx_i^t(\frac{A}{a_i})^t - dx_i^t(\frac{\nabla a_i x_i^t}{a_i^2})A^t \tag{A.6}$$

$$\approx dx_i^t \frac{1}{a_i}\left(I - \frac{\nabla a_i x_i^t}{a_i}\right)A^t \tag{A.7}$$

$$= dx_i^t R^t(x_i) \tag{A.8}$$

En remplaçant dans A.5, nous obtenons la relation entre I_{Li} et I_{Ls} :

$$I_{Li}(f_i, x_i) = \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi(u-x_i)^t(R^t(x_i)f_s - f_i)} du df_s \tag{A.9}$$

$$= \int_{f_s} I_{Ls}(f_s)\delta(R^t(x_i)f_s - f_i)df_s \tag{A.10}$$

$$= \frac{1}{|det(R)|} I_{Ls}(R^{-t}(x_i)f_i) \tag{A.11}$$

L'évaluation du dirac δ donne la relation finale entre f_i et f_s :

$$f_i = R^t(x_i)f_s \approx \frac{1}{a_i}\left(I - \frac{\nabla a_i x_i^t}{a_i}\right)A^t f_s \tag{A.12}$$

Le lecteur pourra remarquer la présence d'un facteur de correction $\left(-\frac{\nabla a_i x_i^t}{a_i^2}\right)$ supplémentaire correspondant à l'estimation au premier ordre de la variation de position. Bien que plus faible que $\frac{1}{a_i}$ seul, ce terme n'est cependant pas négligeable.

A.2 Calcul de la variation d'orientation

La modulation d'orientation est simplement donnée par l'orientation d'une droite après sa projection dans l'image.

Nous considérons une droite sur la surface 3D initiale orientée d'un angle α. Son équation peut s'écrire :

$$cos(\alpha)x_s + sin(\alpha)y_s - ps = 0 \tag{A.13}$$

où ps est égale à $xs * cos(\alpha) + ys * sin(\alpha))/(d + zw0aux)$, l'équation initiale de la droite sur la surface 3D. Si nous remplaçons les coordonnées (x_s, y_s) par leur projection sur l'image, notées (x_i, y_i), nous obtenons :

$$\begin{aligned}&(dcos(\alpha)cos(\sigma)cos(\tau) - dsin(\alpha)sin(\tau) + pssin(\sigma)sin(\tau))x_i + \\ &(dcos(\alpha)cos(\sigma)sin(\tau) + dsin(\alpha)cos(\tau) - pssin(\sigma)cos(\tau))y_i - pscos(\sigma) \\ &= Cx_i + Sy_i - pscos(\sigma) = 0\end{aligned} \tag{A.14}$$

où τ est le tilt, σ est le slant, d est le paramètre de la projection perspective (la distance entre le centre de la projection et la surface).
L'équation A.14 correspond à l'équation de la droite projetée sur l'image.

Son orientation β à la position (x_i, y_i) est donnée par :

$$\beta = arctan\left(\frac{S}{C}\right) \tag{A.15}$$

avec C et S dépendant de α, σ et τ.

A.3 Commentaires sur les stimuli

La vérification des équation 5.3 et 5.4 peut aussi se faire simultanément sur une surface totale. La figure A.1 présente une comparaison avec une texture complète formée par un ensemble de masques de Gabor avec une fréquence identique et une orientation aléatoire avant projection de la surface. L'image de gauche présente le résultat obtenu par projection de la surface. La seconde image présente le résultat obtenu par simulation de l'inclinaison pour la même texture initiale (masques possédant les mêmes positions spatiales et les mêmes orientations). A la fois la fréquence et l'orientation des stimuli sont manipulées individuellement pour chaque masque. La troisième image représente leur différence et celle-ci est quasiment nulle. Cette vérification montre que les stimuli obtenus sont très similaires à ceux qui seraient obtenus par une projection réelle.

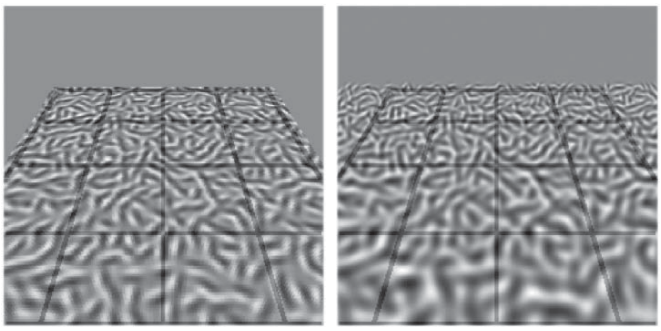


FIG. A.1 – Vérifications simultanées des deux équations 5.3 et 5.4 sur une texture complète ; à gauche : projection réelle ; à droite : simulation en modifiant individuellement la fréquence et l'orientation de chaque masque.

Les figures 5.4, 5.8 et A.1 ne doivent cependant pas tromper le lecteur. Ces figures permettent de vérifier si les équations 5.3 et 5.4 donnent correctement la valeur de la fréquence pour une position spatiale et une projection données. La projection perspective subie par un masque de Gabor modifie à la fois sa taille et ses proportions (effet de compression orthogonalement au tilt). Ce sont ces modifications qui induisent un changement de fréquence locale dans la zone de ce masque. Par contre lorsque la fréquence du masque est modifiée, le masque obtenu n'est pas déformé (pas de compression), son enveloppe reste circulaire mais celle-ci change bien de taille pour garder constante la largeur de bande relative (Equation 5.2). C'est la fréquence qui caractérise le changement du à la projection, de manière équivalent à l'association de l'information de taille et de compression pour la véritable projection d'un texel. Il est important de noter que les stimuli ainsi construits ne sont pas composés d'un ensemble de masques circulaires projetés individuellement représentant chacun un texel. Ils correspondent à une surface couverte par une variation de fréquence continue. En d'autres termes il n'y a

pas de différence entre le fond et les éléments du premier plan. Pour illustrer ce propos, la figure A.2 présente une comparaison entre une texture composée de points Polka aléatoires, la projection d'une texture composée de masques de Gabor et la simulation de la même surface inclinée en appliquant les variations de fréquence et d'orientation aux masques de Gabor en fonction de leur position spatiale.

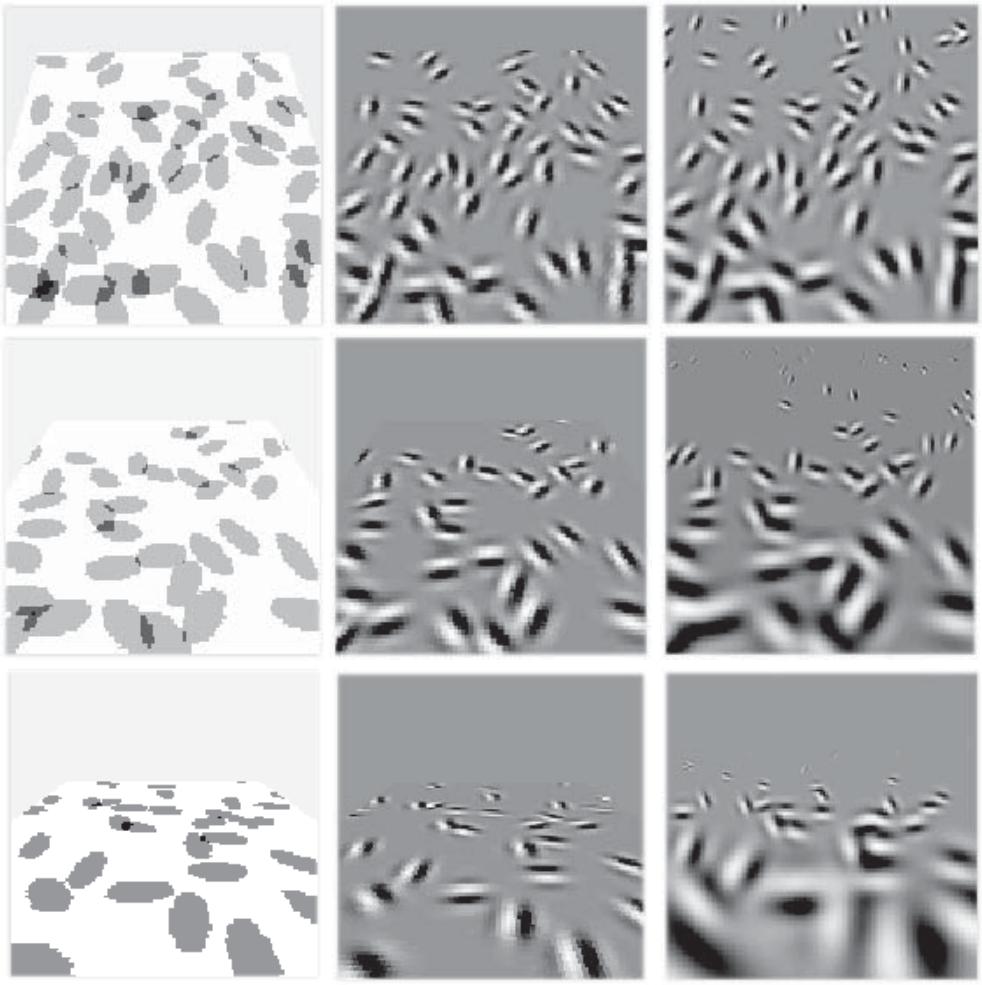


FIG. A.2 – De gauche à droite : comparaison entre la projection d'une texture formée de points Polka , une texture formée de masques de Gabor déformés par la projection de la surface et une texture de masques de Gabor obtenue par la simulation simultanée de la variation de fréquence et de la variation d'orientation ; de haut en bas : chaque ligne correspond à la projection d'une surface plane avec une inclinaison croissante.

Afin de vérifier la séparation entre les deux indices, la figure 5.33 affiche les spectres d'amplitude locaux à différentes positions sur nos stimuli. Pour les stimuli présentant uniquement une variation de fréquence, nous observons bien une expansion du spectre (correspondant au décalage vers les hautes fréquences) sans apparition d'orientations. Pour les stimuli présentant uniquement une variation d'orientation, nous observons bien la présence de deux orientation

principales (une dans la direction du tilt et l'autre orthogonalement à cette direction) sans aucun effet d'expansion du spectre. Pour les stimuli présentant uniquement les deux types de variation, nous observons bien les deux effets simultanément : l'expansion du spectre et la présence de deux orientation principales dans les deux spectres locaux. Ainsi dans les stimuli que nous avons construit, les informations de variation de fréquence et de variation d'orientation sont bien rendues indépendantes.

Publications

[1] C. Massot and J. Hérault. Model of frequency analysis in the visual cortex and the shape from texture problem. *International Journal of Computer Vision*, 2006 (en relecture).

[2] C. Massot and J. Hérault. Extraction de la forme et de la perspective dans des textures artificielles et naturelles par modeles corticaux. *Revue Taitement du Signal*, 2006 (en relecture).

[3] C. Massot and J. Hérault. Cortical area v1 and log-normal filters, application to shape perception from texture gradients. *Conférence internationale Brain Inspired Cognitive Systems (BICS), Grèce*, 2006.

[4] C. Massot and J. Hérault. Recovering the shape from texture using lognormal filters. *Advanced Concepts for Intelligent Vision Systems (ACIVS), Anvers, Belgique*, 2005.

[5] C. Massot and J. Hérault. Extraction de la forme et de la perspective dans des textures artificielles et naturelles par modèles corticaux. *Colloque GRETSI, Louvain-la-Neuve, Belgique*, 2005.

[6] C. Massot and J. Hérault. Extraction d'indices de perspective dans les scènes naturelles par modèles corticaux. *Colloque GRETSI, Paris, France*, 2003.

[7] C. Massot, J. Hérault, and P. Mamassian. Analysis of the combination of frequency and orientation cue in texture orientation perception. *European Conference on Visual Perception (ECVP), La Corogne, Espagne*, 2005.

[8] N. Guyader, C. Massot, and J. Hérault. Modelling global to local cortical interaction for scene analysis and perspective retrieval. *European Conference on Visual Perception (ECVP), Budapest, Hongrie*, 2004.

[9] Z. Hammal, C. Massot, G. Bedoya, and A. Caplier. Eyes segmentation applied to gaze direction and vigilance estimation. *3rd International Conference on Advances in Pattern Recognition (ICAPR), Bath, Royaume-Unis*, 2005.

Bibliographie

[ABS98] G.J. Andersen, M.L. Braunstein, and A. Saidpour. The perception of depth and slant from texture in three-dimensional scenes. *Perception*, 27(9) :1087–1106, 1998.

[All99] D. Alleysson. *Le Traitement Chromatique dans la Rétine : un Modèle de Base pour la perception humaine des couleurs*. Thèse de doctorat, laboratoire de Traitement des Images et de Représentation des Formes (TIRF), Grenoble, France, 1999.

[Alo88] J. Aloimonos. Shape from texture. *Biological Cybernetics*, 58 :345–360, 1988.

[AM04] W.J. Adams and P. Mamassian. Bayesian combination of ambiguous shape cues. *Journal of Vision*, 4 :921–929, 2004.

[AR92] J. Atick and A. Redlick. What does the retina know about natural scenes ? *Neural Computation*, 4(2) :196–210, 1992.

[BA88] J.R. Bergen and E.H. Adelson. Early vision and texture perception. *Nature*, 333 :363364, 1988.

[BBS93] A. Blake, H. Bulthoff, and D. Sheinberg. Shape from texture : Ideal observers and human psychophysics. *Vision Research*, 33 :1723–1737, 1993.

[Bea94] W.H. Beaudot. *The neural information in the vertebra retina : a melting pot of ideas for artificial vision*. Thèse de doctorat, laboratoire de Traitement des Images et de Représentation des Formes (TIRF), Grenoble, France, 1994.

[BI87] P. Buser and M. Imbert. *Vision -. Neurophysiologie fonctionnelle IV*. Hermann Paris, 1987.

[BJ83] J.R. Bergen and B. Julesz. Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man, and Cybernetics*, 13 :857863, 1983.

[BM90] A. Blake and C. Marinos. Shape from texture : Estimation, isotropy and moments. *Artificial Intelligence*, 45 :323–380, 1990.

[BNB04] D. Boukerroui, J.A. Noble, and M. Brady. On the choice of band-pass quadrature filters. *Journal of Mathematical Imaging and Vision*, 21(1) :53–80, 2004.

[BR95] M.J. Black and R. Rosenholtz. Robust estimation of multiple surface shapes from occluded textures. *International Symposium on Computer Vision, Miami, FL*, pages 485–490, 1995.

[Bra68] M.L. Braunstein. Motion and texture as sources of slant information. *Journal of Experimental Psychology*, 78 :247–253, 1968.

[Bra97] D.H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10 :433–436, 1997.

[Bro66] P. Brodatz. *Textures : A Photographic Album for Artists and Designers*. Dover, New York, 1966.

[BS90] L. G. Brown and H. Shvaytser. Surface orientation from projective foreshortening of isotropic texture autocorrelation. *IEEE Trans. PAMI*, 12 :584–588, 1990.

[Bul98] J. Bullier. *La vision : aspects perceptifs et cognitifs*. Boucart, Hennaff & Belin (eds),Edition SOLAL Neuropsychologie, 1998.

[CDM+05] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, and N.C. Rust. Do we know what the early visual system does ? *The Journal of Neuroscience*, 25(46) :1057710597, 2005.

[Cha03] A. Chauvin. *Perception des scènes naturelles : étude et simulation du rôle de l'amplitude, de la phase et de la saillance dans la catégorisation et l'exploration des scènes naturelles*. Thèse de doctorat, laboratoire des Images et des Signaux (LIS), Grenoble, France, 2003.

[CM84] J.E. Cutting and R.T. Millard. Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology*, 113(2) :198–216, 1984.

[CM02] M. Clerc and S. Mallat. The texture gradient equation for recovering shape from texture. *IEEE Trans. PAMI*, 24(4) :536–549, 2002.

[Cru97] C. Avilez Cruz. *Analyse de texture par statistique d'ordre supérieur : caractérisation et performances*. Thèse de doctorat, laboratoire de Traitement des Images et de Représentation des Formes (TIRF), Grenoble, France, 1997.

[CS88] E.L. Crow and K. Shimizu. *Lognormal Distributions : Theory and Application*. Dekker, New York., 1988.

[CTB+99] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik. Blobworld : A system for region-based image indexing and retrieval. *Proc. of the Third International Conference on Visual Information and Information Systems*, pages 509–516, 1999.

[DAOF95] G.C. DeAngelis, A. Anzai, I. Ohzawa, and R.D. Freeman. Receptive field structure in the visual cortex : does selective stimulation induce plasticity ? *Proc. Natl. Acad. Sci. USA*, 92 :9682–9686, 1995.

[DAT82] R.L. DeValois, D.G. Albrecht, and L.G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22 :545–559, 1982.

[Dau80] J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20 :847–856, 1980.

[DeV91] R.L. DeValois. *Orientation and spatial frequency selectivity : Properties and modular organization*. In A. Valberg & B.B. Lee (Eds), From pigments to perception,New York : Plenum, 1991.

[DOF95] C.G. DeAngelis, I. Ohzawa, and R.D. Freeman. Receptive-field dynamics in the central visual pathways. *Trends in Neuroscience*, 18 :451–458, 1995.

[Dur05] B. Durette. *Traitements visuels bio-mimétiques pour la suppléance perceptive.* Rapport de master, Laboratoire des Images et des Signaux (LIS), Grenoble, France, 2005.

[eLL76] R. Bajcsy et L. Lieberman. Texture gradient as a depth cue. *CGIP*, 5 :52–67, 1976.

[ES99] U.T. Eysel and G. Schweigart. Increased receptive field size in the surround of chronic lesions in the adult cat visual cortex. *Cerabral Cortex*, 9 :101–109, 1999.

[Fie87] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4 :2379–2394, 1987.

[FM64] H.R. Flock and A. Moscatelli. Variables of surface texture and accuracy of space perceptions. *Perceptual and Motor Skills*, 19 :327–334, 1964.

[Gar92] J. Garding. Shape from texture for smoothed curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2 :327–350, 1992.

[Gar93] J. Garding. Shape from texture and contour by weak isotropy. *Journal of Artificial Intelligence*, 64 :243–297, 1993.

[GCP+04] N. Guyader, A. Chauvin, C. Peyrina, J. Hérault, and C. Marendaz. Image phase or amplitude ? rapid scene categorization is an amplitude-based process. *C. R. Biologies*, 327 :313318, 2004.

[GDE99] A. Guerin-Dugue and M. Elghadi. Shape from texture by local frequencies estimation. *SCIA, Kangerlussuaq (Greenland)*, pages 533–544, 1999.

[GDO00] A. Guérin-Dugué and A. Oliva. Classification of scene photographs from local orientations features. *Pattern Recognition Letter*, 21 :11351140, 2000.

[GEN95] J.L. Gallant, D.C. Van Essen, and H.C. Nothdurft. Two-dimensional and three-dimensional texture processing in visual cortex of the macaque monkey. *Early Vision and Beyond, eds. Papathomas, Chubb, Gorea and Kowler, MIT Press*, pages 89–98, 1995.

[GH03] T. Gautama and M.M. Van Hulle. *Modeling motion processing in macaque area mt/v5 : From single cells to population codes*, pages 282–305. In G.T. Buracas, O. Ruksenas, Geoffry M. Boynton and T. Albright, editors, IOS Press, NATO Science Series, 2003.

[Gib50a] J.J. Gibson. *The perception of the visual world.* Boston, Houghton Mifflin, 1950.

[Gib50b] J.J. Gibson. The perception of visual surfaces. *American Journal of Psychology*, 63 :367–384, 1950.

[Gib79] J.J. Gibson. *The ecological approach to visual perception.* Boston : Houghton Mifflin Company, 1979.

[GK95] G.H. Granlund and H. Knutsson. *Signal Processing for Computer Vision.* Kluwer Academic Publishers, Norwell, MA, USA, 1995.

[GL96] J. Garding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2) :163–191, 1996.

[H96́] 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.

[H99́] J. Hérault. *Retine et cortex visuel : formalisation et application au traitement des images*. Cerveaux et Machines, Vincent Bloch Ed. Hermes, Paris., 1999.

[H01́] J. Hérault. *De la rétine biologique aux circuits neuromorphiques*. Traité IC2, Les Systemes de Vision, J-M Jolion ed. Hermès, Paris., 2001.

[H05́] J. Hérault. *Neural networks for vision*. World Scientific Corporation, Londres, Singapour, In press., 2005.

[Har79] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67 :786–804, 1979.

[Hee93] D.J. Heeger. Modeling simple-cell direction selectivity with normalized,half-squared, linear operators. *Journal of Neurophysiology*, 70 :1885–1898, 1993.

[HH99] J.M. Henderson and A. Hollingworth. High-level scene perception. *Annual Review of Psychology*, 50 :243271, 1999.

[HLC98] W.S. Hwang, C.S. Lu, and P.C. Chung. Shape from texture estimation of planar surface orientation throught the ridge surfaces of continuous wavelets transform. *IEEE Trans. in Image Processing*, 7(5) :773–780, 1998.

[HW62] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160 :106–154, 1962.

[HW68] D.H. Hubel and T.N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195 :215–243, 1968.

[HW74] D.H. Hubel and T.N. Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Computational Neurology*, 158 :267–294, 1974.

[JCP93] E.B. Johnston, B.G. Cumming, and A.J. Parker. Integration of depth modules : Stereo and texture. *Vision Research*, 33 :813–882, 1993.

[JP87] J. Jones and L. Palmer. An evaluation of the two-dimensional gabor filter model of simple cell receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58 :1233–1258, 1987.

[Jul81] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290 :91–97, 1981.

[Kni98a] D.C. Knill. Discriminating surface slant from texture : Comparing human and ideal observers. *Vision Research*, 38(11) :1683–1711, 1998.

[Kni98b] D.C. Knill. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, 38(17) :2635–2656, 1998.

[Kni98c] D.C. Knill. Surface orientation from texture : Ideal observers, generic observers and the information content of texture cues. *Vision Research*, 38(11) :1655–1682, 1998.

[Kni01] D.C. Knill. Contour into texture : The information content of surface contours and texture flow. *Journal of the Optical Society A*, 18(1) :12–36, 2001.

[Kovns] P. Kovesi. Surface normals to surfaces via shapelets. *Proc. Australia-Japan Advanced Workshop on Computer Vision, Adelaide*, 2003,URL : http ://www.csse.uwa.edu.au/ pk/Research/MatlabFns/.

[KSJ91] E.R. Kandel, J.H. Schwartz, and T.M. Jessell. *Principles of Neural Science*. Third ed., New York : Elsevier Science Publishing Co., 1991.

[KWG94] 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.

[KZB04] M. L. Kherfi, D. Ziou, and A. Bernardi. Image retrieval from the world wide web : Issues, techniques, and systems. *ACM Comput. Surv.*, 36(1) :35–67, 2004.

[LBP05] S. Lelandais, L. Boutté, and J. Plantier. Shape from texture : Local scales and vanishing line computation to improve results for macrotextures. *International Journal of Image Graphics*, 5(2) :329–350, 2005.

[Leb04] H. Leborgne. *Analyse de scènes naturelles par composantes indépendantes*. Thèse de doctorat, laboratoire des Images et des Signaux (LIS), Grenoble, France, 2004.

[LK05] A. Loh and P. Kovesi. Shape from texture without estimating transformations. Technical report, UWA-CSSE-05-001, July 2005.

[LSA01] E. Limpert, W.A. Stahel, and M. Abbt. Lognormal distributions across the sciences : keys and clues. *Bioscience*, 51(5) :341–352, 2001.

[LZ00] A. Li and Q. Zaidi. Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision research*, 40 :217–242, 2000.

[LZ01a] A. Li and Q. Zaidi. Erratum to information limitations in perception of shape from texture. *Vision Research*, 41(22) :29272942, 2001.

[LZ01b] A. Li and Q. Zaidi. Information limitations in perception of shape from texture. *Vision Research*, 41(12) :1519–1533, 2001.

[LZ01c] A. Li and Q. Zaidi. Veridically of three-dimensional shape perception predicted from amplitude spectra of natural textures. *Journal of Optical Society of America A*, 18(10) :2430–2447, 2001.

[LZ03] A. Li and Q. Zaidi. Observer strategies in perception of 3-d shape from isotropic textures : developable surfaces. *Vision Research*, 43(26) :2741–2758, 2003.

[LZ04] A. Li and Q. Zaidi. Three-dimensional shape from non-homogeneous textures : carved and stretched surfaces. *Journal of Vision*, 4(10(3)) :860–878, 2004.

[Mar80] S. Marcelja. Mathematical description of the response of simple cortical cells. *Journal of Optical Society of America, A*, 70(11) :1297–1300, 1980.

[MB99] M. Meister and M. J. Berry. The neural code of the retina. *Neuron*, 22 :435–450, 1999.

[McI96] J.T. McIlwain. *An Introduction to the Biology of Vision*. Cambridge, UK : Cambridge University Press, 1996.

[MMH91] J. Ross M.J. Morgan and A. Hayes. he relative importance of local phase and local amplitude in patchwise image reconstruction. *Biol. Cybern.*, 65 :113119, 1991.

[MR97] J. Malik and R. Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23(2) :149–168, 1997.

[OML03] I. Oruc, L.T. Maloney, and M.S. Landy. Weighted linear cue combination with possibly correlated error. *Vision Research*, 43 :2451–2468, 2003.

[OPM02] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. PAMI*, 24(7) :971–987, 2002.

[OT01] A. Oliva and A. Torralba. Modeling the shape of the scene : A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3) :145–175, 2001.

[OTGDH99] A. Oliva, A. Torralba, A. Guerin-Dugue, and J. Herault. Global semantic classification using power spectrum templates. *Proc. of The Challenge of Image Retrieval. Electronic Workshops in Computing series, Springer-Verlag*, Newcastle, 1999.

[Pel97] D.G. Pelli. The video toolbox software for visual psychophysics : Transforming numbers into movies. *Spatial Vision*, 10(4), 1997.

[Per03] L. Perrinet. *Comment déchiffrer le code impulsionnel de la Vision ? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide.* Thèse de doctorat, Université Paul Sabatier, Toulouse, France, 2003.

[PG92] M.W. Pettet and C.D. Gilbert. Dynamic changes in receptive-field size in cat primary visual cortex. *Proc. Natl. Acad. Sci. USA*, 89 :8366–8370, 1992.

[PK02] N. Prins and F.A.A. Kingdom. Orientation- and frequency-modulated textures at low depths of modulation are processed by off-orientation and off-frequency texture mechanisms. *Vision Research*, 42(6) :705–713, 2002.

[Pot75] M.C Potter. Meaning in visual search. *Journal of Experimental Psychology*, 2 :509–522, 1975.

[RH01] E. Ribeiro and E.R. Hancock. Shape from periodic texture using the eigen vectors of local affine distortion. *IEEE Trans. PAMI*, 23(12) :1459–1465, 2001.

[Rin02] D.L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88(1) :455–463, 2002.

[RM94] R. Rosenholtz and J. Malik. An ideal observer model for shape from texture. *Investigative Opthalmology and Visual Science, Supplemental Issue*, 35(4) :1668, 1994.

[RM97] R. Rosenholtz and J. Malik. Surface orientation from texture : isotropy or homogeneity (or both) ? *Vision Research*, 37(16) :2283–2293, 1997.

[RM04] L.W. Renninger and J. Malik. When is scene identification just texture recognition ? *Vision Research*, 44 :23012311, 2004.

[RWW04] P. Rosas, F.A. Wichmann, and J. Wagemans. Some observations on the effects of slant and texture type on slant-from-texture. *Vision Research*, 44(13) :1511–1535, 2004.

[SB95a] B.J. Super and A.C. Bovik. Planar surface orientation from texture spatial frequencies. *Pattern Recognition*, 28(5) :728–743, 1995.

[SB95b] B.J. Super and A.C. Bovik. Shape from texture using local spectral moments. *IEEE Trans. PAMI*, 17(4) :333–343, 1995.

[SBft] J.A. Saunders and B.T. Backus. Perception of surface slant from oriented textures. *Journal of Vision*, 2006 (revised draft).

[Sch80] E.L. Schwartz. Computational anatomy and functional architecture of striate cortex : A spatial mapping approach to perceptual coding. *Vision Research*, 20 :645–670, 1980.

[SF95] K. Sakai and L.H. Finkel. Characterisation of spatial frequency in the perception of shape from texture. *J. Opt. Soc. Am.*, A 12 :1208–1224, 1995.

[SF97] K. Sakai and H. Finkel. Spatial-frequency analysis in the perception of perspective depth. *Network : Computation in Neural Systems*, 8(3) :335–352, 1997.

[SHS+97] D. Shoham, M. Hubener, S. Schulze, A. Grinvald, and T. Bonhoeffer. Spatio-temporal frequency domains and their relations to cytochrome oxydase staining in cat visual cortex. *Nature*, 385 :529–533, 1997.

[Skl78] J. Sklansky. Image segmentation and feature extraction. *IEEE Transactions on Systems, Man, and Cybernetics*, 8 :237–247, 1978.

[SKS01] P.R. Schrater, D.C. Knill, and E.P. Simoncelli. Perceiving visual expansion without optic flow. *Nature*, 410 :816–819, 2001.

[SM00] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :888–905, 2000.

[SO94] P.G. Schyns and A. Oliva. From blobs to boundary edges : evidence for time and spatial scale dependent scene recognition. *Psychological Science*, 5 :195200, 1994.

[STAL02] M. E. Sereno, T. Trinath, M. Augath, and N. K. Logothetis. Three-dimensional shape representation in monkey cortex. *Neuron*, 33 :635–652, 2002.

[Ste81] K.A. Stevens. The information content of texture gradients. *Biological Cybernetics*, 42 :95–105, 1981.

[Ste84] K.A. Stevens. On gradients and texture "gradients". *Journal of Experimental Psychology : General*, 113 :217–220, 1984.

[STM96] D. Fize S.J. Thorpe and C. Marlot. Speed of processing in the human visual system. *Nature*, 381 :520522, 1996.

[Sto93] J.V. Stone. Shape from local and global analysis of texture. *Philosophical Transaction of the Royal Society London, Series B*, pages 53–65, 1993.

[SvdM02] L. Shams and C. von der Malsburg. The role of complex cells in object recognition. *Vision*, 42(22) :2547–2554, 2002.

[SW90] L. Spillmann and J.S. Werner. *Visual Perception : The Neurophysiological Foundations*. Academic Press, Inc., 1990.

[TA87] J.T. Todd and R.A. Akerstrom. Perception of three dimensional form from patterns of optical texture. *Journal of Experimental Psychology : Human Perception and Performance*, 13(2) :242–255, 1987.

[TGB91] M.R. Turner, G.L. Gerstein, and R. Bajcsy. Underestimation of visual texture slant by human observers : a model. *Biological Cybernetics*, 65 :215–226, 1991.

[TH99] A.B. 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(2), 1999.

[TJ98] M. Tuceryan and A.K. Jain. Texture analysis. *The Handbook of Pattern Recognition and Computer Vision, C. H. Chen, L. F. Pau, P. S. P. Wang (eds.)*, pages 207–248, 1998.

[TMY78] H. Tamura, S. Mori, and Y. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, 8 :460–473, 1978.

[TO02a] J.T. Todd and A.H.J. Oomes. Generic and non-generic conditions for the perception of surface shape from texture. *Vision Research*, 42 :837–850, 2002.

[TO02b] A. Torralba and A. Oliva. Depth estimation from image structure. *IEEE Trans. PAMI*, 24 :1226–1238, 2002.

[TOKK04] J.T. Todd, A.H.J. Oomes, J.J. Koenderink, and A.M.L. Kappers. Perception of doubly curved surfaces from anisotropic textures. *Psychological Science*, 15 :40–46, 2004.

[Tor03] A. Torralba. Modeling global scene factors in attention. *Journal of Optical Society of America A*, 20(7) :1407–1418, 2003.

[TP00] A. Turiel and N. Parga. The multi-fractal structure of contrast changes in natural images : from sharp edges to texture. *Neural Computation*, 12 :763–793, 2000.

[TSNT02] K.I. Tsutsui, H. Sakata, T. Naganuma, and M. Taira. Neural correlates for perception of 3d surface orientation from texture gradient. *Science*, 298(5592) :409–412, 2002.

[TSSV82] R.B. Tootell, M.S. Silverman, E. Switkes, and R.L. De Valois. Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575) :902–904, 1982.

[TTD05] J.T. Todd, L. Thaler, and T. Dijkstra. The effects of field of view on the perception of 3d slant from texture. *Vision Research*, 45 :1501–1517, 2005.

[TYST01] K.I. Tsutsui, M. Yara, H. Sakata, and M. Taira. Integration of perspective and disparity cues in surface-orientation-selective neurons of area cip. *The Journal of Neurophysiology*, 86(6) :2856–2867, 2001.

[VV90] R.L. De Valois and K. De Valois. *Spatial Vision*. Oxford University Press, 1990.

[Wal01] G. Wallis. Linear models of simple cells : Correspondence to real cell responses and space spanning properties. *Spatial Vision*, 14(3,4) :237–260, 2001.

[WH1a] F.A. Wichmann and N.J. Hill. The psychometric function i : fitting, sampling and goodness-of-fit. *Perception & Psychophysics*, 63 :1293–1313, 2001a.

[WH1b] F.A. Wichmann and N.J. Hill. The psychometric function ii : Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63 :1314–1329, 2001b.

[Wit81] A.P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17 :17–45, 1981.

[WM03] P.A. Warren and P. Mamassian. The dependence of slant perception on texture orientation statistics. *Journal of Vision*, 3(abstract no.847), 2003.

[WMeda] P.A. Warren and P. Mamassian. A bayesian model of human slant recovery under perspective projection of orientation information. *Draft version*, 2004 (to be updated).

[WMedb] P.A. Warren and P. Mamassian. Human slant estimation from texture orientation statistics : a test of the isotropy assumption. *Draft version*, 2004 (to be updated).

[WSG02] F.A. Wichmann, L.T. Sharpe, and K.R. Gegenfurtner. The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 28(3) :509–520, 2002.

Texture and 3D Perception in Natural Scenes : Biologically Inspired Models and Psychophysical Experiments.

In this work we are interested in the analysis and the extraction of the 3D information (orientation and shape) contained in natural scenes and homogeneous textures. For this, we adopt a multidisciplinary approach of the modeling of the visual system.

We first present psychophysical experiments aiming at evaluating the relative contribution of the frequency variation and of the linear perspective cues involved in 3D perception. To do so we have created purposely designed stimuli representing homogeneous textures made of Gabor patches displayed on a planar surface. The plane is viewed under perspective projection with particular slant and tilt angles. The frequency and the orientation of each Gabor patch are set according to the local frequency gradient and the local linear perspective defined by the projection. We synthesise textures presenting a frequency variation alone or an orientation variation alone or both kind of variations (in combination or in conflict). For each texture, a tilt and a slant discrimination task are performed. The frequency variation cue appeared to dominate over the linear perspective cue for slant estimation. However both cues are involved in the tilt estimation. These results validate the use of our stimuli for 3D perception study and the decomposition of the texture cue into elementary components.

Based on this approach, we present a biologically plausible model of the frequency variation analysis in the cortical area V1. We model the complex cells responses with log-normal filters which present different theoretical and practical advantages against the classical Gabor filters. The algorithm is composed of a pre-treatment stage corresponding to a retinal filtering allowing to keep only the texture information and of a decomposition of the image into local patches similarly to the cortical cells receptive fields. A robust technique aiming at estimating the local mean frequency, independently of the orientation information, and corresponding to a simple combination of the whole set of filters is applied to every patch. The measure of the local frequency variation between each patch allows to estimate the tilt and slant angles of the studied surface and its shape. The method is evaluated on different images and textures databases. It appears to be comparable in precision with the best known techniques and can be applied to irregular textures with a lower computational cost.
key words : 3D perception, natural scenes, texture, frequency, linear perspective

Laboratoire des Images et des Signaux
46 Avenue Felix Viallet
38031 Grenoble Cedex

Texture et Perception 3D dans les Scènes Naturelles : Modèles d'Inspiration Biologique et Expérimentations Psychophysiques.

Dans ce travail nous nous intéressons à l'analyse et l'extraction de l'information 3D (orientation et forme) contenue dans les images de scènes naturelles et des textures homogènes. Pour cela nous adoptons une approche pluridisciplinaire de la modélisation du système visuel.

Nous présentons d'abord des expérimentations psychophysiques où nous avons cherché à évaluer la contribution relative des indices de variation de fréquence et de perspective linéaire pour la perception 3D. Pour cela nous avons créé des stimuli spécifiques représentant des textures homogènes composées de masques de Gabor disposés sur une surface plane. Le plan est vu en projection perspective suivant une inclinaison (slant) et une orientation (tilt) particulière. La fréquence et l'orientation de chaque masque de Gabor sont déterminées en fonction du gradient de fréquence local et de la perspective linéaire locale définis par la projection. Nous synthétisons ainsi des textures présentant uniquement une variation de fréquence ou une variation d'orientation ou les deux types de variation (en combinaison ou en conflit). Pour chaque texture, une tâche de discrimination du slant et du tilt est effectuée. L'indice de variation de fréquence apparaît prépondérant dans l'estimation de l'inclinaison d'une surface par rapport à la perspective linéaire. Par contre les deux indices jouent un rôle dans l'estimation de l'orientation. Ces résultats valident l'utilisation de nos stimuli pour la perception 3D et permettent de préciser la décomposition de l'indice de texture en composantes élémentaires.

Basé sur cette approche, nous présentons un modèle biologiquement plausible d'analyse de la variation de fréquence au niveau de V1. Nous modélisons la réponse des cellules complexes par des filtres log-normaux à variables séparables présentant différents avantages théoriques et pratiques par rapport aux filtres de Gabor classiquement utilisés. L'algorithme se compose d'une étape de prétraitement composé d'un filtrage rétinien pour ne conserver que les informations de texture et d'une décomposition de l'image en un ensemble d'imagettes similairement aux champs récepteurs des cellules corticales. Une technique robuste d'estimation de la fréquence moyenne locale, indépendante de l'information d'orientation et correspondant à une combinaison simple de l'ensemble des filtres est appliquée à chaque imagette. La mesure de la variation locale de fréquence entre chaque imagette permet d'estimer le tilt et le slant de la surface étudiée ainsi que sa forme. La méthode est évaluée sur différentes bases d'images et de textures. Elle s'avère comparable en précision aux autres techniques et s'applique à des textures irrégulières avec une moindre complexité calculatoire.

mots clefs : perception 3D, scènes naturelles, texture, fréquence, perspective linéaire

Laboratoire des Images et des Signaux
46 Avenue Felix Viallet
38031 Grenoble Cedex