Model of Frequency Analysis in the Visual Cortex and the Shape from Texture Problem

Corentin Massot, Jeanny Hérault

HAL Id: hal-00342144
https://hal.science/hal-00342144

Model of Frequency Analysis in the Visual Cortex and the Shape from Texture Problem.

Corentin Massot, Jeanny Hérault

Laboratory of Images and Signals, Grenoble,France

corentinmassot@yahoo.fr

Abstract

This paper addresses the question: how does the visual cortex extract local perspective information from texture variations? Starting from a model of complex cells in visual area V1, we propose a biologically plausible algorithm for frequency analysis applied to the shape from texture problem. First, specific log-normal filters are designed in replacement of the classical Gabor filters because of their theoretical properties and of their biological plausibility. These filters are separable in frequency and orientation and they better sample the image spectrum which makes them appropriate for any pattern analysis technique. A method to estimate the local frequency in the image, which discards the need to choose the best local scale, is designed. Based on this frequency analysis model, a local decomposition of the image into patches leads to the estimation of the local frequency variation which is used to solve the problem of recovering the shape from the texture. From the analytical relation between the local frequency and the geometrical parameters, under perspective projection, it is possible to recover the orientation and the shape of the original image. The accuracy of the method is evaluated and discussed on different kind of textures, both regular and irregular, with planar and curved surfaces and also on natural scenes and psychophysical stimuli. It compares favorably to the best existing methods, with in addition, a low computational cost. The biological plausibility of the model is finally discussed.

keyword: Model of V1, frequency estimation, log-normal filters, shape from texture, biological model

1 Introduction

Nowadays, the biological mechanisms of the low-level vision are rather well identified [1] [2]. First the retina realizes a set of preprocessings which leads to the separation between shading and texture information by an adaptive local equalization and a spectral whitening [3] [2]. Then the output is projected to the primary visual area V1 where it is filtered by simple (in-phase and in quadrature) and complex (energy dependent) cells. Complex cells are the most numerous cells in the visual cortex and their responses have widely been modelized by Gabor filters responses with specific central frequencies and orientations [4] [5] [6]. In area V1, the complex cells are organized according to micro-columns of similar frequency and orientation and, approximately, divided into 6-7 central frequencies (on a log-scale) and 15-16 orientations (between 0° and 180°) [7]. Micro-columns are organized in a pin-wheel fashion within a macro-column. The surface of V1 is divided into several filters banks, each one being devoted to a visual field subregion called receptive field. Briefly, V1 processing by complex cells can be viewed as a log-polar sampling of the energy spectra of local patches by bandpass oriented filters over the whole visual field, after retinal preprocessing (namely the retinal input image).

The projection of a real image on the retinal surface induces affine deformations to the pattern structure (the texture) covering the viewed 3D surface. We will show how the visual cortex recover a 3D information from the analysis of the texture deformations from the model described before and using only the local energy spectrum.

Since the seminal work of Gibson [8], a significant number of studies have considered the use of the texture gradients as important monocular cues to estimate the orientation and the shape of the surface underlying the texture. Since then neurophysiological [9] [10] as well as psychophysical [11] [12] evidences have been found in support of the process of texture gradients in the visual system. This problem is usually reduced to the estimation of two projection parameters: the slant (inclination of the surface in depth) and the tilt (direction of slant). Since the early 90's the use of spectral analysis has led to several efficient shape

from texture algorithms. In [13], [14], [15], [16], [17], [18] the texture deformation is measured using the affine distortion of the pattern of spectral components (e.g. energy peaks, local moments, inertia). These methods are accurate but need the presence of at least two distinct energy components in the spectrum and so usually perform best in presence of at least two orientations in regular or weakly irregular texture.

Alternative techniques do not make any assumption on the spectral components. In [19], [20],[21], [22] the proposed methods are based on the local estimation of the spatial scale using different kind of filters (e.g. gaussian derivatives, wavelets). They are based on the choice of the best local scale, explicitly assuming that there is only one at the local spatial position considered, which is not always true in specific cases of multiple, occluded or transparent textures [23]. In order to handle more or less the irregularities of the texture, all these methods require heavy computation such as parabolic approximation or variance minimization techniques to finally recover the geometrical parameters of surfaces (usually planar).

In [24], the local migration of wavelets coefficients is related to the local shape parameters. In order to obtain results on irregular textures, the problem is treated as the recovery of a stationary process which imposes a strong constraint on the coefficients variations. However due to ergodicity hypothesis, each wavelet is taken at a central frequency in a relatively high spatial frequency and so does not take explicitly advantage of the whole set of available frequencies.

Finally none of these techniques make a systematic comparison with the biological visual processing, excepted for Sakai and Finkel work [16] which developed a model able to handle both frequency peak variations as well as mean frequency variations, misleading moment based methods.

In the present work, we present a model of frequency analysis by the visual cortex and applied to the shape from texture problem. It is based on the local decomposition of an homogeneous texture using local patches (similarly to the macro columns organization). The spectral analysis of the patches in a local neighborhood ensures the independence of the spatial frequency with small translations all over the surface as well as some smoothing of the local texture characteristics. Section 3 presents a new model of cortical filters used to realize the sampling of the spectrum. They are based on log-normal filters and appear theoretically to be more suitable than the classical Gabor filters, and biologically more plausible.

These filters are separable in frequency and orientation. This property makes them very appropriate to independently handle changes in frequency or in orientation, corresponding to a zoom or a rotation transformation, and so to local affine transformations within the different patches. Their expression on a log-scale frequency also induces a better coverage of the spectrum than with the use of Gabor filters. Based on the use of these new filters a method, derived from Knutsson *et al* technique [25], is introduced to estimate the local frequency in the considered patch (section 4). It prevents from the need to choose the best local scale and takes advantage of the whole set of available local frequencies. An accurate estimation is obtained by the modification of the relative frequency bandwidth, giving an empirical explanation of the cortical cells organization. Section 5 presents the projective model and the analytical relations between the local frequency and the geometrical parameters of the viewed surface. It discusses the approach by local decomposition into patches and proposes an algorithm to recover the orientation and the shape of the original image. Section 6 presents estimation results obtained on different kinds of textures, both regular and irregular, with planar and curved surfaces and also on natural scenes. They are compared to other existing techniques. Last section finally discusses the biological plausibility of the algorithm.

2   Retinal preprocessing

The analysis of textures corresponding to subregions of natural scenes leads to two problems: spatially variable illumination conditions and characteristic $1/f^2$ amplitude spectrum slope of images corresponding to an over-presence of energy in low frequencies [26].

To avoid these problems we introduce a special preprocessing stage based on the model of the human retina [3][27][28]. The retina represents the first set of filtering processes which transforms and prepares the visual input information to be analyzed by higher cortical areas such as V1. The overall effect of this filtering is the removal of shading variations and the enhancement of texture information.

The retina processes the spatial variation of illumination by a non-linear adaptation (see figure 1). In this adaptation, the gain of each photoreceptors is modulated according to the mean illumination of its neighborhood. It is realized by a compression process $y = \dfrac{x}{x + x_0}$, x_0 being proportional to the smoothed

neighborhood (Michaelis-Menten compression law). This compressed input is then low-pass filtered. The compressed input and its low-passed filtered version are then combined into a positive and a negative difference which leads to two components (the ON and OFF cells). These two components are then subtracted leading to the final filtered output. The main filter transfer function corresponds to a global high-pass filter applied to the local luminance equalization, realizing a spectral whitening of the $1/f^2$ amplitude spectrum.

As explained in [3] this filter can be considered as a good approximation of the spatial and static filtering realized by the retina and of the X parvo cellular pathway.
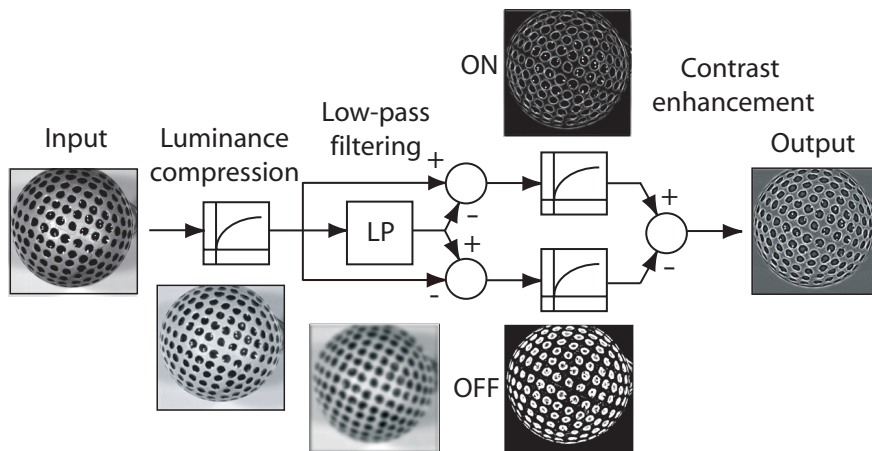

Figure 1: Retinal filter steps: luminance compression of the input image; low-pass filtering and combination with the compressed input; second compression leading to contrast enhancement; final output where the shading information has been removed and the texture information, enhanced.
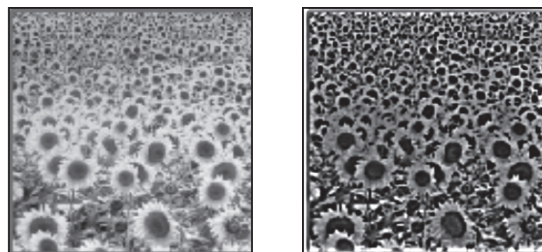

Figure 2: Left: Image of a sunflower field; right: same image after retinal preprocessing

3   Spectral sampling using Log-Normal filters

Different kind of filters are commonly used in pattern recognition technique, especially Gabor filters which present the advantages to be localized in space and frequency and so to be easily tuned to a specific frequency and orientation. However they have the drawbacks not to be separable filter and to have a continuous component. In this paper specific filters, derived from the log-normal distribution, are introduced and a method to extract the local frequency is presented.

3.1   How to choose appropriate filters

Spectral filtering (or equivalently spatial filtering) is a technique commonly used in a great number of pattern analysis problem (e.g. object recognition, face recognition, texture analysis). Facing the large number of plausible filters [6], we have to choose the most appropriate one for the specific application. A possible good starting point is to get inspiration from the visual cortex processing. As described before, it is organized in columns of orientations and frequencies, leading to the sampling of the local spatial spectrum within a given receptive field. In order to reproduce this decomposition of the image into local patches, the chosen model of filters has to share the Gabor property of being localized both in space and frequency (e.g.inside a defined spatial region). The use of the energy spectrum associated to the local patch leads to be insensitive to local spatial translations and so to the local phase and, at the same time, to keep the local the second order statistics. Another advantage is the local smoothing of these statistics leading to a more robust estimation. Gabor filters can also be easily tuned in frequency and orientation leading to a polar sampling of the spectrum. However they are not completely satisfying as models of complex cells responses (see [6]). They also present some drawbacks like not to be separable filters in orientation and frequency. A frequency modulation in the patch corresponds to a change of zoom factor. Similarly a change in orientation in the spectrum corresponds to a rotation of the pattern. These two geometrical transformations represent two free independent variables that have to be taken into account once studying visual scenes in order to obtain a robust analysis. Instead of the classical Gabor filters, other kind of filters have been developed notably by Field (log-gabor filters) [29] and Knutsson *et al* (log-normal filters)

[25] and have been successfully employed in pattern recognition techniques. In this paper we propose new filters which are a specific version of the log-normal filters, better adapted to affine transformations such as zoom and rotation.

3.2 Log-Normal filters

In this section we present new filters specifically designed for zoom and rotation transformations. They are derived from the log-normal function:

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

where $G_{i,j}$ is the transfer function of the filter, $G_i(f)$ and $G_j(\theta)$ respectively represent the frequency (radial) and the orientation (tangential) components of the filter, their shape are gaussian in log-frequency and in orientation. f_i is the central frequency, θ_j, the central orientation, σ_r, the frequency bandwidth,. n controls the orientation bandwidth similarly to a gaussian shape. A is a normalization factor such as $||G_{i,j}(f,\theta)||^2 = 1$ ($A = \frac{2^{2n}}{2\pi C_{2n}^n}\frac{1}{\sigma_r\sqrt{2\pi}}$).

These log-normal filters share the same properties of the Gabor filters: spatial localization; tuned around a central frequency and a central orientation. In addition they are designed so as to be separable in frequency and orientation (equation 1). The radial component $G_i(f)$ represents the log-normal function applied to frequency. The orientation component $G_j(\theta)$ is a cosine wave ensuring a 2π-periodic orientation bandwidth and a finite support (null value at the angular limits of the filter). It approximates a gaussian shape with an accuracy better than 0.1%.

Figure 3 presents a comparison between the Gabor and the log-normal filters which are represented in 3D, cartesian and log-polar coordinates. In cartesian representation the log-normal filter is asymmetrical and this property increases with low central frequencies. Associated to the fact that the filter gain is null at $f = 0$ whatever the bandwidth, the log-normal filter is well defined even at very low central frequencies (Equation 3.2 imposes a null continuous component). This is not the case with Gabor filters which can lead to spurious responses in very low frequencies, hence prohibiting the robust analysis of low frequency components.
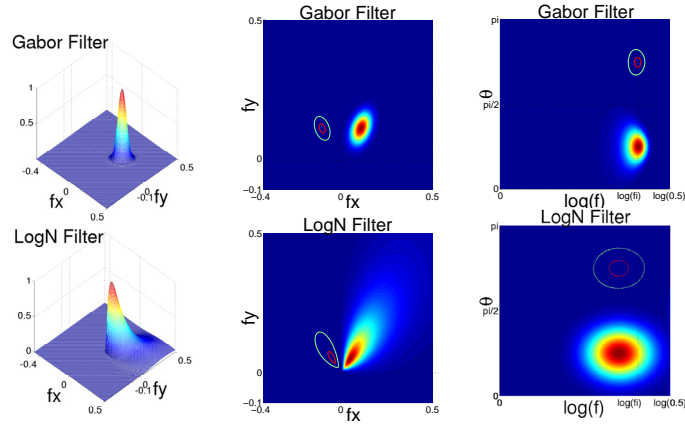
Figure 3: Comparison between Gabor filters (first row) and log-normal filters (second row); left: 3D representation; middle: filters and contours representing 50% and 90% of maximum energy in cartesian coordinates; right: same filters but represented in log-polar coordinates.

One can also observe that in both cartesian and log-polar representations log-normal filters are well adapted to a log-polar sampling of the spectrum. In log-polar coordinates log-normal filters become symmetric (with a gaussian profile) and lead to a better coverage of the spectrum than the Gabor filters which become asymmetric.
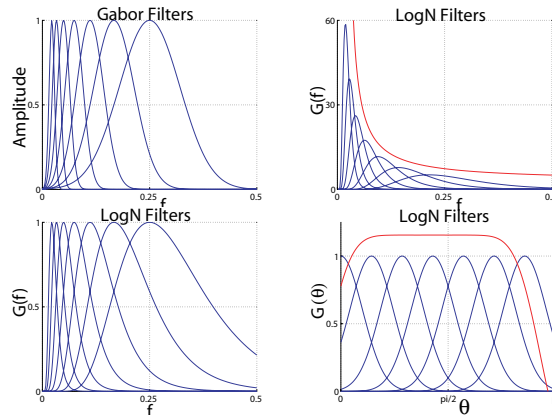

Figure 4: First column: comparison between Gabor (top row) and log-normal (bottom row) filters profiles; second column: log-normal frequency profiles with $1/f$ amplitude factor (top row) and log-normal orientation profiles (bottom row); the filters coverage is represented by their sum which follows a $1/f$ law on amplitude and is constant on orientation.

Figure 4 presents the different profiles of a bank of Gabor filters and of a bank of log-normal filters. For this latter the profiles are plotted with and without the $1/f^2$ amplitude factor so as to observe

8

the asymmetry in low frequencies. In addition, here we indicate two other specific properties of this implementation of the log-normal filters (not directly exploited in this paper).

As described before, a frequency modulation can be induced by a zoom transformation of the image and can be written as $i(x,y) \rightarrow i(\alpha x, \alpha y)$ which gives in the Fourier domain $S(f,\theta) \rightarrow \frac{1}{\alpha^4}S(\frac{f}{\alpha},\theta)$ where α is the zoom factor and S the power spectrum of the image. In the following section, the developped method is free from this $1/\alpha^4$ coefficient.

If we express the response of the filter $C_{i,j}$ on a log-polar scale with $v = ln(f)$ we get:

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

with $v = ln(f)$. It has to be noted that the $1/f^2$ term cancels through the change of variable. Applying equation 2 on $S(\frac{f}{\alpha},\theta)$ with $v_m = v + ln(\alpha)$ we get:

$$C_{ij}(\alpha) = A \int_{\theta=0}^{2\pi} G_\theta^2(\theta - \theta_j) \int_{v_m=-\infty}^{+\infty} \frac{1}{\alpha^4} S(e^{v_m},\theta) exp\left(-\frac{1}{2}\left(\frac{v_m - (v_i - ln(\alpha))}{\sigma_r}\right)^2\right) dv_m d\theta \tag{3}$$

So a zoom transformation of the image can be perfectly tracked by the change in the filter response thanks to the presence of the $1/f^2$ term and to the logarithmic frequency scale. In the case of Gabor filter, the response to a frequency modulation is still dependent of an extra frequency factor and so does not give a direct access to the zoom transformation. This property is not possible with Gabor, Log-Gabor or Knutsson's version of log-normal filters.

Figure 4 finally shows the good coverage on orientations of the filters with a constant sum of the different profiles. Similarly to the frequency variable, the expression of a rotation of the image leads to a perfect tracking of the transformation by the migration of the log-normal coefficients. A rotation of angle β only induces a modification in the orientation component of the filter:

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

9

To sample the half symmetric power spectrum, we design a log-normal filters bank with constant relative frequency and orientation bandwidths (see section 4.1 for more details). On figure 5 right, it is possible to check that this design leads to a good coverage of the spectrum in cartesian coordinates and to a more regular sampling in log-polar coordinates than a corresponding Gabor filter bank with similar bandwidth.
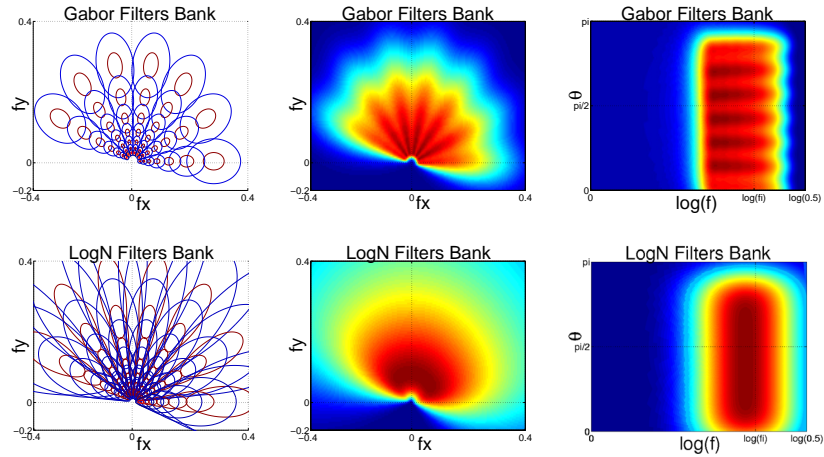

Figure 5: Comparison between Gabor filters (first row) and log-normal filters (second row); left: contours representing 50% and 90% of maximum energy;middle: bank of filters of the whole power spectrum in cartesian coordinates; right: bank of filters (50% and 90% of maximum energy) in log-polar coordinates.

Finally the log-normal filters present other algebraic properties that make them very suitable for image processing: the transfer function is C^{∞}; the real and imaginary parts are in quadrature; the function is self-similar and can be used as a mother wavelet.

According to [6], all the presented characteristics of the log-normal filters are shared with the cortical complex cells properties. The log-normal filters can then be considered as a good approximation of their spatial responses.

4   Local frequency estimation

In this section we will take advantage of the separability property of the log-normal filter so as to estimate the local frequency of the image. Let's take the response of the ith log-normal filter i:

$$G_i^2(f) = \frac{1}{f^2} exp\left(-\frac{1}{2}\left(\frac{ln(f/f_i)}{\sigma_r}\right)^2\right) \tag{5}$$

Similarly to Knutsson *et al* in [25], the ratio of two adjacent filters responses is expressed by:

$$\frac{G_{i+1}^2(f)}{G_i^2(f)} = exp\left(-\frac{1}{2\sigma_r^2}[(ln(f/f_{i+1})^2 - (ln(f/f_i))^2]\right) \tag{6}$$

$$= \left(f/\sqrt{f_i f_{i+1}}\right)^{\frac{ln(f_{i+1}/f_i)}{\sigma_r^2}} \tag{7}$$

If $\sigma_r^2 = ln(f_{i+1}/f_i)$, we finally get a simple relation between the filters responses:

$$G_{i+1}^2(f) = \frac{f}{\sqrt{f_i f_{i+1}}} G_i^2(f) \tag{8}$$

Now, in order to extract a scale information independently of the local orientations, we consider frequency band tuned filters which result from the summation over all orientations j of the filters response tuned to the same frequency i on the image spectra $S(f, \theta)$:

$$C_i = \int_f G_i^2(f) \int_\theta S(f, \theta) \sum_j G_j^2(\theta) f df d\theta \tag{9}$$

The ratio of the responses of two adjacent frequency band filters C_{i+1} and C_i gives:

$$\frac{C_{i+1}}{C_i} = \frac{1}{\sqrt{f_i f_{i+1}}} < f >_i \tag{10}$$

where $< f >_i$ represents the narrow band local mean frequency estimated at the ith frequency band. Equation 10 shows it can simply be extracted by the ratio of two adjacent filters given their two central frequencies f_i and f_{i+1}. This ratio allows also to be independent of the $1/\alpha^4$ coefficient appearing in a zoom transform of the image (see section 3.2).

11

Finally summing over all the local estimations at different frequency bands i of mean frequency $< f >_i$ we obtain the overall wide-band mean frequency $< f >$:

$$< f >= \sum_i \frac{C_i}{\sum_i C_i} < f >_i \tag{11}$$

The final estimation $< f >$ of equation 11 is based on the summation over the whole range of available central frequencies. In this way it leads to an estimation that may be more robust with respect to noise. The response C_i in equation 9 are chosen as the weighting coefficients. This choice is appropriate because they reflect the presence of the frequency $< f >_i$ in the image observed around the central frequency f_i, e.g it corresponds to the extraction of a frequency component of the signal (see [25]).

Contrary to [20] and [24], this approach does not require any restriction to one best local scale (i.e central frequency). The final estimation, using equation 11, is performed under the assumption that there is only one local frequency to estimate. In other cases, equation 10 may be used to estimate the different mean frequencies that can appear in the image at different central frequencies (e.g. case of transparency or occlusion) [23]. Finally a very simplified model of cells combination for frequency extraction is obtained which takes advantage of the whole range of available frequencies.

4.1 Non-constant relative frequency bandwidth

In order to evaluate the precision of our method we applied it on gratings where the local frequency was perfectly known (figure 6 left). These gratings have a linear increase in frequency and the whole set of gratings covers the frequency space from 0 to 0.33.

The final curves (figure 6 right) are obtained after curve fitting (simplex method) on all the estimations performed along different spatial positions on the image grating. Each local frequency estimation is performed inside a sliding window (2D patch) with a specific size (96X96 pixels). The dash curve represents the estimation curve obtained with the bank of log-normal filters described in section 3.2. It can be observed that the estimation is good in low frequencies but tends quickly to an asymptotic value.

One characteristic of the log-normal filters is that their relative frequency bandwidth does not depend on
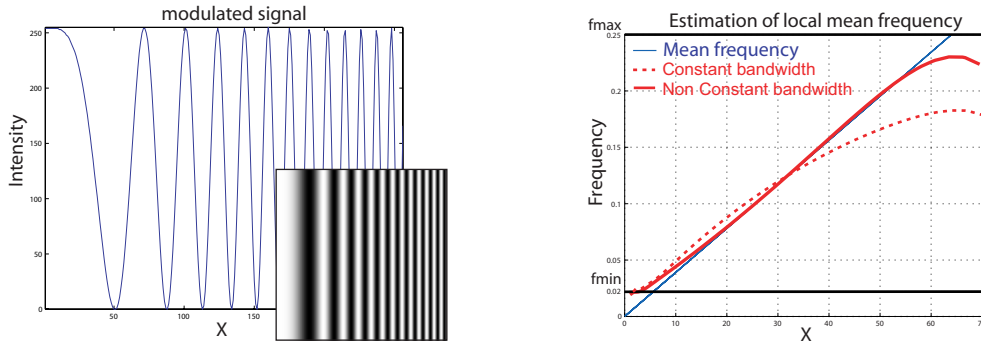
Figure 6: Estimation of the local frequency on a set of gratings; left: projected grating with the 1D signal representing the corresponding modulated signal; right: straight line: real mean frequency of the grating; dash curve: estimated mean frequency with constant relative bandwidth of the filters; solid curve: estimated mean frequency with linearly decreasing relative bandwidth of the filters.

their central frequency f_0. Indeed the mean frequency of one filter is $u_f = e^{\sigma_r^2/2} f_0$ (first order moment) and its bandwidth is $\Delta_f = \sqrt{(e^{\sigma_r^2} - 1)} e^{\sigma_r^2/2} f_0$ (second order moment). Hence the relative frequency bandwidth is given by $\Delta_f/u_f = \sqrt{(e^{\sigma_r^2} - 1)}$ which only depends on the width of the gaussian σ_r. Here, equation 8 imposes σ_r to be equal to $ln(f_{i+1}/f_i)$ in order to keep their ratio constant and equal to 1. Then the filters of the first bank described in section 3.2 were defined with a constant relative frequency bandwidth equal to 1.4 (with $f_{i+1}/f_i = 1.5$).

If we relax this constraint, the analysis of equation 8 shows that a decrease of the ratio counterbalances for the quick saturation of the estimation precision for high frequencies. In this case this ratio becomes slightly superior to 1 but allows to keep equation 9 as a valid approximation of the local mean frequency. We found that for a linear decrease of σ_r with a coefficient equal to 2, the method yields to a very accurate estimation of local frequencies (solid curve on figure 6 right) over the whole range of available frequencies (from minimum central frequency $f_{min} = 0.02$, to maximum central frequency $f_{max} = 0.25$ of our bank of filters). Figure 7 right presents estimations realized with different size of patches (128X128, 96X96, 64X64 and 3X32 pixels) and it can be observed that the accuracy remain closely the same.

This non-constant relative frequency bandwidth has been observed in the cortical cells. Data collected in the human cortex (figure 7 left from [30]) and in the macaque cortex [2] clearly shows that the relative bandwidth of the cortical filters of V1 is not constant but decreases linearly with the central frequencies.
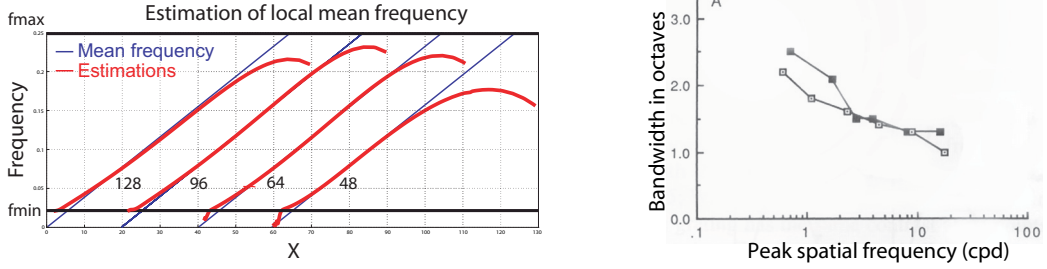
13

Figure 7: Left: the modified bank of log-normal filters with a non-constant relative frequency bandwidth; right: data from [30] collected from human cortex and presenting a linear decrease of the frequency bandwidth.

Hence the presented model of V1 analysis and extraction of the local frequency suggests an empirical justification of the particular configuration of the cortical cells, a phenomenon never explained up to now. The obtained bank of log-normal filters appears to be a very close approximation of the complex cells organization and at the same time leads to very accurate local frequency estimations.

4.2   Local analysis of the amplitude information

In order to estimate the local frequency over the whole surface, a decomposition into patches is realized (figure 8). A hamming window is also applied on each patch so as to remove the border effect in the Fourier transform. Defining the optimal constant size of the patches to study the local properties of a texture is still an open problem. In our case, $64X64$ or $96X96$ pixels wide patches are found to appropriately capture the statistical properties of $256X256$ pixels wide images. The spatial accuracy can be adapted varying the shift between patches. Generally we choose a shift of 8 or 4 pixels corresponding to a decomposition of the image into $21X21$ or $42X42$ patches.

This local decomposition into patches can be viewed as a model of the sampling operated by each cortical cells where the covered spatial region is called the receptive field of the cell [2]. The local mean frequency extraction described previously is performed on the amplitude spectrum of the local patch. Hence through this decomposition, it is possible to recover the information on spatial location without the use of any phase information.
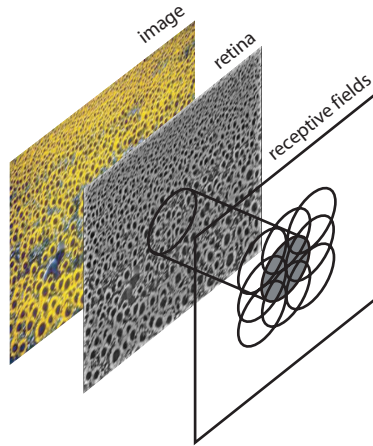
14

Figure 8: Different steps of the pretreatment: initial image of a sunflowers field; retinal preprocessing of the image; local decomposition into patches of the whole image similarly to cortical cells analysis.

4.3 Cortical normalization

The frequency band tuned filters, corresponding to the C_i coefficients in equation 10 and 11, are obtained after the summation over all the filters tuned to the same central frequency and tuned to different central orientation. In order to compensate local irregularities in the texture, a local normalization is applied. More precisely from one spatial location to another, variations in energy may occur due to local texture non stationnarities (i.e size, position or density of the texture patterns). So as to reduce these variations the filters tuned to the same orientation are normalized according to the sum of their responses over all the central frequencies. Filter response G_i^2 can be rewritten as follow:

$$G_{i,norm}^2(f_k, \theta_k) = \frac{G_i^2(f_k, \theta_k)}{\sum_j G_j^2(f_k, \theta_k) + \varepsilon} \tag{12}$$

Constant ε avoids the enhancement of the responses in case of global low energy responses for the considered orientation band. Introducing $G_{i,norm}^2$ instead of G_i^2 leaves equation 10 and 11 unchanged.

This normalization enhances the presence of characteristic spectral components at the different orientations. Such cortical normalization processes have been severally reported and different models have been proposed [31]. This computation of frequency band tuned filters may represent a simple way to separate the frequency from the orientation information directly in cortical area V1. The frequency variations can be analyzed independently to the orientation variations. This has to be related with the work of Li and

zaidi [11] which advocate for the existence of two separate mechanisms involved in the analysis of the frequency and the orientation for 3D shape perception.

5   Local surface representation and final shape recovery

This section describes how to recover the geometrical parameters (tilt and slant angles) information from spatial frequency variation. Different strategies have been adopted which could be applied to our frequency extraction method. In [16] and [17], the method consists in following paths of maximum frequency variations starting from a reference point supposed to be frontoparallel; however such a point does not always exist, for example in the case of planar surfaces. In [19] the method is based on an exhaustive research of tilt and slant angles and on the minimization of the parameters deviation in a backprojected version of the surface. This method depends on the discretization of the parameters in the research step and requires repetitive computation. In [21] and [22] methods are based on robust curve fitting of the local scales. The tilt and slant angles can be directly extracted from the obtained parabolic curve.

Here we present another method to recover the geometrical parameters of the surfaces. First different geometrical relationships are established: the relationship between the surface (viewed) local frequency and the image local frequency under perspective projection and the expressions of the tilt and slant angles. Secondly, the estimation of the orientation of local subregions made of several local patches is performed and allows to solve the shape from texture problem without intensive computation.

5.1   Geometrical relationships

Figure 9 presents the coordinate system of the perspective model of a plane surface. (x_w, y_w, z_w) represents the world coordinates, (x_s, y_s), the surface coordinates and (x_i, y_i), the image coordinates. The z_w axis corresponds to the intersection between the center of projection, the origin of the image coordinates and the origin of the surface coordinates. d (resp. $zw0$) is the coordinate of the image (resp. the surface) on the z_w axis. Let's $dzw0$ be the distance between the image and the surface. τ represents the tilt which is the angle between x_i and the projection of the normal z_s on the image plane. σ is the slant which is the

16

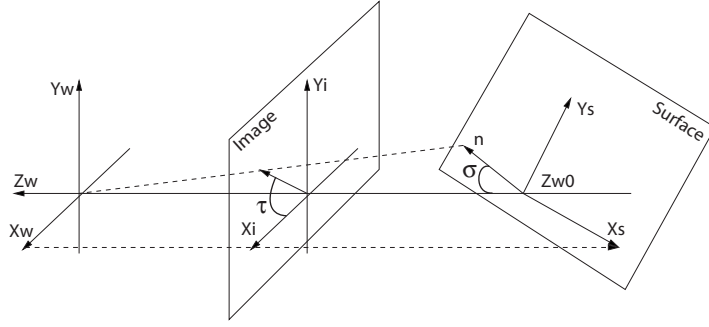angle between axis z_w and the normal of the surface at $zw0$ and its values fall between 0 and $\pi/2$

Figure 9: Model of perspective projection

The relationship between the coordinates (x_s, y_s) of the surface and the coordinates (x_i, y_i) of the image under perspective projection can be expressed as (see [19]):

$$
\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{\begin{bmatrix} cos(\sigma) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} cos(\tau) & sin(\tau) \\ -sin(\tau) & cos(\tau) \end{bmatrix}}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{A}{a_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix}
\tag{13}
$$

where $a_i = \dfrac{-sin(\sigma)sin(\tau)x_i + cos(\tau)sin(\sigma)y_i + dcos(\sigma)}{d + dzw0}$ which can be seen as a zoom factor depending on the spatial position (x_i, y_i).

Let's take a spatial region L_i centered on the position x_i, the corresponding local Fourier transform I_{Li} is expressed by:

$$
I_{Li}(f_i, x_i) = \int_u i_i(u) w_i(u - x_i) e^{-j2\pi(u-x_i)^t f_i} du
\tag{14}
$$

where i_i is the image and w_i, a spatial window in this image. Let's take an equivalent spatial window w_s in the surface i_s, assuming that $v = T^{-1}u$ and $x_s = T^{-1}x_i$, regions of the image and of the surface are linked by: $i_s(v)w_s(v - x_s) = i_i(u)w_i(u - x_i)$.

17

Using the inverse Fourier transform associated to $i_s(v)w_s(v - x_s)$ defined in equation 14, we obtain:

$$I_{Li}(f_i, x_i) = \int_u \left(\int_{f_s} I_{Ls}(f_s) e^{j2\pi(v-x_s)^t f_s} \right) e^{-j2\pi(u-x_i)^t f_i} du df_s \tag{15}$$

$$= \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi((v-x_s)^t f_s - (u-x_i)^t f_i)} du df_s \tag{16}$$

Using equation 13, the first order approximation of $(v - x_s)^t$ is given by:

$$(v - x_s)^t = \frac{1}{a_i} \left(I - \frac{\nabla a_i x_i^t}{a_i} \right) A^t s(u - x_i)^t = R^t (u - x_i)^t \tag{17}$$

Replacing in equation 16, we obtain a relation between I_{Li} and I_{Ls}:

$$I_{Li}(f_i, x_i) = \int_{f_s} I_{Ls}(f_s) \int_u e^{j2\pi(u-x_i)^t (R^t(x_i)f_s - f_i)} du df_s \tag{18}$$

$$= \int_{f_s} I_{Ls}(f_s) \delta(R^t(x_i)f_s - f_i) df_s = \frac{1}{|det(R)|} I_{Ls}(R^{-t}(x_i)f_i) \tag{19}$$

Finally we obtain the relation between f_i and f_s (for δ not null):

$$f_i = R^t(x_i)f_s \approx \frac{1}{a_i} \left(I - \frac{\nabla a_i x_i^t}{a_i} \right) A^t f_s \tag{20}$$

So as to link the frequency variation with the shape of the surface, an homogeneity hypothesis on the texture is required, as described before. Then a frequency variation on hte image can be directly related to the inclination of the surface before projection.

Using equation 20, the expression of the local frequency variation is:

$$df_i = -\frac{1}{ai} \left[\nabla^t a_i dx_i + \nabla a_i dx_i^t \right] f_i \tag{21}$$

The frequency of the image f_i can be expressed in polar coordinates by $f_i = v_i[cos(\varphi_i) \; sin(\varphi_i)]^t$. Equation

21 becomes:

$$df_i = dv_i[cos(\varphi_i)\ sin(\varphi_i)]^t + v_i[-sin(\varphi_i)\ cos(\varphi_i)]^t d\varphi_i \tag{22}$$

$$= -\frac{\nabla^t a_i dx}{a_i} v_i[cos(\varphi_i)\ sin(\varphi_i)]^t - \frac{1}{a_i}\nabla a_i dx_i^t v_i[cos(\varphi_i)\ sin(\varphi_i)]^t \tag{23}$$

Considering the gradient df_i in the direction φ_i, multiplying by $[cos(\varphi_i)\ sin(\varphi_i)]^t$, equation 23 becomes:

$$dv_i = -\frac{\nabla^t a_i dx}{a_i} v_i - \frac{1}{a_i}[cos(\varphi_i)\ sin(\varphi_i)]\nabla a_i dx_i^t[cos(\varphi_i)\ sin(\varphi_i)]^t v_i \tag{24}$$

Summing over φ_i (on $[0, 2\pi]$) we obtain:

$$\frac{dv_i}{v_i} = -\frac{3}{2}\frac{\nabla a_i^t dx_i}{a_i} \tag{25}$$

and then

$$dln(v_i) = -\frac{3}{2}\frac{sin(\sigma)\ [-sin(\tau)cos(\tau)]^t\ [dx_i dy_i]}{a_i} \tag{26}$$

Finally for an homogeneous textured region, the tilt angle corresponds to the direction of the frequency gradient (ratio between the components of v_i according to axis x_i and y_i) and the slant angle is proportional to the norm of the log-frequency gradient $|dln(v_i)|$ in the tilt direction:

$$tan(\sigma) = \frac{|dln(v_i)|f}{\frac{3}{2}|dx| - |dln(v_i)|(-sin(\tau)x_i + cos(\tau)y_i)} \tag{27}$$

with $|dx|$ corresponding to the unitary spatial shift.

5.2 Shape recovery

The obtained geometrical relationships allow to compute the tilt and slant angles at any local position providing the computation of the associated frequency gradient. The estimation method is based on local

extraction of mean frequency gradients in subregions of the final frequency estimation surface (i.e averaging over subsets of patches) (figure 10). We suppose that the surface covered by the subregion is planar that is the slant and tilt angles are supposed constant all over the subregion. The tilt and slant angles are computed using 27 and this operation is carried out on all the subregions over the whole image. The size of the subregions can be adjusted as a smoothing parameter while integrating over the local estimations. In case of a planar surface, the mean tilt and mean slant are computed from the whole set of local estimations. In case of a curved surface, the local estimations are used to recover the normal to the surface at each position (e.g. center of the subregion).
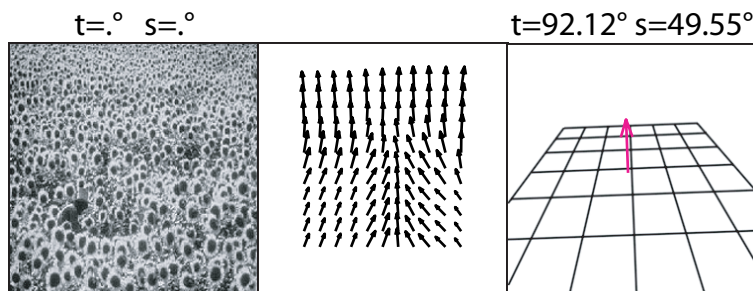

Figure 10: Estimations all over the image. Left: image of a sunflower field; middle: local estimations of the tilt (direction of arrows) and slant angle (length of arrows) located at the center of the patches; right: final planar estimation (average on all the local estimations) represented by an equivalent projected grid (using the available projection parameters).

The shape information is obtained without any assumption on the orientation statistics, e.g. without any isotropy assumption. Only a local homogeneity assumption is required, which means a local stationarity (or weak stationarity) in the spatial statistics of the frequency components. The presented method does not rely on an optimisation procedure. The shape is retrieve in a complete feedforward way. This approach can be related to known cortical cells structure dedicated to the estimation of gradients such as texture gradients [9] or optical flow [32].

6 Results

The proposed algorithm to extract the shape from the texture based on a local frequency estimation can be viewed as a successive combination of filters responses and local averaging. The complexity is linear

for the retinal filtering stage, the filters combination stage and the final shape recovery stage. The most expensive part is the computation of the filters responses which requires a Fourier transform on all the patches. This one depends on the total number of local patches (here, $21X21 = 441$ corresponding to a size of $96X96$ pixels with a shift of 8 pixels). An asymptotic precision level is reached for $7X7 = 49$ filters (frequency and orientation) which is equivalent to the reported number of filters in V1 [2] . The size of the subregions for the computation of the local normals has been fixed to $10X10$ patches. For all the presented results and tests, all these parameters have been fixed to these values. With a 2GHz processor, estimating the shape of a $256X256$ pixels large texture takes approximately 1 minute with a basic Matlab implementation.

First we present an evaluation of the accuracy of the proposed method in comparison with two other techniques developped by Super and Bovik [19] and Hwang *et al* [21]. Figure 11 presents the results obtained on the database of Super and Bovik. In [21], a comparison of the accuracy of the two methods can be found for each image. Super and Bovik obtained mean error on the estimation of the or tilt angle and on the estimation of the slant angle respectively of 3.70^o and of 2.84^o; Hwang *et al* obtained a mean error respectively of 1.75^o and of 2.18^o. The proposed algorithm reaches a mean error of 2.41^o on the tilt estimation (without taking into account the results for a very weak inclination), which is comparable to the other two techniques, and of 4.95^o on the slant estimation, which corresponds to a slightly inferior accuracy.

To evaluate the robustness of our method to different texture statistics, we created a texture database made of 208 samples. Most of the samples come from the Brodatz texture collection, some of them are artificially created, the remaining ones come from natural scenes. [1]. Each sample is primarily projected with a tilt equal to 0° with three slant values: 30°, 45° and 60°. Similarly each sample is projected for a fixed slant equal to 45° with three tilt values: 0°, 45° and 90°.

The estimation error is averaged over all the samples over all the different orientation configurations. We obtain a mean precision of 18.15^o (deviation 25.61^o) on the tilt estimation and of 12.35^o (deviation 8.10^o) on the slant estimation. Figure 12 presents estimation results on texture samples from the used

---

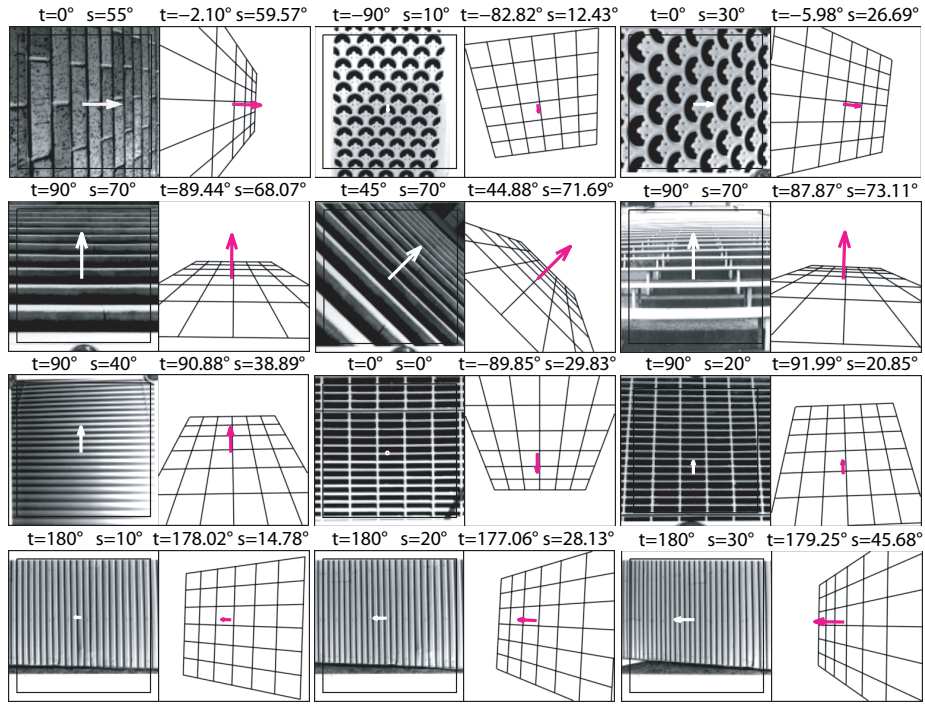[1]this database can be downloaded at this URL: http://www.lis.inpg.fr/massot

Figure 11: Results obtained on the Super and Bovik texture database [19].

database exhibiting different types of irregularities. Figure 13 presents the table detailing the results for the different orientation configurations. On figure 13 right, textures samples which lead to the worst estimations are presented. Bad estimations are obtained on textures presenting only very high frequencies which, when slanted in space, may induce aliasing problem on the spectral analysis (first row). The algorithm reaches also its limits when applied on textures which contain only very low spatial frequencies (second row). As the patches have a constant and predefined size, these spatial variations may not be included by them. Finally it is sensible to strong position and size irregularity, meaning local non-stationnarity, creating *missing* frequency information (third row left)and to the violation of the constant mean frequency hypothesis (third row right). To overcome these two last limitations two mechanisms can be combined to the frequency estimation: an automatic adaptation of the patch size to the mean frequency; a regularization procedure allowing to robustly smooth the resulting surface of local frequency estimation before the shape recovery.

The quality of the performances are difficult to evaluate since very few methods have been fully tested on a great number of natural textures presenting a large variety of irregularities. Nevertheless we compare them with the performances obtained by Lelandais *et al* [22] technique and by Hwang *et al* [21] technique on about ten natural textures [22]. These two methods obtain a precision around 1° on the tilt estimation. However it has to be noted that the textures are, in majority, relatively regular with strong orientations or structure patterns. On slant estimation, the two methods obtain respectively a precision of 15.90^o (deviation 14.90^o) and of 31.40^o (deviation 23.30^o). The mean precision obtained by our method is then comparable with, in addition, less dispersion (lower deviation). These results show that the accuracy of the propose method is comparable with those obtained by techniques especially developed to solve the shape from texture problem when these ones are evaluated on a great number of natural textures.

Figure 14 presents results obtained on natural scenes. The first four images are taken from [20] on which the proposed technique obtains very similar results. The remaining images are unreferenced examples. Figure 15 presents also estimations on different regions of a multi-textured scene. It has to be noted that the results have been obtained with exactly the same parameters as in the preceding tests excepted the size of the analyzed region which is adapted so as to respect the homogeneity assumption. All these results show the ability of the method to deal with textures directly extracted from natural scenes so as to retrieve a perspective information.

Figure 16 presents estimation results of the shape of curved surfaces. The two first samples are taken from the Super and Bovik texture database. The four following samples present cylinders with increasing irregularity. The second and the fourth one come from Sakai and Finkel [16] work and were designed to mislead moment based methods. The algorithm is able to retrieve a cylinder shape in each case. The last two samples are taken from Maureen and Mallat work [24] and the shape are closely similar to their estimations.

Finally figure 17 presents results on samples created for psychophysical experiments. The first sample comes from Knill work [12] and represents a slanted planar surface (tilt is equal to 90° and slant, to 45°)

covered by ellipses with random size and position. The following four samples are taken from Li and Zaidi work [11]. All shapes are strictly similar, only the texture characteristics are modified. For the last sample the algorithm is unable to recover the shape similarly to the visual perception. It can also be observed that the algorithm is perfectly able to recover the shape on the third sample, representing only a frequency variation with all orientation cue removed. However in this case the visual system fails at recovering the shape. As discussed in [11], this would suggest a specific overweighting of the orientation cue for the estimation of the surface curvature in a cue fusion process [34]. In this case the visual system may be strongly biased by the lack of orientation information, overriding the frequency one.

7 A model of the frequency analysis in the visual cortex

In the preceding sections a model of local frequency extraction on artificial and natural images has been presented and tested. It can be seen as a biologically plausible model of the cortical analysis. Indeed it is based on a the analysis of the amplitude spectrum which avoids the use of the phase information. Locally it is robust to translations and realized a smoothing of the texture statistics. The sampling of the spectrum is realized by specific log-normal filters which are adapted to zoom transformations and are a good approximation of the cortical cells responses. This cortical model, dedicated to the extraction of the local mean frequency, realizes the separation between the frequency and the orientation information. It allows the independent analysis of the zoom and of the rotation transformations. Two independent mechanisms can then be developed each one dedicated to the study of each corresponding gradient [11]. It is applied to the shape from texture problem and, only under an homogeneity assumption, manages to retrieve an accurate shape information even on irregular textures and reproduces the perception on some of the main psychophysical stimuli in 3D monocular perception.

Figure 18 presents the cortical model of frequency analysis on images. This one can be seen as a simple feedforward cells combination. It is divided into four stages: (1) the model of V1, (2) the frequency bands computation, (3) the filters combination and (4) the final frequency estimation over the whole image.

In each patch the responses of all filters are computed. This modelizes the cortical decomposition of

the image into its local frequency and orientation components. Figure 18.1 shows the response of each individual filter tuned to one specific central frequency and orientation and spatially localized at the center of the studied patch. It is then possible to observe the spatial evolution of the energy response at every frequency and orientation. A combination of each pair of adjacent frequency band filters, using equation 10, gives an estimation of the mean frequency at each patch center on the whole image (figure 18.3). The final combination is performed according to equation 11 (figure 18.4). This leads to a very simplified model of cells combination for frequency extraction and takes advantage of the global range of available frequencies leading to a robust estimation without discarding any information.

It is interesting to observe that this model of 3D monocular shape from texture perception can be associated to other 3D perception models, also based on bank of filters analysis, such as Qian stereoscopic model [35], Gautama *et al* shape from motion model [32] and Fleming *et al* shape from shading model [36]. It is then possible to see the emergence of similar filter based models dedicated to 3D perception each one dedicated to a specific modality but elaborated with the same biological substrate. This common basis may also allow the emergence of a common way to represent the 3D information through the different modalities and an easy way to implement their fusion (for example based on the modified weak fusion model [34]) in order to encode the visual space (i.e objects shape or natural scenes spatial layout).

Finally this frequency analysis is a general purpose model which can be adapted to other visual mechanisms such as attentional model with the creation of saliency maps and local feature extraction for object recognition and scene categorization tasks [37] [38].

8   conclusion

In this paper we have presented new filters which are a specific version of log-normal filters. Their properties (e.g., frequency orientation separability, good spectral coverage) have been stated. We have derived a method to estimate the local frequency of textured regions, which is based on the use of the whole range of frequency tuned filters. We derived the analytical estimation of the tilt and slant angle from the frequency variation, and finally, we have proposed a new algorithm to recover the shape from the texture based on

the local decomposition of the image into patches, and on the combination of the whole bank of filters. The method has been proved to be robust and accurate for both regular and irregular textures. On natural scenes, it is able to retrieve a perspective information which can be integrated as a feature in the categorization or the classification of images. The model can be improved by the integration of a robust regularization process on the estimation of the frequency all over the surface and by an automatic adaptation of the patch size to the mean frequency. To go further, this model can be transposed so as to extract the vanishing points of the projected texture by an analog analysis of the variation of the texture orientations which would lead to an improvement of the estimation performances.

References

[1] 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.

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

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

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

[5] 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.

[6] G. Wallis. Linear models of simple cells: Correspondence to real cell responses and space spanning properties. *Spatial Vision*, 14(3,4):237–260, 2001.

[7] 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.

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

[9] 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.

[10] Y. Liu, R. Vogels, and G.A. Orban. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *The Journal of Neuroscience*, 24(15):37953800, 2004.

[11] 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.

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

[13] B.J. Super and A.C. Bovik. Shape from texture using local spectral moments. *IEEE Trans. PAMI*, 17(4):333–343, 1995.

[14] J. Malik and R. holtz. Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23(2):149–168, 1997.

[15] 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.

[16] K. Sakai and H. Finkel. Spatial-frequency analysis in the perception of perspective depth. *Network: Computation in Neural Systems*, 8(3):335–352, 1997.

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

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

[19] B.J. Super and A.C. Bovik. Planar surface orientation from texture spatial frequencies. *Pattern Recognition*, 28(5):728–743, 1995.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] M. Carandini, D.J. Heeger, and J.A. Movshon. *Linearity and Gain Control in V1 Simple Cells.* CEREBRAL CORTEX, vol. XIII: CORTICAL MODELS, 1999.

[32] 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.

[33] 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/.

[34] M.S. Landy, L.T. Maloney, E.B. Johnston, and M. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389–412, 1995.

[35] N. Qian. Binocular disparity and the perception of depth. *Neuron*, 18(3):359–368, 1997.

[36] R.W. Fleming, A. Torralba, and E.H. Adelson. Specular reflections and the perception of shape. *Journal of Vision*, 4(9):798–820, 2004.

[37] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998.

[38] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Trans. PAMI*, 27(8):1265–1278, 2005.
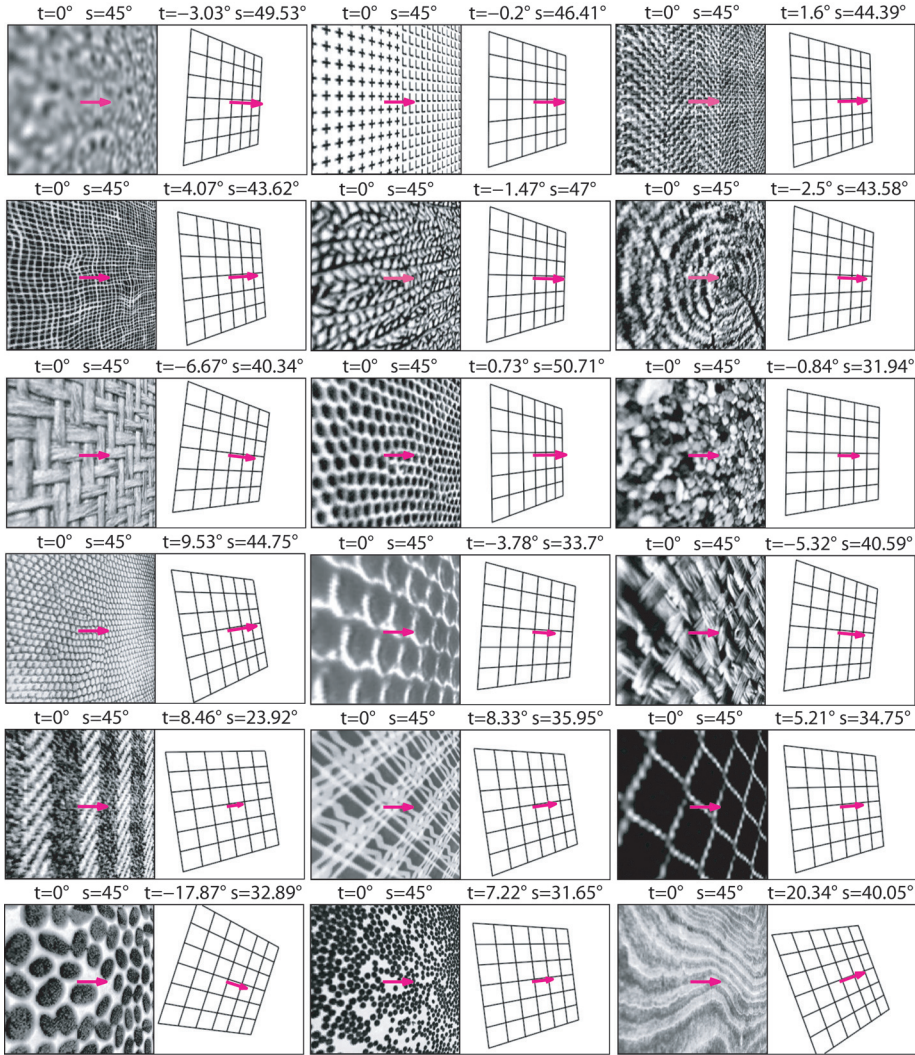
Figure 12: Examples of results obtained on our texture database (208 samples, each one projected according to five different tilt and slant configurations leading to a total of 1040 textures); these examples have been chosen so as to give an overview on the different irregularities contained in the database; results are organized such as the precision decreases from the top row to the bottom row.

σ	$\tau = 0^o$	
	tilt	slant
30°	26.65 (34.62)	7.21 (5.25)
45°	17.31 (27.85)	11.13 (7.90)
60°	15.21 (23.12)	18.83 (11.23)
τ	$\sigma = 45^o$	
	tilt	slant
45°	16.97 (22.26)	12.66 (8.27)
90°	14.61 (20.19)	11.96 (7.77)
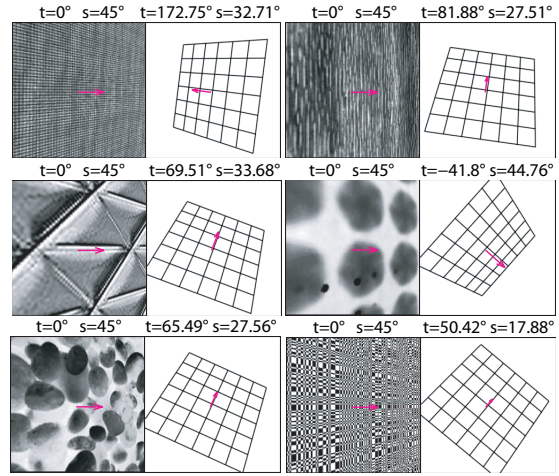
Figure 13: Mean estimation errors obtained on the five different orientation configurations over all the sample of the texture database; standard deviation is indicated in parenthesis.
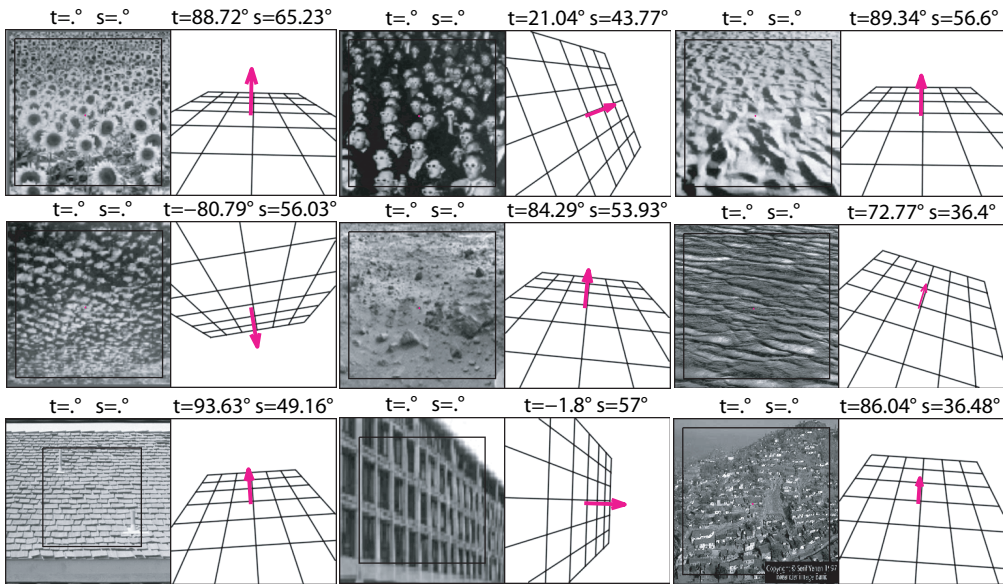

Figure 14: Results obtained on natural scenes taken from [20] and on unreferenced images; no theoretical estimation is provided for these images; from the first to the fourth image, results compare favorably with results obtained by Lindeberg and Garding in ([20]).
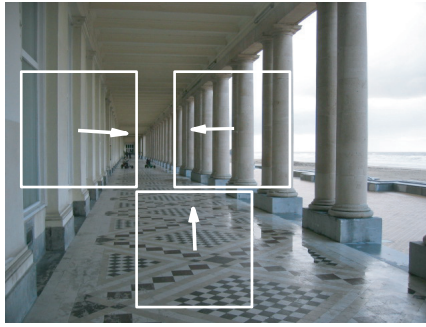
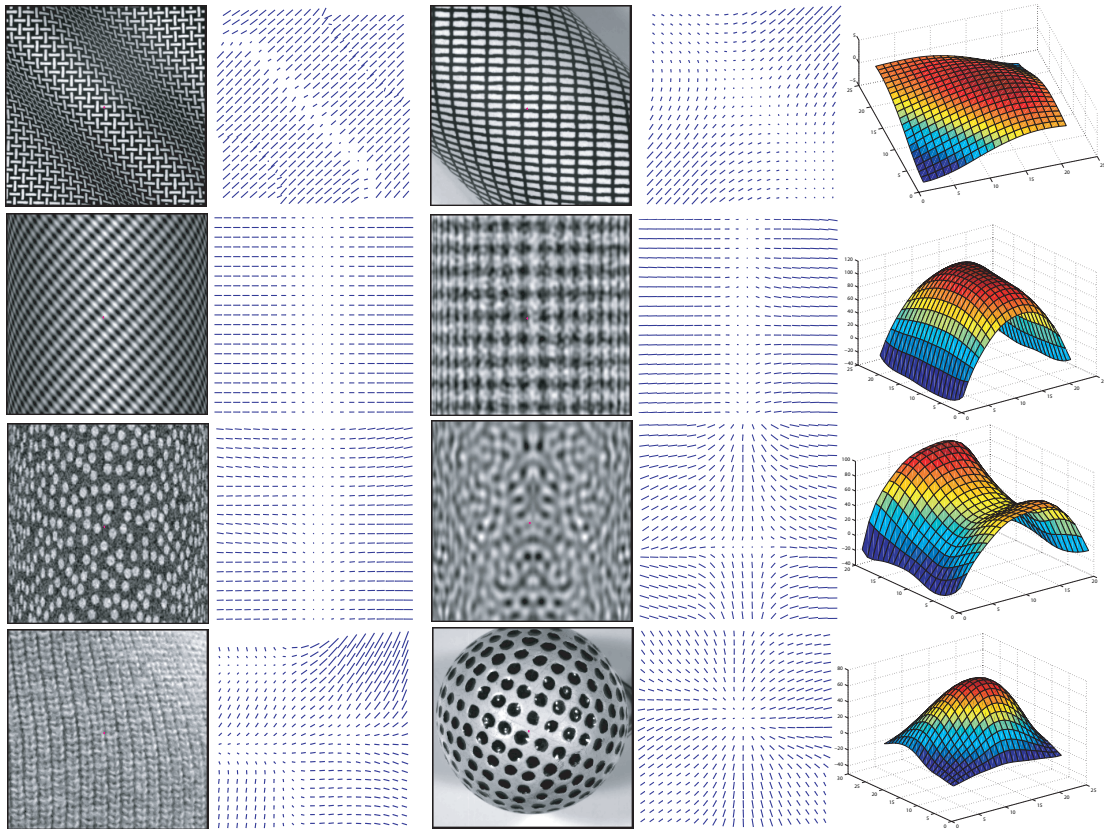Figure 15: Result obtained on a multitextured scene.

Figure 16: Results obtained on curved surfaces; each sample result presents the original texture, the needle map corresponding to the estimation and, for the samples of the second column, the shape reconstruction of the surface from the tilt and slant information (needle map and shapelets reconstruction code provided by Peter Kovesi [33]).
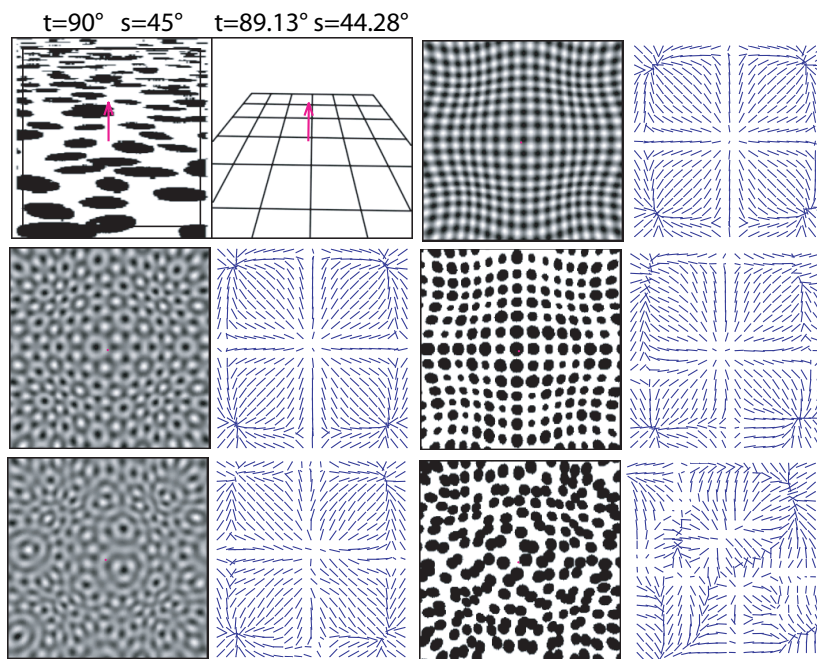
Figure 17: Results obtained on planar and curved surfaces from the psychophysical litterature [12] [11].
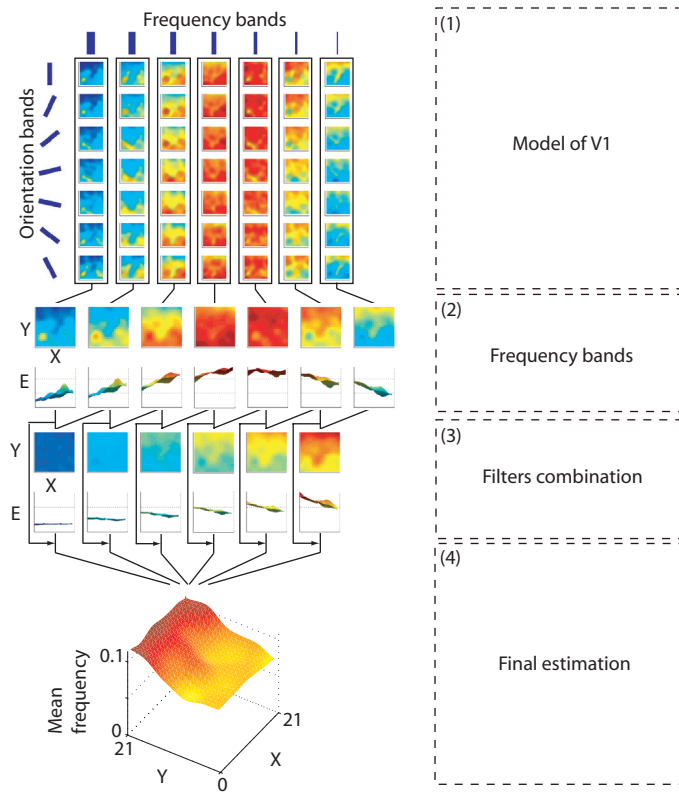
Figure 18: Cortical modeling of frequency analysis: (1) decomposition into frequencies and orientations: each small image represent the response of the individual filter on each patch over the whole image; (2) frequency bands computation after orientation normalization; (3) filters combination using equation 10; (4) final frequency estimation on the whole image using equation 11, due to the orientation normalization, each scale weight is directly the frequency band response.