Shape From Texture: Direct Estimation of Planar Surface Orientations Using Continuous Wavelet Transform*

Wen L. Hwang†, Chun-Shien Lu‡, and Pau-Choo Chung‡

† Institute of Information Science, Academia Sinica, Taipei, Taiwan

‡ Department of Electrical Engineering, National Cheng Kung University, Taiwan

Abstract

In this paper, we propose a new method to estimate the orientations of a planar surface under perspective projection. Texture variations are characterized by the ridges of continuous wavelet transform. The ridges of a continuous wavelet transform mark the places in the spatial frequency where the energy of an image is mostly concentrated. We demonstrate that the ridges contain crucial information in determining the planar surface orientations. An algorithm to determine the ridge points of a textured image is presented. The ridge points are organized into disjoint ridge surfaces. We show that textural information giving the perception of surface orientations can be characterized from a ridge surface of the textured image. The scale in the ridge surface is a parabolic function of the tilt and slant angles. This provides a closed-form solution to determine the surface orientations directly with information only from the restriction of the wavelet transform to the ridge surface. We have used our algorithm on several textured real-world images in noise-free and noisy environments. The resultant estimation of the tilt and slant angles is accurate up to a 3° average error even in low SNR environments.

Corresponding address:

Wen-Liang Hwang

Institute of Information Science, Academia Sinica,

Nankang, Taipei, Taiwan, R.O.C.

E-mail: whwang@iis.sinica.edu.tw

*EDICS: IP 1.6 Multiresolution Processing

1 Introduction

A. *Shape From Texture*

It has been pointed out that texture variations provide important cues in visual perception of three dimensional structures from a monocular image. Gibson proposed the texture gradient as a cue to derive surface orientations and subsequently verified this through psychophysical studies. In computer vision and image processing, this problem has received much attention and is called shape from texture, even though it is the orientations of a planar surface instead of the shape of a curved surface that is being computed. The reason may be that a curved surface can be approximately recovered provided that the orientations of the planes tangent to the surface are estimated. The orientations of a planar surface are the slant angle, which determines the degree of obliqueness of the surface, and the tilt angle, which is the direction towards which the texture in the image plane recedes most rapidly. A planar surface texture subjected to projective distortions causes texture compression (called the foreshortening effect) in the tilt direction, and the degree of compression in that direction is indicated by the slant angle. Fig. 1 gives an example of the impression of a surface plane with flower patterns slanted in three-dimensional space. One observes the decreasing size of the flower patterns towards the tilt direction. This corresponds to the increasing local spatial frequency towards that direction.

Several difficulties in solving the shape from texture problem are encountered, such as that the notation of the texture is not well-defined, and the surface orientations, although perceived clearly by us, are not as clear-defined. Usually, four stages are involved in solving the problem. First, a texture model is required. Secondly, a projection model which establishes the coordinate relationship between the surface plane and the image plane is built. Third, texture features describing variations due to projection model are extracted. Finally, a computational algorithm is proposed to estimate the surface orientations from the variations of texture features. We propose, in this study, a closed-form solution on the planar surface orientations subjected to perspective projection. Texture features are described as variations of local spatial frequency. Our idea is simply as follows: We suppose that the energy of a surface texture is concentrated at some frequency bands. When a surface texture is projected onto the image plane, the appearance of the surface texture in the image plane is distorted. The distortion causes the energy concentration in the resultant texture to change spatially in frequency as well. The change of energy concentration in terms of spatial frequency is caused by projective effects. If we are

able to characterize the change of spatial frequency as a function of the slant angle and tilt angle, the surface orientations can be determined.

We use the two-dimensional continuous wavelet transform with rotations (CWT), in particular the Morlet wavelet, to represent a textured image, thus ensuring that we have enough scale and rotation resolutions to characterize the frequency variations. The result is a representation with four parameters: two for spatial translations, one for rotation, and one for scale. It has been shown that the representation is sensitive to rotation, and that it gives useful information about the local frequency of an image. More precisely, when an image contains a component of the local frequency, it is sufficient to study the neighborhood of a curve of the local modulus maxima (called the ridge of the continuous wavelet transform) to characterize the frequency. The curve is located at the scale inversely proportional to the magnitude of the frequency, and it is oriented towards the direction of the frequency. We show, in this study, that the dominant spatial frequency variations that give the perceived surface orientations form a continuous ridge surface. Also, the scale of the ridge surface is a parabolic function, depending on both the tilt angle τ and the slant angle ρ. After rotating the parabola by exactly τ degrees, the parabola becomes a canonical form, depending only on the slant angle. The slant angle then can be estimated directly using some efficient methods. It is worthing noting that our algorithm estimates the tilt angle first, followed by the slant angle. Furthermore, our algorithm has been tested and shown to be a robust estimator of surface orientations in noisy environments with varying levels of white noise.

B. *Related Works*

In the literature, many algorithms have been successfully proposed[1-18]. In general, isotropic or homogeneous textures are assumed. Much of the work describes texture features in terms of texture elements. However, up to now, there have been no reliable approaches to detecting texture elements as well as describing their attributes. Alternatively, algorithms based on space/frequency representation of textured images are widely applied in describing texture distortions. These approaches do not require computing of textural elements and their attributes. So far, the backprojection technique, which finds solutions that minimize some criteria from texture variations, has been adopted in many algorithms to compute surface orientations. Usually, an exhausted search in the solution space is required.

In the following, we will briefly review some previous work about the shape from texture problem. Witkin[2] used maximum likelihood to estimate the orientations of isotropic textures under

orthographic projection. Edge directions are used as texture features in his method. Aloimonos[4] presented a backprojection method to determine the orientations of planar textured surfaces under perspective models. Homogeneous textures were assumed. His method requires finding the boundaries of texels (edges). Blostein and Ahuja[6] identified texture elements as blobs and recovered the orientations of planar textured surfaces under perspective projection based on exhaustively searching the slant and tilt angles to find the best planar fit. Jau and Chin[7] used the texture density map obtained from the Wigner distribution to characterize the texture distortion under perspective projection. A plane is then fitted to the texture density map to finish the estimation. Blake and Lieberman[8] proposed a moment-based approach to analyze textured surfaces of planar and curved types. Their statistical estimation of surface orientations is based on an optimal backprojection method, where the texels need to be identified before estimation. Krumm and Shafer[11] used the space/frequency representation to characterize the distortions of planar textured surfaces. They related local spatial frequencies between the projected texture and frontal texture by means of an affine transformation matrix. An exhaustive search was performed to find the orientations by minimizing a cost function. The main difference between theirs and other space/frequency representation methods was that they characterized a single texture by using multiple frequency peaks. Garding[12] specified texture distortions by means of foreshortening effects and texture gradients and related these distortions directly to surface orientations and curvature. Malik and Rosenholtz[13] used the affine transformation matrix to develop the relationship between local spectra. Their method was based on Garding's differential geometry. Super and Bovik[16] used the Gabor transform of homogeneous textures. They related the local spatial frequencies between the projected image and the surface texture. A backprojection method, which minimizes the variance of the frequencies of a surface texture, is used to find the surface orientations. The backprojection method was a type of coarse-to-fine sampling of the slant-tilt space.

The rest of this paper is arranged as following. In Section 2, some background for the two-dimensional continuous wavelet transform is briefly reviewed. In Section 3, we give a modulation model of textures and then derive algorithms about detecting ridge points and constructing ridge surfaces. In Section 4, we explain why the surface orientations is derivable from ridge surfaces. A computational model is proposed. Effective methods in computing the tilt and the slant angles are provided in Section 5. The Section 6 is intended to cast issue concerning robust estimation in the framework of wavelet transform. We elaborate the implementation in Section 7. Several experimental results are reported. Finally, the conclusions are summarized in Section 8.

2 The 2D CWT and Local Frequency Analysis

This section reviews briefly the two-dimensional continuous wavelet transform of images and gives the description of the specific wavelet used in this paper[21][24]. To indicate notations, we use boldface letters for vectors or matrices, and T denotes the transpose operation.

We call a complex-valued function $\psi(\mathbf{x}) \in L^2(\mathcal{R}^2)$ a wavelet whose double integral is zero:

$$\int_{\mathcal{R}^2} \psi(\mathbf{x}) d\mathbf{x} = 0.$$

Let the Fourier transform of an image $f(\mathbf{x}) \in L^2(\mathcal{R}^2)$ be

$$\hat{f}(\mathbf{w}) = \int_{\mathcal{R}^2} f(\mathbf{x}) e^{-j(\mathbf{w}^T \mathbf{x})} d\mathbf{x},$$

and let $\psi_{(\mathbf{b},\, s,\, \theta)}(\mathbf{x})$ be obtained by the translation, scaling, and rotation of $\psi(\mathbf{x})$:

$$\psi_{(\mathbf{b},\, s,\, \theta)}(\mathbf{x}) = \frac{1}{s^2} \psi(\mathbf{r}_{-\theta} \frac{\mathbf{x} - \mathbf{b}}{s}),$$

where $\mathbf{b} \in \mathcal{R}^2$, $s > 0$, and $\theta \in [0, 2\pi)$ are translation, scaling, and rotation parameters, respectively. Also,

$$\mathbf{r}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$$

is the rotation matrix of the angle θ. The two-dimensional continuous wavelet transform of $f(\mathbf{x})$ is defined simply as the convolution product with $\psi_{(\mathbf{b},\, s,\, \theta)}(\mathbf{x})$:

$$\mathcal{W}f(\mathbf{b}, s, \theta) = \frac{1}{s^2} \int_{I\!R^2} f(\mathbf{x}) \overline{\psi(\mathbf{r}_{-\theta} \frac{\mathbf{x} - \mathbf{b}}{s})} d\mathbf{x} = \int_{I\!R^2} \hat{f}(\mathbf{w}) \overline{\hat{\psi}(s\mathbf{r}_{-\theta}\mathbf{w})} e^{-j(\mathbf{w}^T \mathbf{b})} d\mathbf{w},$$

where $\overline{\psi(\mathbf{x})}$ denotes the complex conjugate of $\psi(\mathbf{x})$. If the Fourier transform of the wavelet $\psi(\mathbf{x})$ is localized at the frequency $\mathbf{k_0}$, the wavelet transform coefficient at $(\mathbf{b}, s, \theta)$ measures the frequency content of $f(\mathbf{b})$ at $\mathbf{w} = \mathbf{r}_\theta \mathbf{k_0}/s$. In other words, there is a natural association of the the polar coordinate (s^{-1}, θ) and the frequency of $\psi_{(\mathbf{b},\, s,\, \theta)}(\mathbf{x})$ at $\mathbf{b}$.

The following partial isometry allows us to interpret the squared-modulus of the continuous wavelet transform as an energy density:

$$\int_{I\!R^2} |f(\mathbf{x})|^2 \, d\mathbf{x} = \frac{1}{c_\psi} \int_{I\!R^2} d\mathbf{b} \int_0^{2\pi} d\theta \int_{0+}^{\infty} \frac{ds}{s} \, |\mathcal{W}f(\mathbf{b}, s, \theta)|^2 \,, \tag{1}$$

where c_ψ is a non-zero constant. Its value is set to 1 for convenience. One can prove that the continuous wavelet transform is invertible. This means that the image $f(\mathbf{x})$ is recovered from its wavelet transform by the formula

$$f(\mathbf{x}) = \frac{1}{c_\psi} \int_{I\!R^2} d\mathbf{b} \int_0^{2\pi} d\theta \int_{0+}^{\infty} \frac{ds}{s} \mathcal{W}f(\mathbf{b}, s, \theta) \psi_{(\mathbf{b},\, s,\, \theta)}(\mathbf{x}).$$

In implementation, the scale parameter s is taken to be a discrete value, in the form of

$$s = 2^{o + \frac{v}{n}},$$

where o is the octave, v is the voice, and n is the number of voices per octave. In particular, we use the 2D Morlet wavelet for the shape from texture problem. The 2D Morlet wavelet (also called the Gabor function) is defined as

$$\psi_{\text{M}}(\mathbf{x}) = e^{j{\mathbf{k_0}}^T \mathbf{x}} e^{-|\mathbf{x}|^2/2}$$

and is equal to

$$\hat{\psi}_{\text{M}}(\mathbf{w}) = e^{-|\mathbf{w}-\mathbf{k_0}|^2/2}$$

in the frequency domain. The Morlet wavelet is well adapted to our application since it optimizes both spatial resolution and frequency resolution simultaneously. This property is extremely important in characterizing the local spatial frequency. Strictly speaking, the function $\psi_{\text{M}}(\mathbf{x})$ is not a wavelet because of the non-zero value of $\hat{\psi}_{\text{M}}(0)$. However, when $\mathbf{k_0}$ is chosen large enough, $\hat{\psi}_{\text{M}}(0)$ is numerically negligible, and $\psi_{\text{M}}(\mathbf{x})$ becomes admissible as a wavelet. Fig. 2 shows the squared-modulus of the Morlet wavelet in polar coordinates (s^{-1}, θ). The modulus is a Gaussian function. The more voices and rotations are used, the denser the wavelet covers the polar plane.

3 Texture Description and Ridge Points

Although the notion of texture is not well-defined, we are still required to describe textures so that textural information can be quantified by a computational model. A texture model is usually proposed which contains adequate parameters such that the required information can be estimated reliably. In this study, we aim at measuring planar surface orientations from the spatial frequency variations of a textured image. A class of textures feasible in practice and suitable for our needs is given in the following[22][23].

3.1 Texture Model

We start with a reasonable texture model as a superposition of several monochromatic sinusoids with frequency $\{\mathbf{\Omega}_k\}$ and phase $\{p_k\}$:

$$f(\mathbf{x}) = \sum_{k \in I(\mathbf{x})} A_k \cos(\mathbf{\Omega}_k^T \mathbf{x} + p_k), \tag{2}$$

where $I(\mathbf{x})$ is an index subset at $\mathbf{x}$. cIf $I(\mathbf{x})$ contains an element k, this implies the kth frequency in $\{\mathbf{\Omega}_k\}$ and the kth phase in $\{p_k\}$ at $f(\mathbf{x})$. Such a model has been used successfully in many applications[22][23][24]. It should be noted that it is far from being unique to represent $f(\mathbf{x})$ as a superposition of several monochromatic sinusoids. The unique representation $(\mathcal{Z}f)(\mathbf{x})$ can be obtained by projecting $f(\mathbf{x})$ orthogonally to Hardy space $\mathbf{H^2}(R^2)$ (a collection of functions in $L^2(R^2)$ with non-negative frequency)[20]. One can show that $\mathcal{W}(\mathcal{Z}f)(\mathbf{b}, s, \theta) = 2\ \mathcal{W}f(\mathbf{b}, s, \theta)$, provided that the wavelet $\psi(\mathbf{x})$ is also in Hardy space. This is the case of the Morlet wavelet. Another point to note that an image always takes a non-negative value, but our model can be negative. Obviously, this is avoided by adding a large enough positive value to $f(\mathbf{x})$. Since the wavelet transform of a constant is zero, adding a constant does not affect the analysis of an image at all.

3.2 Ridge Points Extraction

We use the continuous wavelet transform to reliably extract the different frequency components of the modulation model. Let us take the wavelet transform of $f(\mathbf{x})$:

$$\begin{aligned}
\mathcal{W}f(\mathbf{b}, s, \theta) &= \sum_{k \in I(\mathbf{b})} A_k \frac{1}{s^2} \int_{I\!R^2} \cos(\mathbf{\Omega}_k^T \mathbf{x} + p_k) \overline{\psi_M(\mathbf{r}_{-\theta} \frac{\mathbf{x} - \mathbf{b}}{s})} d\mathbf{x} \\
&= \sum_{k \in I(\mathbf{b})} \frac{A_k}{2} \int_{I\!R^2} (\delta(\mathbf{w} - \mathbf{\Omega}_k) e^{-jp_k} + \delta(\mathbf{w} + \mathbf{\Omega}_k) e^{jp_k}) \hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{w}) e^{-j(\mathbf{w}^T \mathbf{b})} d\mathbf{w} \\
&= \sum_{k \in I(\mathbf{b})} \frac{A_k}{2} \hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k) e^{-j(\mathbf{\Omega}_k^T \mathbf{b} + p_k)}.
\end{aligned}$$

Since $\hat{\psi}_M(\mathbf{w})$ is concentrated at the frequency $\mathbf{k_0}$, the k-th frequency component $\mathbf{\Omega}_k$ of $f(\mathbf{x})$ will be concentrated around $\mathbf{r}_{\theta_k}\mathbf{k_0}/s_k$, where θ_k is the angle between $\mathbf{k_0}$ and $\mathbf{\Omega_k}$, and s_k^{-1} is the magnitude multiplier in order to scale $||\mathbf{k_0}||$ to $||\mathbf{\Omega_k}||$.

Equation (1) allows us to interpret the squared-modulus of CWT coefficients as energy density (up to a multiplication constant). Thus, the texture density at $(\mathbf{b}, s, \theta)$ is the summation of the energy of

each individual component plus $Cross\ \ terms$:

$$\begin{aligned}|\mathcal{W}f(\mathbf{b},s,\theta)|^2 &= (\sum_{k\in I(\mathbf{b})}\frac{A_k}{2}\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k)e^{-j(\mathbf{\Omega}_k^T\mathbf{b}+p_k)})(\sum_{l\in I(\mathbf{b})}\frac{A_l}{2}\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_l)e^{j(\mathbf{\Omega}_l^T\mathbf{b}+p_l)})\\ &= \sum_{k\in I(\mathbf{b})}\frac{A_k^2}{4}\left|\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k)\right|^2 + Cross\ \ terms,\end{aligned}$$

where the $Cross\ \ terms$ comes from interference between different frequency components. We shall make the terms, more precisely,

$$Cross\ \ terms = \sum_{k,l\in I(\mathbf{b}),k\neq l}\frac{A_lA_k}{4}\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_l)\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k)\cos((\mathbf{\Omega}_l-\mathbf{\Omega}_k)^T\mathbf{b}+p_l-p_k).$$

The $Cross\ \ terms$ will not be completely absent provided that there are multiple frequency components in an image. However, they will be small as long as the wavelet $\hat{\psi}(\mathbf{w})$ is well localized in frequency and the different frequency components are separated far enough such that

$$\forall l\ \ \forall k\neq l\ \ \forall s\ \ \forall\theta\ \ \hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_l)\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k)\approx 0.$$

In this case, the texture energy at $(\mathbf{b},s,\theta)$ is approximately the summation of the energy of each individual component:

$$|\mathcal{W}f(\mathbf{b},s,\theta)|^2 \approx \sum_{k\in I(\mathbf{b})}\frac{A_k^2}{4}\left|\hat{\psi}_M(s\mathbf{r}_{-\theta}\mathbf{\Omega}_k)\right|^2. \tag{3}$$

Equation 3 indicates that the texture energy at $\mathbf{b}$ is concentrated around $|I(\mathbf{b})|$ different frequency components. In polar coordinate (s^{-1},θ) , they are concentrated around $(\frac{||\mathbf{\Omega}_k||}{||\mathbf{k_0}||},\theta_k)$ correspondingly. In this paper, we use the following simple algorithm to effectively extract these points. We select points which are the squared-modulus local maxima among θ and s for any fixed $\mathbf{b}$. Let $\mathcal{N}(x)$ contain the neighborhood of the argument x, including x; then $(\mathbf{b},s_0,\theta_0)$ is selected if

$$\forall\,\mathcal{N}(s_0)\ \ \forall\,\mathcal{N}\ (\theta_0)\ \ |\mathcal{W}f(\mathbf{b},s_0,\theta_0)|^2\geq|\mathcal{W}f(\mathbf{b},\mathcal{N}(s_0),\mathcal{N}(\theta_0))|^2.$$

It is not hard to realize that the set of energy aggregation points $(\mathbf{b},s=\frac{||\mathbf{k_0}||}{||\mathbf{\Omega}_k||},\theta_k)$ is kept by using this simple process. One can then read off from these points important local parameters about the spatial frequency $\mathbf{\Omega}_k$. It is worth noting that the energy aggregation points obtained by our algorithm characterize exactly the local frequency of our texture model. Furthermore, it is a good approximation for a more general texture model of the form

$$f(\mathbf{x})=A(\mathbf{x})\cos(\phi(\mathbf{x})). \tag{4}$$

For a much more precise way of extracting the local frequency of the above model, we refer to an extension of the so-called "ridge" method from signal to image[20][24]. The "ridge" method, based on analysis of the phase of CWT, was developed by the "Marseilles group" for asymptotic signals (signals whose amplitudes vary slowly compared with their phases). In this method, ridge points are singled out where they have stationary phase. However, extracting the local frequency from the phase is relatively unreliable than compared to extracting it from the modulus in a noisy environment[25][26]. Thus, we characterize the spatial frequency from the squared-modulus. It is not hard to realize that for the specific texture model we use, the energy aggregation points found by our method are the same as those found by the extension of the "ridge" method to an image. Without abusing the terminology, we will call the energy aggregation points as ridge points hereafter. Also, it has been shown that if the original image can be represented as in Equation 4, then the image recovered from the restriction of the wavelet transform on ridges is a good approximation of the original image.

3.3 Multi-Ridge Surface Extraction

It is not unusual in a textured image for neighboring image pixels to carry similar frequency information. To represent this similarity in frequency at neighboring pixels in an image, we shall connect the ridge points nearby in a continuous wavelet transform. Ridge surfaces are, thus, constructed from ridge points using the "continuity" criterion. Obviously, there are at most $|I(\mathbf{b})|$ ridge surfaces which a pixel $\mathbf{b}$ may belong to, given the modulation model in Equation 2. The detailed algorithm connecting the ridge points is described as follows:

1. Unmark all the ridge points.
2. Let $(\mathbf{b}, s, \theta)$ be an unmarked ridge point p and S_i be a new ridge surface.
3. Assign the unmarked neighboring ridge points of p to S_i and mark them. The neighboring ridge points of p are all the ridge points that are neighbors to $(\mathbf{b}, s, \theta)$ in spatial frequency space. In implementation, we adopt the common eight neighborhood for pixel $\mathbf{b}$. The neighborhood of scale s is one voice away from s. Similarly, the neighboring angle of θ is simply $\delta\theta$ degrees away from θ.
4. Repeat the second step until there is no unmarked ridge point.

Based on this process, the frequency information of a textured image forms disjoint ridge surfaces in the continuous wavelet transform. Loosely speaking, the corresponding image patches of these ridge surfaces form the parts of the image that have "coherent" frequency structures. Although ridge surfaces are disjoint, these image patches are not necessarily disjoint. Fig. 3(a) shows a picture of a scene. One observes the slanting building from the decreasing window size. Fig. 3(b) is the scale map where the maximal modulus occurs at each pixel **b** (called single ridge). Figs. 3(c)-(f) show the maps of scale of multiple ridge surfaces. Fig. 3(c) shows the ridge surface corresponding to the slanting building, where the ridge surface tilts horizontally. Fig. 3(d) shows the ridge surface of vertical structures of the building. Figs. 3(e) and (f) correspond to the road and the cars in front of the building.

For the estimation of surface orientations which will be discussed later, we derive the shape of a ridge surface in order to use it in the shape from texture problem. Furthermore, we keep only those ridge surfaces which contain a sizable number of ridge points. This corresponds to retaining the important image patches with information related to surface orientations. Image patches with small amounts of pixels usually come from local structures of texture and, therefore, are deleted. In particular, this is indispensable for robust estimation in noisy environments where the appearance of noise tends to create dotted ridge points randomly spread in spatial frequency space.

4 Planar Texture Projection and Orientations

Our goal is to determine the orientations of a planar textured surface using information about spatial frequency from a monocular image. First, we will establish the coordinate relationship between the surface plane and the image plane. After that, the correspondence between spatial frequency in a surface texture with that in the projected image will be established. Finally, the structure of spatial frequency variations relating to surface orientations will be derived.

4.1 The Viewing Geometry

We choose coordinate systems that have been adopted in [2][16][17]. Let the coordinate systems of the world (x_w, y_w, z_w), of the surface plane (x_s, y_s, z_s), and of the image plane (x_i, y_i, z_i) be depicted in Fig. 4. Furthermore, let the optical axis point to $-z_w$, and the image plane be centered at $z_w = f < 0$. Also, the coordinates system of the image plane is set parallel to that of the world. Let z_0 be the

intersection of the optical axis with the surface plane. To define the coordinate system on the surface plane, we use the conventions by setting x_s pointing to the perceived tilt direction, z_s pointing toward the image plane, and y_s pointing in the direction such that (x_s, y_s, z_s) forms a right-handed coordinate system. The slant angle ρ is defined as the angle between z_s and z_i, which takes non-negative values between 0° and 90°. Furthermore, the angle between the x_i axis and the projection of the surface normal, i.e. z_s, into the image plane is defined as the tilt angle τ which takes values between -180° and 180°. The slant-tilt combination represents the orientations of a planar surface.

The relationship between the coordinate systems of a surface plane and the image plane under the perspective projection model is given by

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{z_w}{f} \begin{bmatrix} \sec\rho & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\tau & \sin\tau \\ -\sin\tau & \cos\tau \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \tag{5}$$

where

$$\frac{z_w}{f} = \frac{z_0}{\tan\rho(x_i\cos\tau + y_i\sin\tau) + f}. \tag{6}$$

Consider the following inversion of Equation 5:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{z_w} \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix} \begin{bmatrix} \cos\rho & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_s \\ y_s \end{bmatrix}. \tag{7}$$

The foreshortening effect is shown in the middle matrix at the right of the Equation 7, where pattern compression is perceived by a factor of $\cos\rho$ in the tilt direction, *i.e.* along the x_s axis, and not at all along the y_s axis. In the case of $\tau = 0$, Equation 5 becomes

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{z_0}{f} \begin{bmatrix} \frac{\sec\rho}{1+\frac{x_i\tan\rho}{f}} x_i \\ \frac{1}{1+\frac{y_i\tan\rho}{f}} y_i \end{bmatrix} \tag{8}$$

In order that the coordinates of $\mathbf{x}_i$ map to the same quadrant of $\mathbf{x}_s$, both $\frac{y_i\tan\rho}{f}$ and $\frac{x_i\tan\rho}{f}$ must be small than 1.

4.2 Projection of Spatial Frequency

The spatial frequency of surface textures when viewed from the projected image will be subjected to certain distortions according to the projective effects. We assume that the tilt angle τ has been

derived for the moment, and that the focal length f has been given. Let $\mathbf{r}_\tau$ denote the rotation matrix

$$\begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix}.$$

Let $\mathbf{x_i} = [x_i \;\; y_i]^T$, and $\mathbf{x_s} = [x_s \;\; y_s]^T$ be coordinates of the image plane and of surface plane, respectively. We denote $\mathbf{x} = [x_i \cos\tau + y_i \sin\tau \;\; x_i \sin\tau + y_i \cos\tau]^T$, the new image plane coordinate obtained by applying the rotation matrix $\mathbf{r}_\tau$ to $\mathbf{x_i}$. After substituting Equation 7 into texture model given in Equation 2, we obtain the projected texture in image plane:

$$g(\mathbf{x}) = \sum_{k \in I(\mathbf{x})} A_k \cos(\mathbf{\Omega}_k^T \frac{z_w}{f} \mathbf{s}_\rho \mathbf{x} + p_k), \tag{9}$$

where $\mathbf{s}_\rho$ denotes the matrix $\begin{bmatrix} \sec\rho & 0 \\ 0 & 1 \end{bmatrix}$. Let the frequency vector $\mathbf{\Omega}_k = [u_k \; v_k]$; we have

$$\begin{aligned} g(\mathbf{x}) &= \sum_{k \in I(\mathbf{x})} A_k \cos(\frac{z_w}{f}(u_k x \sec\rho + y v_k) + p_k) \\ &= \sum_{k \in I(\mathbf{x})} A_k \cos(\phi_k(\mathbf{x})) \\ &\approx \sum_{k \in I(\mathbf{x})} A_k \cos(\nabla\phi_k(\mathbf{x})^T \mathbf{x}). \end{aligned} \tag{10}$$

The perspective projection incurs a non-linear oscillation $\cos(\phi_k(\mathbf{x}))$ in the image. In Equation 10, the local spatial frequency is approximated by taking the gradient of the phase $\phi_k(\mathbf{x})$. After substituting $\frac{z_w}{f}$ from Equation 6, we have the relationship of the local spatial frequency in the image plane with the slant angle:

$$\left[\frac{\partial\phi_k(\mathbf{x})}{\partial x} \;\; \frac{\partial\phi_k(\mathbf{x})}{\partial y}\right] = \left[\frac{z_0(u_k f \sec\rho - v_k y \tan\rho)}{(x\tan\rho + f)^2} \;\; \frac{z_0 v_k}{x\tan\rho + f}\right]. \tag{11}$$

Note that the change of the spatial frequency under perspective projection is applied equally to each frequency component of the image $f(\mathbf{x})$.

4.3 Perception of Surface Orientations

The perception of dominant frequency variations in the projected image gives the tilt direction toward which the surface texture is slanting. Although this direction is obvious in our visual system, to characterize it in a computational way is not very transparent especially in complex textures, where dominant

frequency variations may be the consequence of combining varying surface properties. However, if the perceived surface orientations of a texture come from a sense of strong and smooth variations of the spatial frequency, then this information about spatial frequency variations can certainly be determined from ridge surfaces. Since there are many structural "coherent" components in a surface texture, it is obvious that certain ridge surfaces are irrevalent to the surface orientations. Fig. 5(a) shows a complex texture constructed by adding a slanted sinusoid to a frontal straw cloth. The slanted-sinusoid gives the perceived surface orientations. For simplicity, only two ridge surfaces are shown in Fig. 5(b). Surface orientations can be estimated from the top ridge surface (with larger scale values) while the bottom ridge surface containing detailed local structures of straw cloth is not considered as part of the estimation task.

Suppose that the surface orientations are derivable from the ridge surface $\mathcal{S}$, and $g_{\mathcal{S}}(\mathbf{x})$ is the image reconstructed by restricting the wavelet transform to $\mathcal{S}$. It is worth mentioning that $g_{\mathcal{S}}(\mathbf{x})$ may be very different from our textured image since there are surface properties in a textured image that are not captured by the ridge points in $\mathcal{S}$. To recapitulate what we have derived at Equations 10 and 11, $g_{\mathcal{S}}(\mathbf{x})$ is approximated by

$$g_{\mathcal{S}}(\mathbf{x}) \approx A_{\mathcal{S}}(\mathbf{x}) \cos(\nabla \phi_{\mathcal{S}}(\mathbf{x})^T \mathbf{x}), \tag{12}$$

where

$$\left[\frac{\partial \phi_{\mathcal{S}}(\mathbf{x})}{\partial x} \;\; \frac{\partial \phi_{\mathcal{S}}(\mathbf{x})}{\partial y} \right] = \left[\frac{z_0 (u_{\mathcal{S}}(\mathbf{x}) f \sec\rho - v_{\mathcal{S}}(\mathbf{x}) y \tan\rho)}{(x \tan\rho + f)^2} \;\; \frac{z_0 v_{\mathcal{S}}(\mathbf{x})}{x \tan\rho + f} \right]. \tag{13}$$

The vector $[\frac{\partial \phi_{\mathcal{S}}(\mathbf{x})}{\partial x} \;\; \frac{\partial \phi_{\mathcal{S}}(\mathbf{x})}{\partial y}]^T$ is the frequency giving the impression of the receding textured image under perspective projection. The frequency vector $[u_{\mathcal{S}}(\mathbf{x}) \; v_{\mathcal{S}}(\mathbf{x})]^T$ may be originated from a single texture component, *e.g.* monocomponent sinusoid, or from a more complex textural components *e.g.* flower in Fig. 1. Mentioned in the previous section, the x axis in coordinate system $\mathbf{x} = [x \; y]^T$ points towards the tilt direction, that is, $x = x_i \cos\tau + y_i \sin\tau$. Hence, $\frac{\partial \phi_{\mathcal{S}}(\mathbf{x})}{\partial x}$ is the frequency component in the tilt direction, while $\frac{\partial \phi_{\mathcal{S}}(\mathbf{y})}{\partial y}$ is the frequency component perpendicular to the tile direction.

Since our perception is in the tilt direction, this means that a large number of image pixels in $g_{\mathcal{S}}(\mathbf{x})$ satisfy $||u_{\mathcal{S}}(\mathbf{x})|| > ||v_{\mathcal{S}}(\mathbf{x})||$. In other words, a majority of points in $g_{\mathcal{S}}(\mathbf{x})$ contains the part of the spatial frequency which varies mainly towards the tilt direction. Most of all, this argument is consistent with our perceptual experience that the spatial frequency is receding away the most along the tilt direction. For simplicity in computation, we further assume that $||u_{\mathcal{S}}(\mathbf{x})|| >> ||v_{\mathcal{S}}(\mathbf{x})|| \approx 0$ and that $u_{\mathcal{S}} = u_{\mathcal{S}}(\mathbf{x})$ and $v_{\mathcal{S}} = v_{\mathcal{S}}(\mathbf{x})$ for all $\mathbf{x}$ in the domain of $g_{\mathcal{S}}(\mathbf{x})$. In this case, Equation 13 is

reduced to a simpler form:

$$\begin{aligned}\left[\frac{\partial\phi_{\mathcal{S}}(\mathbf{x})}{\partial x} \quad \frac{\partial\phi_{\mathcal{S}}(\mathbf{x})}{\partial y}\right] &\approx \left[\frac{z_0(u_{\mathcal{S}} f\sec\rho - v_{\mathcal{S}} y\tan\rho)}{(x\tan\rho+f)^2} \quad 0\right] \\ &= \left[(\frac{z_0}{f})\frac{u_{\mathcal{S}}\sec\rho - v_{\mathcal{S}}\frac{y\tan\rho}{f}}{(1+\frac{x\tan\rho}{f})^2} \quad 0\right]. \end{aligned} \tag{14}$$

Based on Equation 8, where $\frac{y\tan\rho}{f} < 1$, and $\frac{x\tan\rho}{f} < 1$, we have $u_{\mathcal{S}}\sec\rho >> v_{\mathcal{S}}\frac{y\tan\rho}{f}$, and Equation 14 will become

$$\left[\frac{\partial\phi_{\mathcal{S}}(\mathbf{x})}{\partial x} \quad \frac{\partial\phi_{\mathcal{S}}(\mathbf{x})}{\partial y}\right] \approx \left[(\frac{z_0}{f})\frac{u_{\mathcal{S}}\sec\rho}{(1+\frac{x\tan\rho}{f})^2} \quad 0\right]. \tag{15}$$

The observed spatial frequency variation is irrelevant to y, the direction perpendicular to the tilt. If we choose $\mathbf{k_0}$ at $(k_0 = ||\mathbf{k_0}||, 0)$ with respect to coordinate system $\mathbf{x_i}$, then the corresponding ridge points $(\mathbf{x}, s(\mathbf{x}), \theta(\mathbf{x}))$ to Equation 12 form a continuous ridge curve with values given respectively by

$$\begin{aligned} s(\mathbf{x}) &= \frac{fk_0(1+\frac{x\tan\rho}{f})^2}{z_0 u_{\mathcal{S}}\sec\rho}, \\ \theta(\mathbf{x}) &= \tau. \end{aligned} \tag{16}$$

It is readily clear that the scale $s(\mathbf{x})$ is a parabolic function of x. Fig. 6(a) shows a picture of a sinusoid slanted horizontally. The scale map of the ridge surface of the picture is shown in Fig. 6(b). Fig. 6(c) shows the 3D profile of the scale map of the ridge surface. The ridge surface is a parabola. The 1D cross section of the scale map of the ridge surface is shown at Fig. 6(d). For orthographic projection, where the focal length f is infinite, the spatial frequency along tilt direction is $u_{\mathcal{S}}\sec\rho$, irrelevant to the location $\mathbf{x}$. For perspective projection, where the focal length f is finite, the spatial frequency along the tilt direction is described as $(\frac{z_0}{f})\frac{u_{\mathcal{S}}\sec\rho}{(1+\frac{x\tan\rho}{f})^2}$, inversely proportional to the scale of the ridges. In the case of $\rho = 0$, this term becomes $(\frac{z_0}{f})u_{\mathcal{S}}$. So, the terms of $\frac{\partial\phi_{\mathcal{S}}(\mathbf{x})}{\partial x}$ have individual meaning: $u_{\mathcal{S}}$ is the spatial frequency induced by the surface orientations along the tilt direction, $\sec\rho$ is the result of slanting the surface plane, and $\frac{z_0}{f}$ and $\frac{1}{(1+\frac{x_i}{f}\tan\rho)^2}$ come from perspective projection.

Once the tilt is obtained, the slant angle ρ can be determined from the scale of ridge curves using Equation 17. This gives us important information in developing an algorithm for estimating the slant and the tilt angles.

5 Estimation of Surface Orientations

We have shown that the scale of the ridge surface compatible with the perceiving surface orientations is a parabolic surface. The description of this surface can be further simplified once the tilt direction is obtained. The slant angle can, therefore, be computed in a convenient way. First of all, however, we need to determine the tilt angle τ.

5.1 Computing the tilt angle

Assume that the ridge surface $\mathcal{S}$ gives the perception of surface orientations. Although the rotation θ of the ridge points $(\mathbf{b}, s, \theta)$ on the ridge surface $\mathcal{S}$ is the tilt angle τ(see Equation 17), it is impossible to obtain this value precisely from the rotation in the ridge points. In implementation, the number of rotations used in computing the continuous wavelet transform is always limited precision by a quantization $\delta\theta$. As a consequence, we can not expect that there will be enough number of rotations available for determining the tilt angle to any precision. However, we know earlier that the scale of ridge surface $\mathcal{S}$ is a parabola, depending on the value of the tilt τ, slant ρ, and coordinates x_i and y_i. The parabola is a canonical form provided that we represent it using the new coordinate system $\mathbf{x}$, obtained by rotating $\mathbf{x_i}$ exactly τ degrees. We denote the ridge points $([x_i, y_i]^T, s_i, .)$ on $\mathcal{S}$. To estimate the tile angle, we approximate the scale on the ridge surface $\mathcal{S}$ by using the quadratic function of x_i and y_i,

$$P(x_i, y_i) = a_1 x_i^2 + a_2 x_i y_i + a_3 y_i^2 + a_4 x_i + a_5 y_i + a_6.$$

The coefficients $a_1, \ldots, a_6$ are obtained by using the least square fitting method, which minimizes the mean-squared error of s_i with $P(x_i, y_i)$ (see Equation 17 below). We know that the above formula can be expressed in the canonical form:

$$P^{'}(x, y) = \tilde{a_1} x^2 + \tilde{a_4} x + \tilde{a_6},$$

where the mixed-product term is eliminated by a rotation of τ degrees of the coordinate system. The new formula $P^{'}(x, y)$ is independent of the axis y, which is pointing in the direction perpendicular to that of the tilt. We thus obtain the tilt angle by a proper change of the coordinate system. It is a simple matter to obtain the tilt angle of $\mathcal{S}$ from the following formula:

$$\tau = \tan^{-1} \frac{a_5}{a_4}.$$

Thus far we have given no explanation how the ridge surface $\mathcal{S}$ is selected. The likelihood of a testing ridge surface $\mathcal{T}$ as a candidate surface is obtained by computing the mean-squared error $\mathcal{E}$ between the scale s_i at (x_i, y_i) on the test surface $\mathcal{T}$ and the least squared fitting result $P_{\mathcal{T}}(x_i, y_i)$:

$$\mathcal{E} = \frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} (P_{\mathcal{T}}(x_i, y_i) - s_i)^2, \tag{17}$$

where $|\mathcal{T}|$ is the size of the domain of ridge $\mathcal{T}$. A set of candidate ridge surfaces is chosen after comparing the resultant mean-squared error $\mathcal{E}$ against a given threshold ϵ. Since the perception of the surface orientations appears as a global phenomenon, the ridge surface $\mathcal{S}$ selected from candidate set is thus the one which covers the largest number of image pixels.

5.2 Computing the slant angle

After the tilt angle is estimated, the slant angle can be calculated by means of the following methods. First of all, we rotate exactly τ degree and obtain the coordinate system with respect to $\mathbf{x} = [x\ y]^T$. The ridge surface $\mathcal{S}$ is composed of many parabolic ridge curves along x. Thus, we can treat each ridge curve separately in the slant angle estimation. The final result is obtained by averaging the slant angles found from each ridge curve separately. We present two methods below for the computation of the slant angle from a ridge curve.

5.2.1 Computation of slant angle by voting method

From Equation 17, the scales $s([x_1, y_0]^T)$ and $s([x_2, y_0]^T)$ of any two points in a ridge curve is given by

$$\frac{s([x_1, y_0]^T)}{s([x_2, y_0]^T)} = \frac{(1 + \frac{x_1}{f} \tan\rho)^2}{(1 + \frac{x_2}{f} \tan\rho)^2} = C^2.$$

Solving the above equation for ρ, we have the slant angle:

$$\rho = \tan^{-1}(\frac{f(1-C)}{Cx_2 - x_1}), \tag{18}$$

where $C = \sqrt{\frac{s([x_1,y_0]^T)}{s([x_2,y_0]^T)}}$. Although, the slant angle can be computed from any two points by using the above equation, for consideration of robust estimation, it is better to obtain the angle by using a voting procedure. We cast a vote to the slant angle obtained from Equation 18 for each pair of points on the same ridge curve. Usually, the slant angle ρ is determined from the ballot for which the maximal number of votes is cast.

5.2.2 Computation of slant angle by curve fitting

An alternative way of estimating the slant angle ρ is to use the curve fitting method. We have shown in Equation 17 that the scale of the ridge curve is a parabolic function:

$$s(\mathbf{x}) = \frac{fk_0(1+\frac{x\tan\rho}{f})^2}{z_0 u_{\mathcal{S}} \sec\rho}.$$

It can be approximated by using the formula $\beta_2 x^2 + \beta_1 x^1 + \beta_0$. Equating the coefficients of the above two equations, the slant angle is related to the three coefficients β_0, β_1, and β_2. After simple calculation, the slant angle ρ is derived by

$$\rho = \tan^{-1}(\frac{f\ \beta_1}{2\beta_0}).$$

Remark. The ridge curve is concave. In practice, to prevent β_2 from changing sign during curve fitting, and therefore the ridge curve becomes convex, we do not use β_2 in computing the slant angle. However, to optimize the curve fitting, the Lagrange multiplier, with the constraint $\beta_2 \geq 0$ is used.

6 Robust Estimation with Additive White Noise

We exploited the robust processing of textural information by first step analysis of the energy distribution of white noise in the wavelet transform. Let $\tilde{n}(\mathbf{x})$ be a wide-sense stationary, white noise random field with variance σ, and let $g(\mathbf{x})$ be the projected image given in Equation 10. The noisy image $\tilde{h}(\mathbf{x})$ is given by

$$\tilde{h}(\mathbf{x}) = g(\mathbf{x}) + \tilde{n}(\mathbf{x}).$$

The wavelet transform of $\tilde{n}(\mathbf{x})$ is $\mathcal{W}\tilde{n}(\mathbf{b}, s, \theta) = \frac{1}{s^2}\int_{I\!R^2} \tilde{n}(\mathbf{x})\overline{\psi(\mathbf{r}_{-\theta}\frac{\mathbf{x}-\mathbf{b}}{s})}d\mathbf{x}$. We denote by $E(X)$ the expected value of the random variable X. It is not hard to see that the decay of $E(|\mathcal{W}\tilde{n}(\mathbf{b}, s, \theta)|^2)$ is proportional to $1/s^2$. Indeed,

$$|\mathcal{W}\tilde{n}(\mathbf{b}, s, \theta)|^2 = \frac{1}{s^4}\int_{I\!R^2}\int_{I\!R^2} \tilde{n}(\mathbf{u})\tilde{n}(\mathbf{v})\overline{\psi(\mathbf{r}_{-\theta}\frac{\mathbf{u}-\mathbf{b}}{s})}\psi(\mathbf{r}_{-\theta}\frac{\mathbf{v}-\mathbf{b}}{s})d\mathbf{u}d\mathbf{v}.$$

Since $\tilde{n}(\mathbf{x})$ is a white noise, $E(\tilde{n}(\mathbf{u})\tilde{n}(\mathbf{v}))$ is $\sigma^2\delta(\mathbf{u}-\mathbf{v})$. Hence,

$$\begin{aligned} E(|\mathcal{W}\tilde{n}(\mathbf{b}, s, \theta)|^2) &= \frac{1}{s^4}\int_0^{2\pi}\int_{I\!R^2}\int_{I\!R^2} E(\tilde{n}(\mathbf{u})\tilde{n}(\mathbf{v}))\overline{\psi(\mathbf{r}_{-\theta}\frac{\mathbf{u}-\mathbf{b}}{s})}\psi(\mathbf{r}_{-\theta}\frac{\mathbf{v}-\mathbf{b}}{s})d\mathbf{u}d\mathbf{v}d\theta \\ &= \frac{\sigma^2}{s^2}||\psi(\mathbf{x})||^2. \end{aligned}$$

This indicates that the average energy of white noise $\tilde{n}(\mathbf{x})$ decays with the increase of the scale s. The energy of white noise is mostly in fine scale. However, the scales where the energy is concentrated around in determining surface orientations is a parabolic strip bending towards smaller scale as the values of x is increasing. Thus, the effect of white noise occurs mostly for larger x. As a consequence of the uneven distribution of the noise component in the wavelet transform, the ridge surface $\mathcal{S}$ is distorted unevenly along x as well. In addition, the ridge points at a finer scale are usually caused by local surface patterns, are relatively unimportant for estimation of surface orientations. We shall avoid using the smaller scales in the ridge surface $\mathcal{S}$ in order to have a reliable estimation.

7 Detailed Implementation and Experimental Results

We will outline the detailed implementation of the algorithm for robust estimation of surface orientations from a monocular textured image. In addition, we will discuss issues in the selection of free parameters in our algorithm. Experiments have been conducted in noisy environments where white noise with varying variance was added to a textured image. The summary of our algorithm is listed as follows.

Algorithm

- Compute the continuous wavelet transform of a textured image.
- Detect ridge points at each fixed $\mathbf{b}$ by selecting the scale and the angle where the squared-modulus at the points is a local maximum. Neighboring ridge points are connected as ridge surfaces.
- The ridge surface $\mathcal{S}$ covering the largest number of image pixels is chosen for computing surface orientations.
- Estimate the tilt angle τ by fitting the scale of $\mathcal{S}$ by using a parabola.
- Rotate $\mathcal{S}$ exactly τ degrees to a new coordinate system, and estimate the slant angle ρ using either the voting or the curve fitting method.

Since perception of three-dimensional structures from texture variations occurs as a global phenomenon in an image, we expect that the spatial frequency variations used to perceive the surface

orientations will be a dominant effect in many places. As such, the ridge surface which covers the largest number of image pixels is selected in our algorithm. Also, discussed earlier the fine scale ridge points are usually local surface patterns, which have relatively unimportant and unreliable information about the surface orientations. Hence, in implementation, we usually restrict the ridge surface $\mathcal{S}$ to the middle part of the image in computing the tilt and slant angles.

Experiments

Two sets of textured images were used to demonstrate our method. The first set of images shown in subfigures 1 of Figs. 7(a)-(i) was Super and Bovik's images[16]. The second set was Brodatz's natural images[19], shown in subfigures 1 of Figs. 7(j)-(p). All the images were 128×128 pixels with 256 gray levels. The Morlet wavelet with $k_0 = 6.9$ was used. In our experiments, we used four octaves, eight voices, and eight rotations to perform the wavelet transform. A ridge surface containing the largest number of image pixels was selected using multi-ridge detection. Subfigures 2 of Figs. 7(a)-(p) show the selected ridge surfaces. The darker the gray level at a position were, the smaller the scale at the position would be. And, the black pixels indicate the places outside the domain of the ridge surface. Each subfigure 2 is framed by a white window, denoting that the surface orientations were estimated by restricting the ridge surface to the enclosed region. Tables 1 and 2 show the estimated results of the tilt and the slant angles from Super and Bovik's images. The average error of the tilt angle as measured by our method was 1.45°, and the average errors of our slant angle using the voting and curve fitting were 2.21° and 2.41°, respectively. On the other hand, the average errors of tilt and slant angles obtained by Super and Bovik's method were 4.08° and 2.99°, respectively. In addition, Table 3 shows the estimated results of the tilt and the slant angles from Brodatz's images. The average errors of our tilt angle were 2.61°, and those of our slant angle using voting and curve fitting were 1.80° and 2.70°, respectively. It is found that the estimated errors using our method for both Super and Bovik's images and Brodatz's images are all within 3°. Compared with Super and Bovik's method, our method is better in the tilt angle and slightly better in the slant angle.

Furthermore, to demonstrate the robustness of our method, four images, Venetian blind, Aluminum wall-1, D52, and D82, were added with white noise with SNR ranging from 20 db to -5 db. We used only one realization of white noise in conducting the experiments. Subfigures 1 of Figs. 8(a)-(d) show the noisy images Venetian blind with SNR 20, 10, 0, and -5 db. Their detected ridges are shown in subfigures 2, respectively. It is empirically observed that the fours ridges were clearly detected, and that the number of ridge points decreased as SNR decreased. Besides, Figs. 8(e1) and (f1) show the

noisy D52 with SNR 0 db and the noisy D82 with SNR 10 db. These are the noisy levels of D52 and D82 that successful estimations were achieved. Their detected ridges are shown in Figs. 8(e2) and (f2), respectively. All the estimated surface orientations of noisy images are shown in Table 4. One observes that the estimated results were almost the same as those of the noise-free images except for D52 with -5 db and D82 for 0 db and -5 db, where our method failed.

8 Conclusions

We have proposed a new method for estimating the orientations of a planar surface under the perspective projection model. Texture variations are effectively characterized by means of the ridge of the continuous wavelet transform. Unlike much previous work, we use the structural variations of local spatial frequency for estimating the surface orientations. Even though a textured image has more than a single dominant frequency component, the proposed multiple ridge detection algorithm is able to choose the ridge surface that describe the surface orientations. We have derived a closed-form formulas for tilt and slant angles such that estimations can be achieved directly. The effectiveness of our method has been demonstrated by empirical results on real-world images. Furthermore, the presented method is highly white noise tolerant even when the SNR of noisy images is as low as 0 db or -5 db. Some natural textures, especially those which can not be accurately described from the ridge surface, will fail in detection using our method. Examples have been found such as the disordered textures in Brodatz's album.

Acknowledgment: Wen L. Hwang would like to thank Dr. Bruno Torrésani and Dr. Rene Carmona for many valuable discussions of the time-frequency method. The authors would like to thank Dr. Super for giving us his textured images and much useful help. Also, our thanks go to Dr. Mark Liao for several suggestions.

References

[1] R. Bajcy and L. Lieberman, "Texture gradients as a depth cue", *Computer Graphics and Image Processing*, 1976, pp. 52-67.

[2] A. P. Witkin, "Recovering surface shape and orientation from texture", *Artificial Intelligence*, Vol. 17, 1981, pp. 17-45.

[3] L. S. Davis, L. Janos, and S. M. Dunn, "Efficient Recovery of Shape from Texture", *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. 5, 1983, pp. 485-492.

[4] J. Aloimonos, "Shape from texture", *Biological Cybernetics*, Vol. 58, 1988, pp. 345-360.

[5] K. Kanatani and T. Chou, "Shape from texture: General principle", *Artificial Intelligence*, Vol. 38, 1989, pp. 1-49.

[6] D. Blostein and N. Ahuja, "Shape from texture: Integrating surface element extraction and surface estimation", *IEEE Trans . Pattern Anal. and Machine Intell.*, Vol. 1, 1989, pp. 1233-1251.

[7] Y. C. Jau and R. T. Chin, "Shape from texture using the Wigner distribution", *Comput. Vision Graphics Image Proces.*, Vol. 52, 1990, pp. 248-263.

[8] A. Blake and L. Lieberman, "Shape from texture: Estimation, isotropy, and moments", *Artificial Intelligence*, Vol. 45, 1990, pp. 323-380.

[9] L. G. Brown and H. Shvaytser, "Surface orientation from projective forshortening of isotropic texture autocorrelation", *IEEE Trans . Pattern Anal. and Machine Intell.*, Vol. 12, 1990, pp. 584-588.

[10] M. R. Turner, G. L. Gerstein, and R. Bajcsy, "Underestimation of visual texture slant by human observers: a model", *Biological Cybernetics*, Vol. 65, 1991, pp. 215-226.

[11] J. Krumm and S. Shafer, "Shape from periodic texture using the spectrogram", *Proc. Conf. Computer Vision and Pattern Recognition* , 1992, pp. 284-289.

[12] J. Garding, "Shape from texture for smoothed curved surfaces in perspective projection", *J. Math. Imaging and Vision*, Vol. 2, 1992, pp. 329-352.

[13] J. Malik and R. Rosenholtz, "A differential method for computing local shape-from-texture for planar and curved surfaces", *Proc. Conf. Computer Vision and Pattern Recognition* , 1993, pp. 267-273.

[14] J. Garding, "Direct estimation of shape from texture", *IEEE Trans. Pattern Anal. Artificial Intell.*, Vol. 15, 1993, pp. 1202-1208.

[15] J. V. Stone and S. D. Isard, "Adaptive scale filtering: A general method for obtaining shape from texture", *IEEE Trans. Pattern Anal. Artificial Intell.*, Vol. 17, 1995, pp. 713-718.

[16] B. Super and A. C. Bovik, "Planar surface orientation from texture spatial frequencies", *Pattern Recognition*, Vol. 28, No. 5, 1995, pp. 729-743.

[17] B. Super and A. C. Bovik, "Shape from texture using local spectral moments", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, No. 4, 1995, pp. 333-343.

[18] C. S. Lu, W. L. Hwang, H. Y. Mark Liao, and P. C. Chung, "Shape from texture based on the ridge of continuous wavelet transform", *IEEE Inter. Conf. on Image processing*, 1996.

[19] P. Brodatz, "Textures: A photographic album for artists and designers", *Dover Publications*, 1966.

[20] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies", *IEEE Trans. Inform. Theory*, Vol. 38, No. 2, 1992, pp. 644-664.

[21] J. -P. Antoine, P. Carrette, R. Murenzi, and B. Piette, "Image analysis with two-dimensional continuous wavelet transform", *Signal Processing*, Vol. 31, 1993, pp. 241-272.

[22] A. C. Bovik, "Analysis of multichannel narrowband filters for image texture segmentation", *IEEE Trans. Signal Processing*, Vol. 39, 1991, pp. 2025-2043.

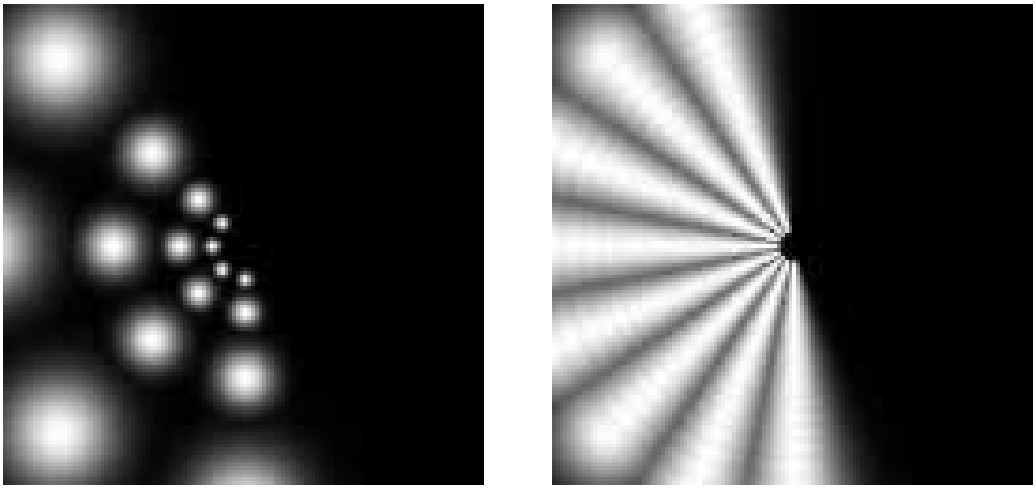
[23] A.C. Bovik, N. Gopal, T. Emmoth, A. Restrepo(Palacios), "Localized measurements of emergent frequencies by Gabor wavelets", *IEEE Trans. Inf. Th.*, Vol. 38, 1992, pp. 691-712.

[24] C. Gonnet and B. Torrésani, "Local frequency analysis with two-dimensional wavelet transform", *Signal Processing*, 1994, pp. 389-404.

[25] R. Carmona, W.L.Hwang, B. Torrésani, "Characterization of signals by the ridges of their wavelet transform", *preprint*, submitted to IEEE Trans. on Signal Processing(1996).

[26] R. Carmona, W.L.Hwang, B. Torrésani, "Multiridge detection and time-frequency reconstruction", *preprint*, submitted to IEEE Trans. on Signal Processing(1996).

Figure 1: Image of flower patterns.



(a) (b)

Figure 2: The squared-modulus of the Morlet wavelet in the polar plane (s^{-1}, θ): the lowest frequency is at the center, the orientation is 0° at the bottom and increases in the clock-direction. (a) The parameters of CWT are 4 octaves, 1 voice/octave, and 4 rotations; (b) The parameters of CWT are 4 octaves, 8 voices/octave, and 8 rotations.

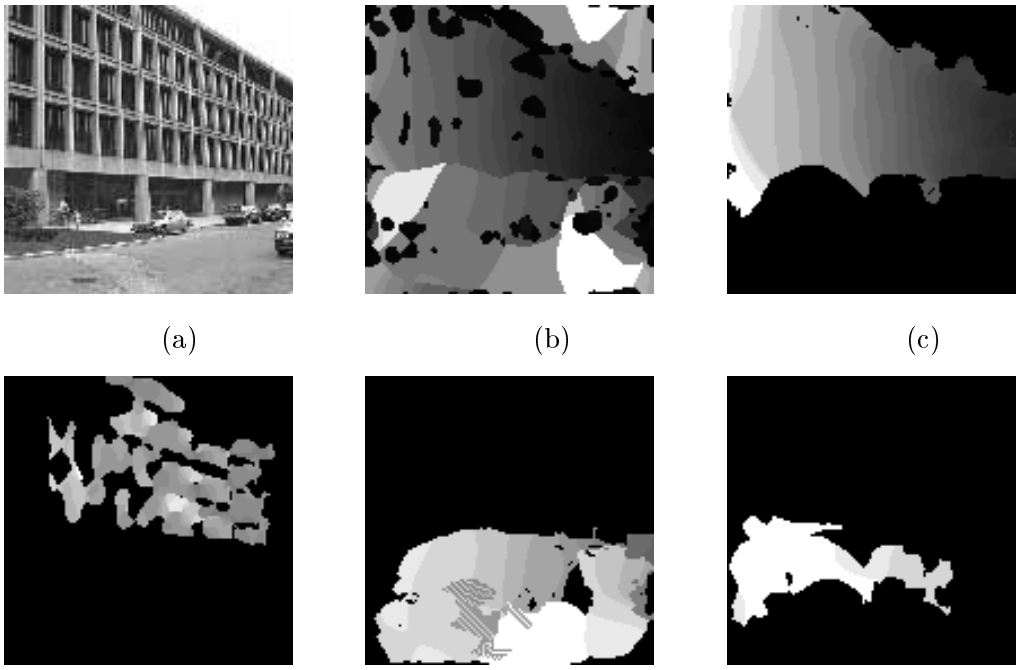


Figure 3: A scene and their ridges: (a) The scene; (b) Single ridge; (c) The ridge surface corresponding to the dominate variations of the slanting building; (d) The ridge surface corresponding to the vertical variations of the building; (e)(f) The ridge surfaces corresponding to the road, cars, and other structures in front of the building.

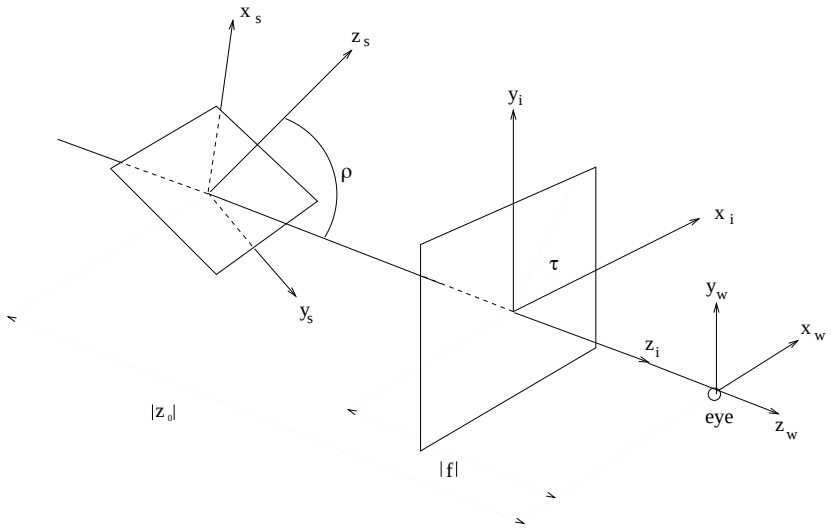
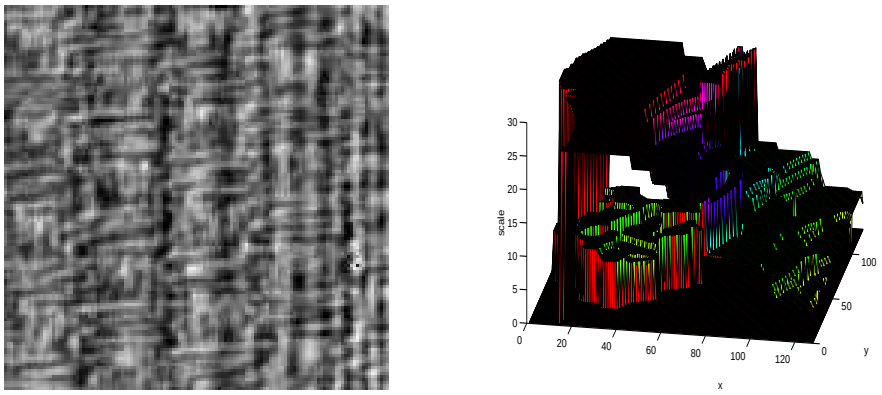
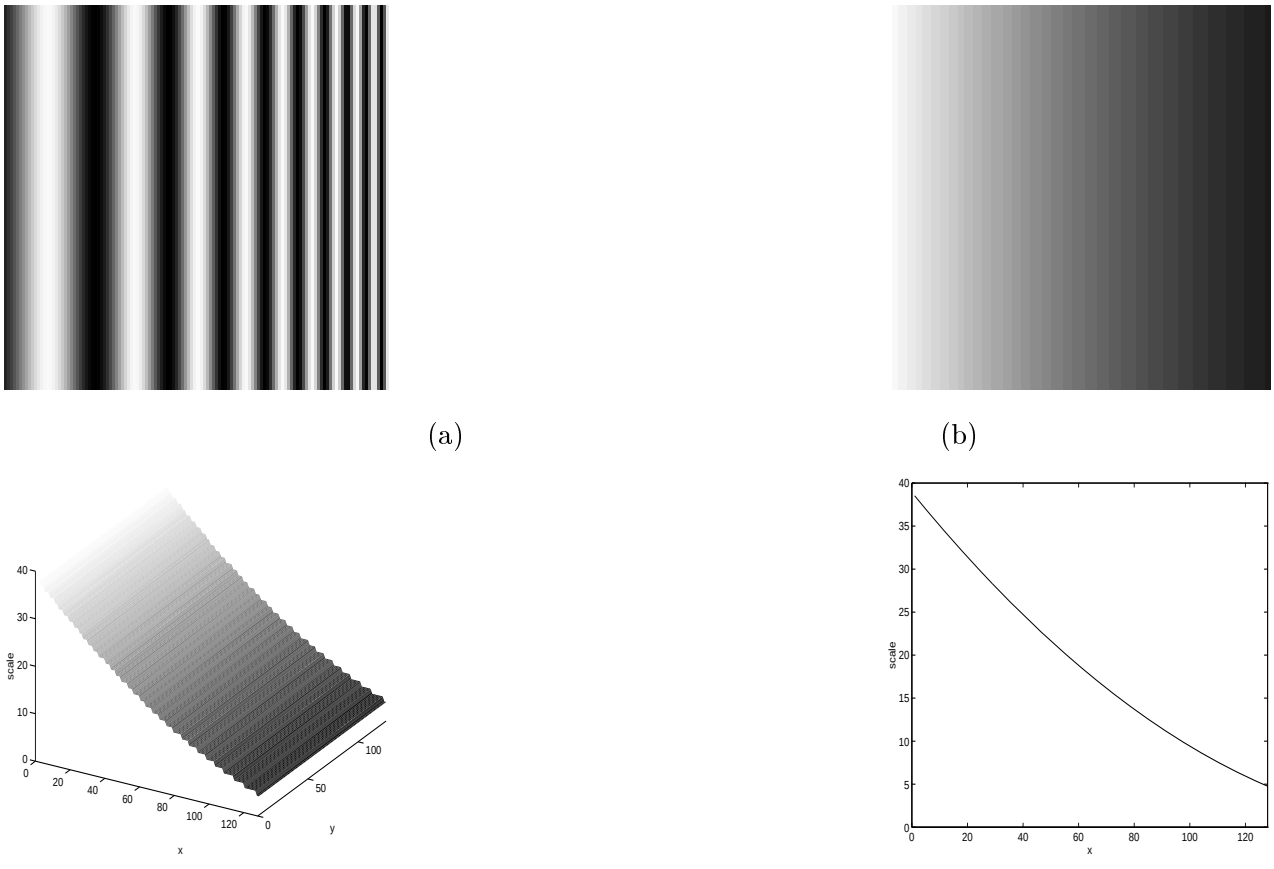


Figure 4: The coordinate relationship between the image plane and the surface plane.



(a) (b)

Figure 5: A hybrid image and its ridges: (a) The hybrid image of a slant cosine and a frontal straw cloth; (b) The two detected ridges, where only the top ridge is useful for surface orientations.

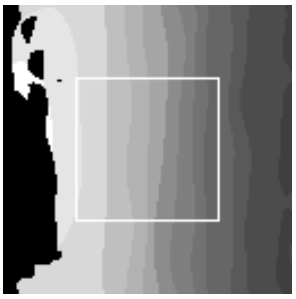


(a) (b)

(c) (d)

Figure 6: Slanted cosine image and 1-D, 2-D, and 3-D plots of its ridge: (a) The slanted image; (b) 2-D scale map of the ridge; (c) 3-D scale map of the ridge; (d) 1-D scale map of the ridge.

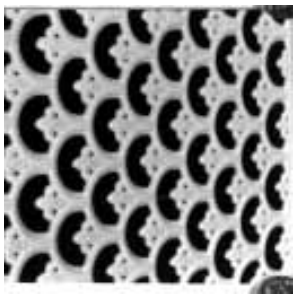
(a1) Brick wall



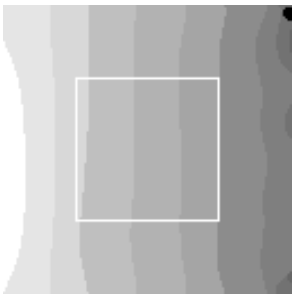
(a2)

(b1) Grillwork-1

(b2)

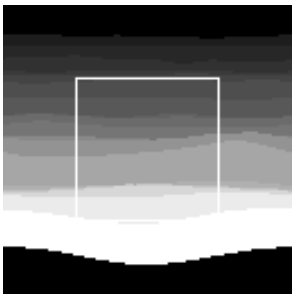


(c1) Grillwork-2

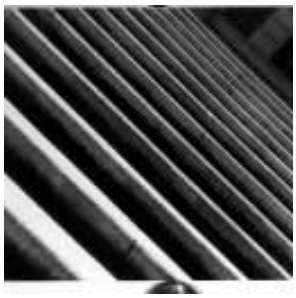


(c2)

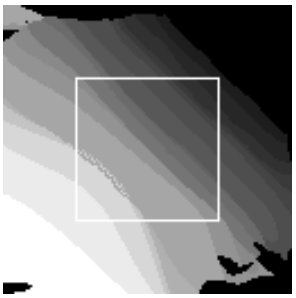
(d1) Steps-1



(d2)

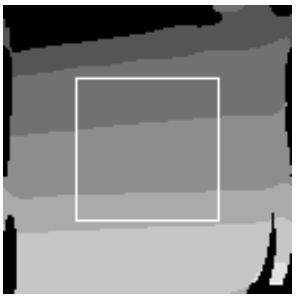


(e1) Steps-2



(e2)

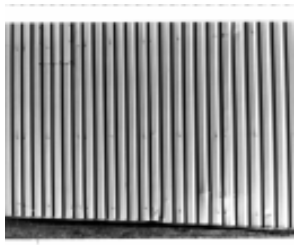
(f1) Venetian blind



(f2)

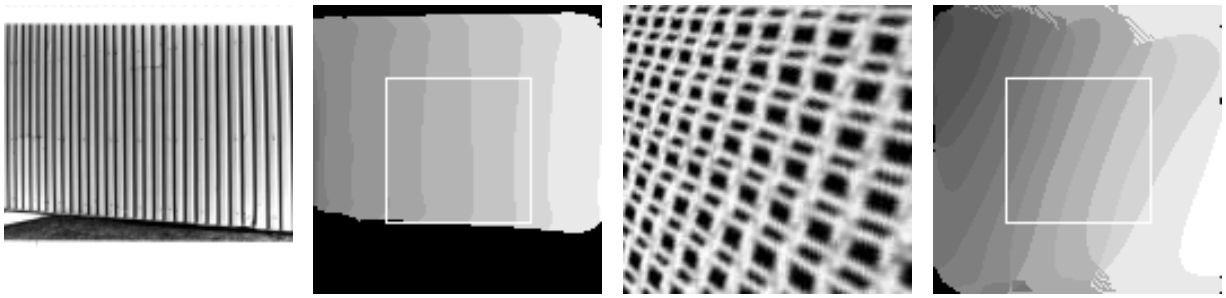
(g1) Ventilator

(g2)

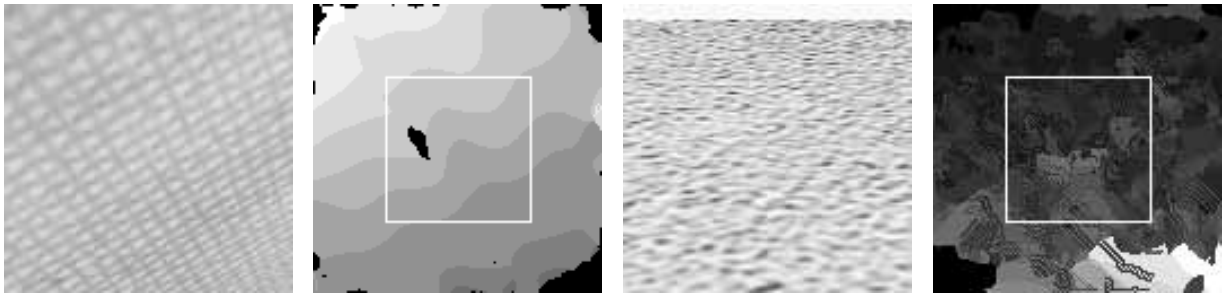


(h1) Aluminum wall-1

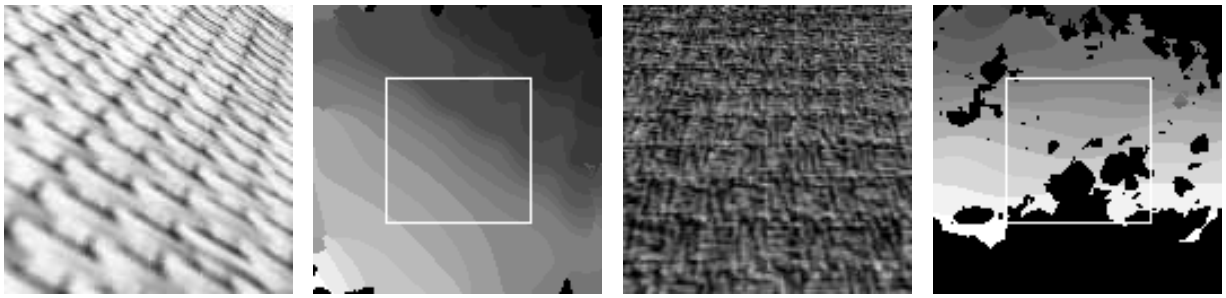
(h2)



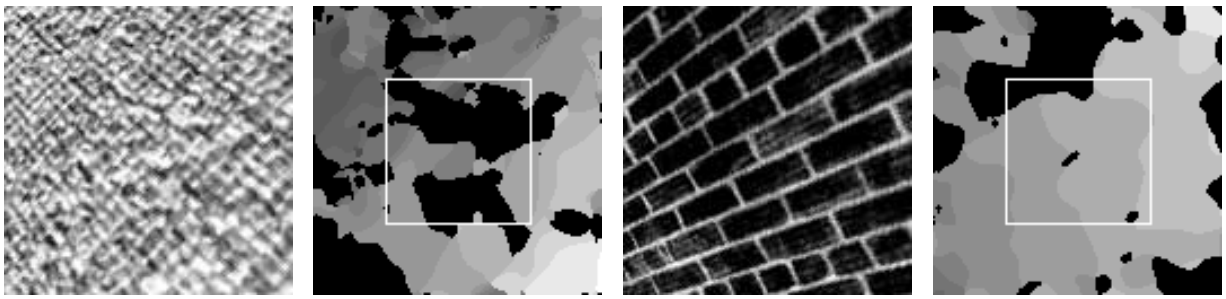
(i1) Aluminum wall-2 (i2) (j1) D20 (j2)



(k1) D52 (k2) (l1) D57 (l2)



(m1) D65 (m2) (n1) D82 (n2)



(o1) D84 (o2) (p1) D95 (p2)

Figure 7: Experiemtal images: (a1)-(i1) are Super and Bovik's images, (a2)-(i2) are the ridges of (a1)-(i1); (j1)-(p1) are Brodatz's images, (j2)-(p2) are the ridges of (j1)-(p1).

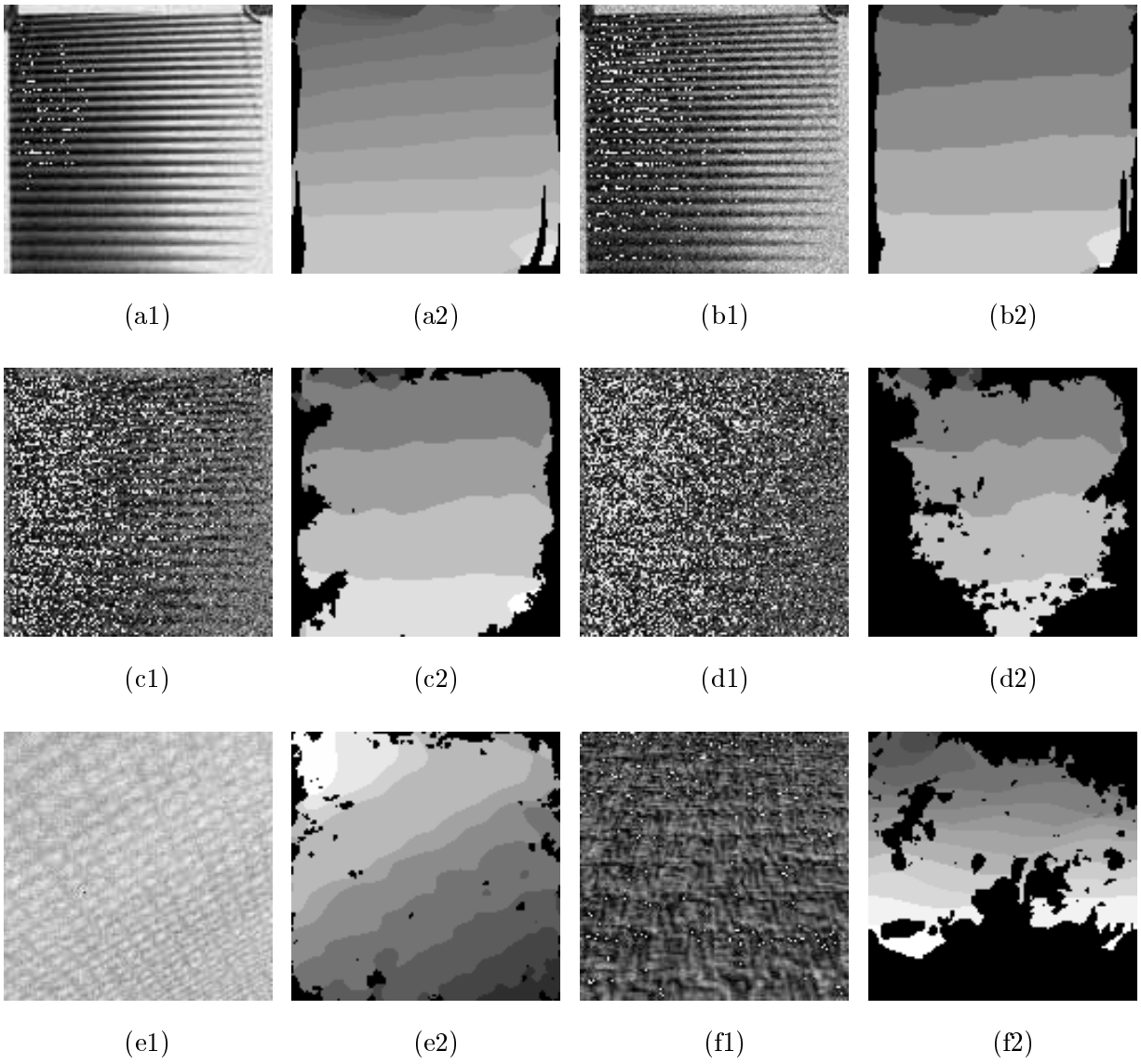


Figure 8: Noisy images and their ridges: Noisy Venetian blind with SNR 20, 10, 0, and -5 db are shown in (a1), (b1), (c1), and (d1); their detected ridges are shown in (a2), (b2), (c2), and (d2), respectively; (e1) D52 with SNR 0 db, (e2) Ridge of (e1); (f1) D82 with SNR 10 db, (f2) Ridge of (f1).

Table 1: **Tilt results of our method in the set of Super and Bovik's images.**

Image	True tilt τ	τ_m	$\mid \tau_m - \tau \mid$	τ_{SB}	$\mid \tau_{SB} - \tau \mid$
Brick wall	0	-1.61	1.61	1.1	1.1
Grillwork-1	-90.0	-91.67	1.67	-99.0	9.0
Grillwork-2	0.0	-1.16	1.16	-15.3	15.3
Steps-1	90	91.45	1.45	88.6	1.40
Steps-2	45	46.02	1.02	45.10	0.10
Venetian blind	90	92.98	2.98	90.9	0.9
Ventilator	90	89.26	0.74	82.7	7.3
Aluminum wall-1	180	-178.05	1.95	178.9	1.1
Aluminum wall-2	180	179.51	0.49	179.5	0.50

τ_m: tilt angle of our method

τ_{SB}: tilt angle of Super and Bovik's method

Table 2: **Slant results of our method in the set of Super and Bovik's images.**

Image	True slant ρ	ρ_v	ρ_c	$\mid \rho_v - \rho \mid$	$\mid \rho_c - \rho \mid$	ρ_{SB}	$\mid \rho_{SB} - \rho \mid$
Brick wall	55	54.80	53.34	0.20	1.66	52.3	2.7
Grillwork-1	10	9.66	10.79	0.34	0.79	11.3	1.3
Grillwork-2	30	25.70	25.13	4.30	4.87	30.3	0.3
Steps-1	70	68.70	68.61	1.30	1.39	68.0	2.0
Steps-2	70	68.80	68.05	1.20	1.95	71.0	1.0
Venetian blind	40	40.83	41.81	0.83	1.81	32.1	7.9
Ventilator	20	16.18	16.28	3.82	3.72	15.2	4.8
Aluminum wall-1	20	17.96	19.47	2.04	0.53	16.7	3.3
Aluminum wall-2	30	24.70	25.00	5.30	5.00	26.4	3.6

ρ_v: slant angle of our voting method

ρ_c: slant angle of our curve fitting method

ρ_{SB}: slant angle of Super and Bovik's method

Table 3: **Estimated surface orientations by our method in the set of Brodatz's images.**

Image	$True\tau$	τ_m	$\mid \tau_m - \tau \mid$	$True\rho$	ρ_v	ρ_c	$\mid \rho_v - \rho \mid$	$\mid \rho_c - \rho \mid$
D20	160	160.05	0.05	37	37.53	35.97	0.53	1.03
D52	-60	-61.08	1.08	50	49.58	48.89	0.42	1.11
D57	90	91.41	1.41	70	67.30	65.09	2.70	4.91
D65	60	55.00	5.00	50	54.38	52.62	4.38	2.62
D82	90	90.71	0.71	50	48.62	46.74	1.38	3.26
D84	135	128.58	6.42	35	33.26	31.58	1.74	3.42
D95	-155	-158.57	3.57	27	28.45	27.58	1.45	0.58

τ_m: tilt angle of our method

ρ_v: slant angle of our voting method

ρ_c: slant angle of our curve fitting method

Table 4: **Estimated surface orientations(tilt/slant of voting, slant of curve fitting) by our method for (true tilt/true slant) of noisy images.**

db	Venetian blind 90/40	Aluminum wall-1 180/20	D52 −60/50	D82 90/50
∞	93.0/(40.8,41.8)	-178.1/(18.0,19.5)	-61.1/(49.6,48.9)	90.7/(48.6,46.7)
20	94.1/(34.5,34.2)	-178.3/(18.0,19.2)	-61.3/(49.2,48.5)	90.7/(48.6,47.0)
10	94.3/(34.8,34.3)	-178.4/(17.8,19.1)	-61.3/(49.1,48.6)	90.7/(48.7,47.3)
0	94.6/(35.2,34.4)	-178.5/(17.8,19.0)	-61.2/(49.2,48.9)	X
−5	95.3/(35.2,34.2)	-176.0/(8.4,18.0)	-72.5/(81.6,58.0)	X