應用MAXIMUM-LIKELIHOOD分群法於三維紋路分割:計算表面方向的第一步*

MAXIMUM LIKELIHOOD CLUSTERING FOR 3D TEXTURE SEGMENTATION: A FIRST STEP OF SHAPE FROM TEXTURE

呂俊賢†, 黃文良‡, 詹寶珠†

† 成功大學電機所
‡ 中央研究院資訊所

摘要

先前計算紋路影像之表面方向的方法大都有一共通的假設,那就是影像主要只包含一種紋路.事實上,我們所看見的影像經常出現兩種以上的紋路.爲了更實際解決表面方向計算的問題,分割三維紋路影像是不可缺少的第一步.我們以連續小波轉換之脊平面來描述透射投影模式下紋路影像支配頻率的變化.脊平面可更進一步推導爲二次多項式.一個階層式的agglomerative分群法被用來合併小區塊.合併之條件由maximum-likelihood函數決定.接著,我們結合不同脊平面所分割出的結果,以完成整個分割.

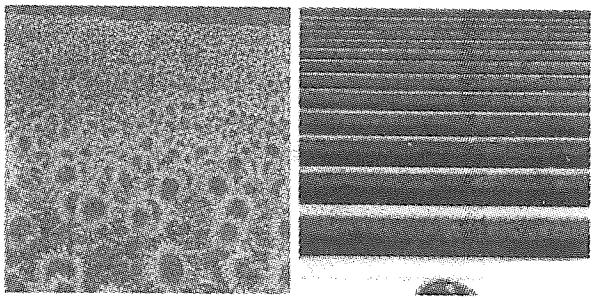
關鍵字: 三維紋路分割　連續小波轉換　脊平面

ABSTRACT

A common assumption of the shape from texture problem is that a perceived image mainly contains only one type of texture with the same surface orientation. Unfortunatelly, a natural image is often composed of more than one textures. In order to solve the shape from texture problem in a practical manner, we need to segment 3D textured images firstly. The dominate frequency variations of projected images are characterized by the ridge surface of continuous wavelet transform. We model the ridge surface as the quadratic polynomial plus white noise. Agglomerative clustering is used for merging similar blocks according to maximum likelihood joint probability. A fusion technique is applied to accomplish the segmentation work. Textured images synthesized from Brodatz's album and several natural images demonstrate the performance of our method.

Keywords: 3D texture segmentation continuous wavelet transform ridge surface maximum-likelihood clustering

*This work is supported by National Science Council under the Grant NSC 87-2213-E-001-027



(a)　(b)

Figure 1: One main texure in the projected images: (a). Sunflowers [2]; (b). Steps [13].

1. INTRODUCTION

Shape from texture is an important problem in recovering 3D scene information from a single image by using texture cue only [2][6][7][9][13]. In the literture, almost all the shape from texture methods assume that the projected image is mainly composed of only one texture, as shown in Fig. 1. Surface orientation is therefore estimated from this texture. However, there are often more than one textures appearing in our perceived images (see an example shown in Fig. 5). The important problem, 3D texture segmentation, is avoided and rarely discussed for previous methods except for [7][11]. In order to solve the shape from texture problem in a practical manner, the 3D texture image segmentation must be taken into account. 3D textured image segmentation differs from its 2-D counterpart in that the features commonly used in 2-D texture segmentation is no longer suitable, the projection effects should be considered. Krumm and Shafer [7] had firstly presented a 3D textured image segmentation approach. They estimated the local frontal frequencies by applying their shape from texture approach to local patches. The local frontal frequencies were used later as the

features for 3D texture segmentation. However, the accuracy of their algorithm depends on the number of patches and their sizes. In addition, their segmentation results have the blocky effects.

We had also previously presented a 3D texture segmentation approach based on a robust fitting technique to extract reliable local surface orientations as the local features [11]. The local surface orientations are estimated based on our shape from texture method [6][9]. Dominate frequency variations of projected images are efficiently characterized by the ridge surface of continuous wavelet transform (CWT). The continuous wavelet transform tuned to various scales and rotations is particular suitable for local frequency analysis. With the scale parameter, we obtain the degrees of frequency variations; with the rotation parameter, we know at which direction the maximal frequency variations is. This is crucial for 3D texture segmentation in capturing texture variations resulted from projection effects. The ridge surface of a continuous wavelet transform marks the places in the spatial frequency domain where the energy is mostly concentrated [5]. The scales of a ridge surface is derived to be a parabolic function of the slant and the tilt angles. The surface orientation is solved by fitting the scale of the ridge surface as a parabola [6]. According to the local surface orientations, a robust clustering was used for coarse segmentation.

It should be emphasized that the above 3D segmentation algorithms either need to estimate the local frequencies [7] or local surface orientations [11] as the features. Owing to the ridge surface had been derived to be a parabola, it is reasonable for us to model the ridge surface as a quadratic polynomial plus white noise. This model is the one adopted by Silverman *et al.* [12] and Lavalle *et al.* [8]. However, their model is based on the assumption that the image or the smooth 3D surface region can be approximately represented by either linear or quadratic polynomials. No derivation for such types of polynomials is given. According to our parabolic model of ridge surface, a ridge surface is initially divided into several blocks. A maximum likelihood clustering is proposed to accomplish the coarse segmentation. A best pair of blocks to be merged or not is determined by a agglomertive clustering iteratively. Finally, all the coarsely segmented results obtained from the multiple ridge surfaces are then fused to get complete segmentation result.

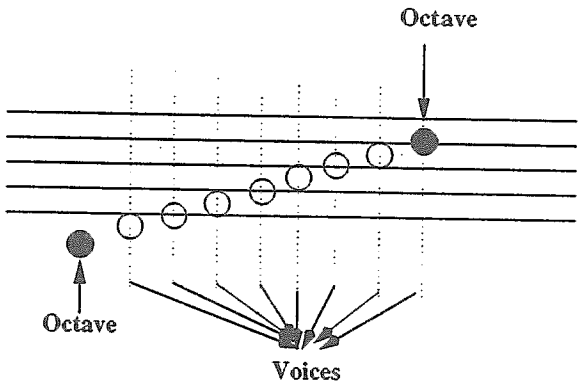


Figure 2: The relationship of octave and voices follows the music.

2. CONTINUOUS WAVELET TRANSFORM AND RIDGE SURFACE

2.1. Continuous Wavelet Transform

A complex-valued function $\psi(\underline{x})$ in $L^2(\mathcal{R}^2)$ is a wavelet if $\int_{\mathcal{R}^2}\psi(\underline{x})d\underline{x}=0$. Let $\psi_{(\mathbf{b},\,s,\,\theta)}(\underline{x})$ be obtained by the translation, scaling, and rotation of $\psi(\underline{x})$: $\psi_{(\mathbf{b},\,s,\,\theta)}(\underline{x}) = \frac{1}{s^2}\psi(\mathbf{r}_{-\theta}\frac{(\underline{x}-\mathbf{b})}{s})$, where $\mathbf{b}\in\mathcal{R}^2$, $s>0$, and $\theta\in[0,2\pi)$ are the translation, scaling, and rotation parameters, respectively; and $\mathbf{r}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$ is the rotation matrix of angle θ [1]. In our implementation, we use the $2D$ Morlet wavelet, defined as $\psi_{\mathrm{M}}(\underline{x}) = e^{j\mathbf{k_0}^T\underline{x}}e^{-|\underline{x}|^2/2}$ in the spatial domain, which is $\hat{\psi}_{\mathrm{M}}(\mathbf{w}) = e^{-|\mathbf{w}-\mathbf{k_0}|^2/2}$ in the frequency domain, where $\mathbf{k_0}$ is the center frequency of Morlet wavelet. Also, following the conventional usage [5], the scale parameter s takes the discretized values, $s=2^{o+\frac{v}{n}}$, where o is the octave, v is the voice, and n is the number of voices per octave. The relationship of octave and voices is shown in Fig. 2.

Fig. 3 shows the functional range of the Morlet wavelet according to its tuning rotations and scales. The more voices and rotations are used, the denser the wavelet covers the frequency plane. The Morlet wavelet optimizes both the spatial resolution and the frequency resolution simultaneously; therefore, it is well adapted for local frequency analysis.

We adopt an idealized or monochromatic texture model [3] to represent an image $f(\underline{x})$ with N components:

$$f(\underline{x}) = \sum_{k=1}^{N} A_k \cos(\Omega_k^T\underline{x} + p_k), \qquad (1)$$

where amplitude A_k, frequency Ω_k, and phase p_k are assumed to be constants. By taking the wavelet

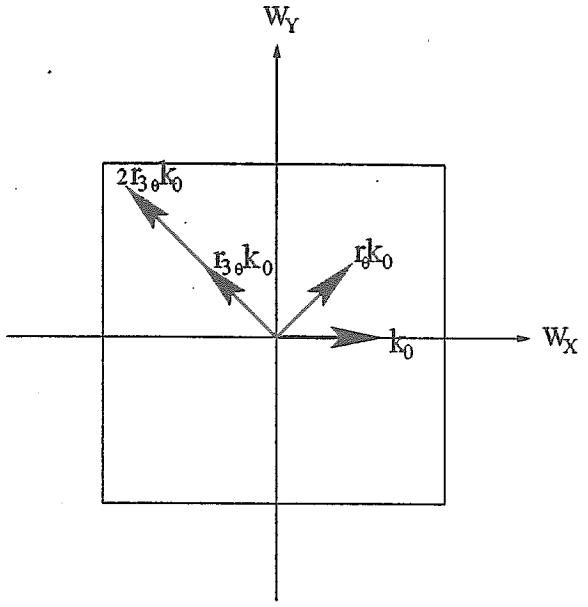


Figure 3: The tunning of CWT in terms of rotations and scales.

transform of $f(\underline{x})$ in the frequency domain, we obtain

$$(\mathcal{W}f)(\mathbf{b},s,\theta)=\sum_{k=1}^{N}\frac{A_k}{2}\hat{\psi}_M(s\mathbf{r}_{-\theta}\Omega_k)e^{-j(\Omega_k^T\mathbf{b}+p_k)}.$$

One can then read off from these points (called ridge points hereafter) important local parameters about the spatial frequency Ω_k. The ridge points at $\mathbf{b}$ can be extracted by selecting the squared-modulus local maxima among θ and s at $\mathbf{b}$. Let $\mathcal{N}(\underline{x})$ contain the neighborhood of argument x, including x; then $(\mathbf{b},s_0,\theta_0)$ is selected if

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

The detail of our ridge surface detection algorithm is described in [6].

2.2. Ridge Surface: A Parabolic Function

Our viewing geometry is adopted as the same as Super and Bovik [13]. 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 those depicted in Fig. 4. 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, onto the image plane is defined as the tilt angle τ, which takes values between $-180°$ and 180°. The slant-tilt combination represents the surface orientation of a planar texture.

The relationship between the coordinate systems of a surface plane and the image plane under the

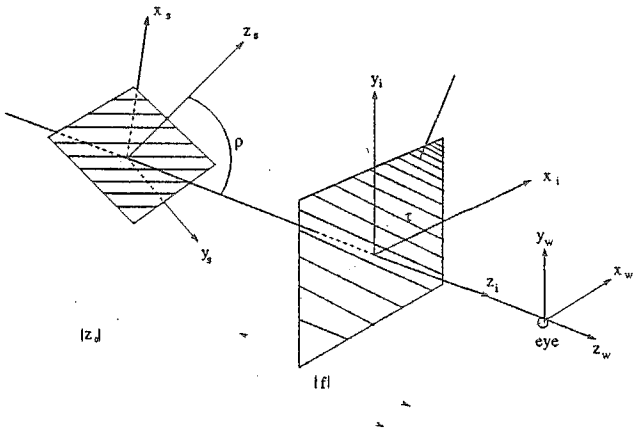


Figure 4: The coordinate relationship between the image plane and the surface plane.

perspective projection model was derived in [13]:

$$\begin{bmatrix}x_s\\y_s\end{bmatrix}=\frac{z_w}{f}\begin{bmatrix}\sec\rho & 0\\0 & 1\end{bmatrix}\begin{bmatrix}\cos\tau & \sin\tau\\-\sin\tau & \cos\tau\end{bmatrix}\begin{bmatrix}x_i\\y_i\end{bmatrix},\tag{2}$$

where

$$\frac{z_w}{f}=\frac{z_0}{\tan\rho(x_i\cos\tau+y_i\sin\tau)+f}.\tag{3}$$

Assume that the tilt angle τ has been derived; hence, the rotation matrix $\mathbf{r}_\tau$ is known. 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 the surface plane, respectively. Then, the new coordinate of the image plane, $\mathbf{x}$, is obtained by applying the rotation matrix $\mathbf{r}_\tau$ to $\mathbf{x}_i$. Substituting $\mathbf{x}_s$ in Eq. (2) into Eq. (1) for $\underline{x}$, we obtain the projected texture in the image plane:

$$g(\mathbf{x})=\sum_{k=1}^{N}A_k\cos(\Omega_k^T\frac{z_w}{f}\begin{bmatrix}\sec\rho & 0\\0 & 1\end{bmatrix}\mathbf{x}+p_k).\tag{4}$$

Let the spatial frequencies of the surface texture be $\Omega_k=[u_k\ v_k]$; we have

$$g(\mathbf{x})=\sum_{k=1}^{N}A_k\cos(\phi_k(\mathbf{x}))\approx\sum_{k=1}^{N}A_k\cos(\nabla\phi_k(\mathbf{x})^T\mathbf{x}+p_k).\tag{5}$$

Thus, the non-linear spatial frequency in the projected image caused from the perspective projection is approximated as the gradient of the phase $\phi_k(\mathbf{x})$. After substituting $\frac{z_w}{f}$ from Eq. (3) into Eq. (5), the local spatial frequency at $\mathbf{x}$ is represented in terms of the slant angle ρ :

$$\begin{aligned}\left[\frac{\partial\phi_k(\mathbf{x})}{\partial x}\ \ \frac{\partial\phi_k(\mathbf{x})}{\partial y}\right]&=\left[\frac{z_0(u_kf\sec\rho-v_ky\tan\rho)}{(x\tan\rho+f)^2}\ \ \frac{z_0v_k}{x\tan\rho+f}\right]\\&=\left[\frac{z_0u_kf\sec\rho}{(x\tan\rho+f)^2}+y\frac{\partial^2\phi_k(\mathbf{x})}{\partial x\partial y}\ \ \frac{\partial\phi_k(\mathbf{x})}{\partial y}\right],\end{aligned}\tag{6}$$

where $\frac{\partial\phi_k(x)}{\partial x}$ and $\frac{\partial\phi_k(x)}{\partial y}$ correspond to the frequencies in the tilt and perpendicular to the tilt, respectively.

Let $g_S(x)$ be the image reconstructed by restricting the wavelet transform to S. It is worth mentioning that $g_S(x)$ may be significantly different from our textured image since surface properties in the textured image could not be completely captured by the ridge points in S. To recapitulate what we have derived from Eqs. (5) and (6), $g_S(x)$ is approximated by

$$g_S(x) \approx A_S \cos(\nabla\phi_S(x)^T x + p), \qquad (7)$$

where p is the phase at coordinate origin, and

$$\left[\frac{\partial\phi_S(x)}{\partial x} \quad \frac{\partial\phi_S(x)}{\partial y}\right] = \left[\frac{z_0 u_S f \sec\rho}{(x\tan\rho + f)^2} + y\frac{\partial^2\phi_S(x)}{\partial x\partial y} \quad \frac{\partial\phi_S(x)}{\partial y}\right]. \qquad (8)$$

Since the variations of $\frac{\partial\phi_S(x)}{\partial x}$ along the direction perpendicular to the tilt, $\frac{\partial^2\phi_S(x)}{\partial x\partial y}$, are relatively less significant, this term can be ignored. In this case, it can be checked that

$$\frac{\partial\phi_S(x)}{\partial x} \approx (\frac{z_0}{f})\frac{u_S\sec\rho}{(1+\frac{x\tan\rho}{f})^2}. \qquad (9)$$

Since the dominant spatial frequencies of a textured image are characterized by ridge points, it is possible to relate $\frac{\partial\phi_S(x)}{\partial x}$ to the contents of ridge points $(x, s(x), \theta(x))$ in ridge surface S. To recall the content of ridge point $(x, s(x), \theta(x))$: the magnitude of the spatial frequency at x is inversely proportional to scale $s(x)$, and $\theta(x)$ gives the direction in which the frequency points. Hence, if we choose k_0 at $(k_0 = ||k_0||, 0)$, then the scale value $s(x)$ of the ridge point $(x, s(x), \theta(x))$ in the ridge surface S is given by

$$s(x) = \frac{fk_0(1+\frac{x\tan\rho}{f})^2}{z_0 u_S\sec\rho}. \qquad (10)$$

It is clear that the scale $s(x)$ of the ridge surface is a parabolic function of x, independent of y.

3. 3D TEXTURED IMAGE SEGMENTATION ALGORITHM

Our segmentation algorithm is directly performed on the ridge surfaces of a textured image. The scale values of a ridge surface is derived to be a parabola in Eq. (10). It is thus reasonable for us to model the ridge surface as a parabolic polynomial plus zero-mean white Gaussian noise of constant variance. This modeling is similar to the data-generation model by using linear polynomial or quadratic polynomial for an image [8][12]. However, the linear or quadratic polynomial model for an image is purely an assumption. Their approaches will fail if high-order polynomial is needed for modeling.

3.1. Coarse Segmentation by Maximum Likelihood Clustering

We first divide each ridge surface into blocks and utilize a region-based clustering to accomplish a 3D segmentation problem. Agglomerative clustering is adopted to achieve coarse segmentation iteratively. During each iteration, a pair of blocks should be determined to be merged or not. The merge criterion is modeled as the maximum likelihood joint probability on the ridge surface. At the end of agglomerative clustering, the number of clusters is generated if no merge is necessary.

The joint likelihood of a ridge surface S given K clusters of the ridge surface and the model coefficients of the parabolic functions corresponding to their clusters, and the noise variances in these clusters, is defined as

$$p(S|C_1, C_2, ..., C_K, \alpha_1, \alpha_2, ..., \alpha_K, \sigma_1, \sigma_2, ..., \sigma_K)$$

$$= \prod_{k=1}^{K}\prod_{(i,j)\in C_k}(2\pi\sigma_k^2)^{\frac{-1}{2}}\cdot exp[\frac{-1}{2}(\frac{s(i,j)-f_{\alpha_k}(i,j)}{\sigma_k})^2]$$

$$= \prod_{k=1}^{K}(2\pi\sigma_k^2)^{\frac{-N(C_k)}{2}}\cdot exp[\frac{-1}{2}\prod_{(i,j)\in C_k}(\frac{s(i,j)-f_{\alpha_k}(i,j)}{\sigma_k})^2], \qquad (11)$$

where α_k and σ_k denote the coefficients of fitted parabolic function and the noise variance, respectively, for the cluster C_k with the number of ridge points $N(C_k)$. $s(i,j)$ is the scale value of the ridge point at the location (i,j) whereas $f_{\alpha_k}(i,j)$ is the fitted value of the parabolic function at (i,j). Fit the scale values of a ridge surface to a parabola is simply a least-squares problem. To simplify analysis, the variances σ_k of noises for all block C_k $(1 \le k \le K)$ are assumed to a equal constant σ.

According to the maximum likelihood clustering model, the best pair of blocks to be merged is determined if Eq. (11) is maximized. Let C_m and C_n be two blocks, and S_{C_m,C_n} be the partial set of ridge surface S containing C_m and C_n. The joint likelihood of the ridge surface over S_{C_m,C_n} given the estimated parameters $\alpha_{m,n}$ is defined as

$$p(S_{C_m,C_n}|\alpha_{m,n}) = \prod_{(i,j)\in S_{C_m,C_n}}(2\pi\sigma^2)^{\frac{-1}{2}}\cdot exp[\frac{-1}{2}(\frac{s(i,j)-f_{\alpha_{m,n}}(i,j)}{\sigma})^2]. \qquad (12)$$

The pair of blocks, C_m and C_n (for $1 \le m, n \le K$), maximizes the joint-likelihood probability in Eq. (12) will be selected as the best pair of blocks to be merged. After the maximum likelihood clustering, a coarse segmentation is obtained for each ridge surface.

3.2. Fusion

A textured image after CWT may be characterized with more than one ridge surfaces. The coarsely segmented results of all the ridge surfaces should be combined to get a complete segmentation result. This combination help us to generate new clusters by fusing the features of different ridge surfaces and to delete fragmentations if a cluster's area is small enough. The Lu *et al.*'s fusion method [10] had been successfully used in 2D texture segmentation to fuse segmentation results come from differnet high-frequency channels of discrete wavelet transform. Therefore, this fusion technique is employed for the similar purpose about the clustering problem in 3D texture segmentation. The fusion step is a re-labeling process by updating the labels of pixels of two to be fused coarsely segmented results. The only parameter needed is the size of a cluster considered to be not a "true" cluster. In the 3D texture segmentation, it is easy to decide since a region having apparent projection effects usually occupy a large enough area.

3.3. Fine Segmentation

After the coarse segmentation and fusion steps, the "true" number of clusters is obtained. Each cluster is then fitted to a parabolic function since it corresponds to a ridge surface. For every un-classified pixel, it is assigned to be a specific cluster if its scale values is most close to the fitted parabola corresponging to the specific cluster. If a pixel does not belong to any ridge surface, it can not be assigned to any clusters. In this case, a connected component technique is then used to classify these hole regions according to local spatial relations.

4. EXPERIMENTAL RESULTS

Some natural images synthesized from Brodatz's album [4] and come from MIT VisTex database are used to demonstrate the performance of our method. Fig. 5(a) is an image composed of two Brodatz's textures, $D101$ and $D102$, inclined to zero tilt angle and different slant angles. Figs. 5(b) and (c) are real-world textured images with one main textured region and the background. Fig. 5(d) is also an real-world image but with three different textures: windows, a brick wall, and a pavement. The window and the brick wall regions have the same surface orientations but have the different frequencies, where the pavement has the surface orientation different from the other two regions. These segmentation results are shown in Fig. 5 with boundaries superimposed on original textured images. It is found that our algorithm correctly separate different textures. For Figs. 5(b)-(c), the buildings and the ladders are successfully extracted, whereas the other parts with less perspective effects are regarded as backgrounds. Experimental results demonstrate the ability of our method in segmenting 3D textured images.

5. CONCLUSIONS

The ridge surfaces of continuous wavelet transform are shown to be powerful in characterizing texure variations. It is reasonable to represent the ridge surfaces as quadratic polynomial since a ridge surface had been derived to be a parabola. By utilizing this merit, the clustering process is modeled as the maximum likelihood clustering by using agglomerative clustering to merge similar blocks. To notice that the proposed method does not need to estimate local (original) frequencies or local surface orientations as the local features. Start with 3D texture segmentation, solving the shape from texture problem becomes practical.

For some textured images with less regular property, this method may fail to accurately determin the true number of clusters. This is due to the ridge surface can not be efficiently used to describe the dominate frequency variations of projected images. This problem can be sovled to some extent by a robust fitting technique [11] to select the best subset of ridge points to represent the dominate frequency variations.

6. REFERENCES

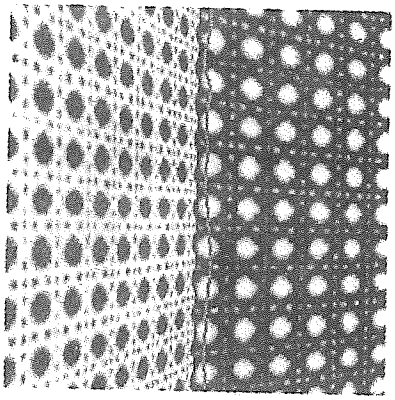
[1] J. -P. Antoine, P. Carrette, R. Murenzi, and B. Piette, "Image Analysis with Two-Dimensional Continuous Wavelet Transform", *Signal Processing*, Vol. 31, 1993, pp. 241-272.

[2] D. Blostein and N. Ahuja, "Shape from Texture: Integrating Texture-Element Extraction and Surface Estimation ", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 12, 1989, pp. 1233- 1251.

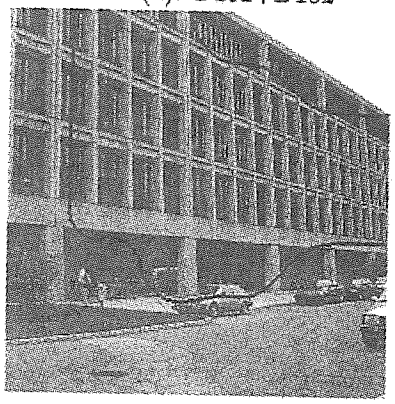
[3] A. C. Bovik, "Analysis of multichannel narrow-band filters for image texture segmentation", *IEEE Trans. Signal Processing*, Vol. 39, 1991, pp. 2025-2043.

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

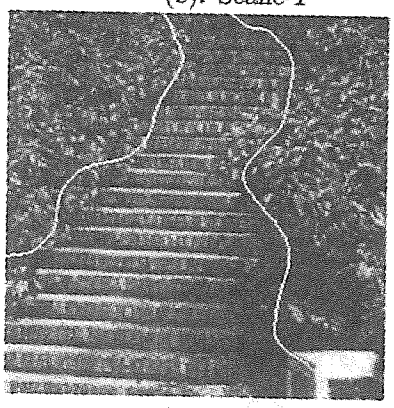
[5] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and Gabor



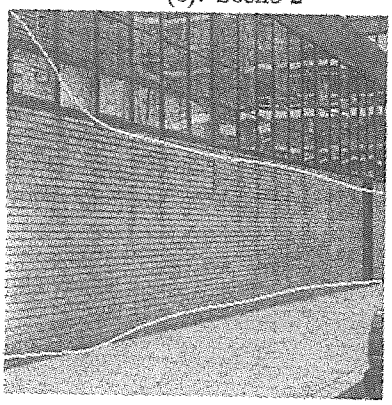
(a). D101+D102



(b). Scane-1



(c). Scene-2



(d). Scane-3

Figure 5: Segmentations with boundaries superimposed.

analysis: Extraction of instantaneous frequencies", *IEEE Trans. Inform. Theory*, Vol. 38, No. 2, 1992, pp. 644-664.

[6] W. L. Hwang, C. S. Lu and P. C. Chung, "Estimation of Planar Surface Orientation Through the Ridge Surface of Continuous Wavelet Transform", *To appear in IEEE Trans. Image processing*, 1997.

[7] J. Krumm and S. Shafer, "Segmenting textured 3D surfaces using the space/frequency representation", *Spatial Vision*, Vol. 8, No. 2, 1994, pp. 281-308.

[8] S. M. LaValle and S. A. Hutchinson, "A Bayesian Segmentation Methodology for Parametric Image Models", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 17, No. 2, 1995, pp. 211-217.

[9] 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 Conf. Image processing*, Vol. I, 1996, pp. 295-298.

[10] C. S. Lu, P. C. Chung and C. F. Chen, "Unsupervised Texture Segmentation Via Wavelet Transform", *Pattern Recognition*, Vol. 5, pp. 729-742, 1997.

[11] C. S. Lu, W. L. Hwang, and P. C. Chung, "Segmentation of 3D Textured Image Using Continuous Wavelet Transform, *To appear in IEEE Conf. Image Processing*, 1997.

[12] J. F. Silverman and D. B. Cooper, "Bayesian Clustering for Unsupervised Estimation of Surface and Texture Models", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 10, 1988, pp. 482-495.x

[13] B. Super and A. C. Bovik, "Planar Surface Orientation from Texture Spatial Frequencies", *Pattern Recognition*, Vol. 28, No. 5, 1995, pp. 729-743.