Local Spatial Frequency Analysis of Image Texture

John Krumm and Steven A. Shafer *
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Real-world scenes contain many interacting phenomena that lead to complex images which are difficult to interpret automatically. Part of this difficulty is due to the dichotomy of useful representations for these phenomena. Some effects are best described in the spatial domain, while others are more naturally expressed in frequency. In order to resolve this dichotomy, we present the combined space/frequency representation which, for each point in an image, shows the spatial frequencies at that point. This representation is useful for developing theories about many important vision phenomena, leading to deeper understanding and better algorithms. In this paper, we show how the representation can be used for the shape from texture problem and to analyze aliasing simply and naturally. The space/frequency representation should be a key aid in untangling the complex interaction of phenomena in images, allowing automatic understanding of real-world scenes.

1 Introduction

Vision can provide a uniquely rich source of information for robots operating in unstructured environments. This is evident from the variety of methods used to analyze images – each exploiting one kind of phenomena, *e.g.* shading, texture, stereo, defocus. However, this very richness greatly complicates the task of interpreting images of real-world scenes, because the exploited effect is often disturbed by other phenomena. For instance, in segmenting 2D textures, 3D shape effects can confound an algorithm that tries to find regions of uniform texture.

We attribute this problem partly to the dichotomy of representations used in computer vision. Effects like perspective projection, shading phenomena, and overall shape are best described in geometric (spatial) terms. The frequency domain (or some close variant) serves best for texture, sampling, and many optical effects. Combinations of effects from different preferred representations usually lead to complicated, unreliable algorithms.

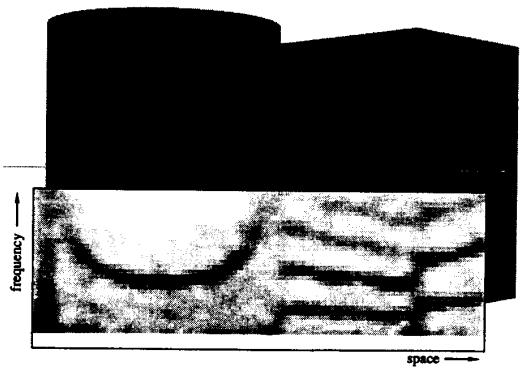

Figure 1: Textured cylinder and cube with spectrogram of center scan-line

In order to solve this problem, we present the combined *space/frequency* representation which ideally shows the spatial frequencies present at each point in a signal. The space/frequency representation of a 1D signal is a 2D function of space and frequency. Each point in the signal has associated with it a 1D frequency profile. For a 2D signal, the representation is a 4D function of two spatial variables and two frequency variables. In a discrete image, it is as if every pixel has associated with it a 2D Fourier transform. In this paper we limit ourselves to the analysis of 1D signals, mostly for the simplicity of displaying the corresponding 2D space/frequency representation. The concepts and algorithms can be easily generalized to images, however.

An example of the space/frequency representation of a signal is shown in Figure 1. The box in the lower half of the image shows the space/frequency representation (the spectrogram) of the image's center scan-line. Because the underlying patterns on the two objects are periodic, there are dark frequency peaks in the spectrogram where the objects occur. The large, "U"-shaped frequency peak on the left shows that the frequency of the texture pattern projected from the cylinder appears higher near the edges than in the middle, as one would expect. At the extreme edges of the cylinder, the projected frequency is so high it cannot be adequately reproduced in the image. This is shown in the spectrogram as the frequency peak bumping into the Nyquist frequency at the top. On the left side of the cube, we see a slowly decreasing fundamental frequency

*This research was supported by the Defense Advanced Research Projects Agency, DoD, through ARPA Order Number 4976, monitored by the Air Force Avionics Laboratory under Contract F33615-87-C-1499 and by the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration under Contract 957989 and by the National Aeronautics and Space Administration under the Graduate Student Researcher's Program, Goddard Space Flight Center, for the first author, grant NGT-50423. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the United States Government or the Jet Propulsion Laboratory.

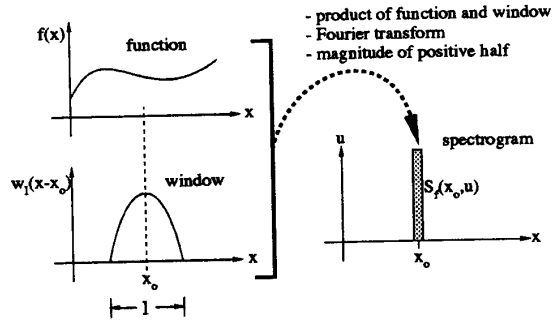

Figure 2: Computing the spectrogram

and overtones which are likewise decreasing. This decrease continues to the corner of the cube, where the fundamental and harmonics begin to increase as the side recedes into the distance. This is a sample of the kind of analysis possible with the spectrogram.

There are many methods of calculating the space/frequency representation, none of which achieve the ideal. We have chosen the spectrogram as our method, because of its simplicity and proven utility. It was first used for computer vision by Bajcsy and Lieberman[1], which is perhaps the earliest use of explicit space/frequency concepts in the field. Other methods of calculating the representation include the Wigner Distribution (used in computer vision by Jau and Chin[6] and by Reed and Wechsler[8]) and localized, bandpass filters, such as Gabor functions (used by Heeger[5] and by Bovik, Clark and Geisler[2]).

The spectrogram of a signal is a series of small-support, Fourier transforms of the signal, each centered around a different point of the signal. For a one-dimensional signal $f(x)$, the spectrogram is $S_f(x,u)$, where u is frequency in cycles/unit distance. $S_f(x,u)$ is an estimate of the power of frequency u at the point x. The continuous spectrogram of the one-dimensional function $f(x)$ is given by

$$S_f(x,u) = \left| \int_{-\infty}^{\infty} w_l(\alpha - x) f(\alpha) e^{-j2\pi u\alpha} d\alpha \right|^2 ,$$

where $w_l(x)$ is a window function with support length l.

The process by which a spectrogram is calculated is shown in Figure 2. To calculate one vertical slice of the spectrogram for a given value of x, say x_o, the signal is first multiplied by a window offset by x_o. This product is Fourier transformed; the magnitude is calculated from the complex values of the Fourier transform; and the non-negative half of the magnitudes serve as $S_f(x_o,u)$, which is one column of the spectrogram. This process is repeated for every x. We consider only the non-negative half of the magnitudes since the Fourier transform of a real signal (the only kind we have) is symmetric in magnitude. The discrete version is computed using the discrete Fourier transform (DFT), which is discrete in both space and frequency.

The window function controls how the rest of the signal contributes to the spectrogram at the point x. There are ongoing questions about the best shape and size of the window $w_l(x)$. Many window shapes are considered by Harris in [4]. He illustrates the compromises involved in the selection, and concludes by recommending the 4-sample Blackman-Harris window. We use the minimum, 4-sample Blackman-Harris window, which for a discrete set of n points is given by

$$w_n(k) = a_0 - a_1 \cos\left(\frac{2\pi}{n-1}k\right) + a_2 \cos\left(\frac{2\pi}{n-1}2k\right) - a_3 \cos\left(\frac{2\pi}{n-1}3k\right)$$

for $k = 0,1,\ldots,n-1$ and $(a_0,a_1,a_2,a_3) = (0.35875,0.48829,0.14128,0.01168)$.

The window size l (or in the discrete case n) affects how much of the signal is included in the Fourier transform at each point. In practice, we have found $n = 63$ to be satisfactory on discrete signals of length 512 (one image scan-line).

2 Three-Dimensional Shape and the Spectrogram

Texture is an important indication of 3D shape, and the connection has been studied extensively in computer vision. The projected, local spatial frequencies on a textured surface change with the surface's depth and orientation, as shown in Figure 1. This is the phenomenon that makes shape from texture possible, and it is the reason that the spectrogram is a natural choice for this kind of analysis. In the following discussion, we describe how to quantitatively extract shape information from the spectrograms of flat, textured surfaces by calculating the effect of depth and orientation on the spatial frequencies of the projected texture pattern.

2.1 Mathematical Formulation

The coordinate system and other quantities are defined as in Figure 3. The pinhole of a pinhole camera is placed at the origin of the right-handed (x_{3D}, y_{3D}, z_{3D}) coordinate system, looking along the $-z_{3D}$ axis. Objects are projected onto the image whose axes are (x,y). The pinhole-to-sensor distance is d, meaning that point (x_{3D}, y_{3D}, z_{3D}) will be projected onto the image plane at the point $(x,y) = (\frac{x_{3D}d}{-z_{3D}}, \frac{y_{3D}d}{-z_{3D}})$ under perspective. In general, there is a surface in front of the camera with a superimposed intensity pattern given by $g(s,t)$, where (s,t) are coordinates of a coordinate system on the surface. We will ignore the y_{3D} and y coordinates, in effect confining our attention to the x_{3D}-z_{3D} plane ($y_{3D} = 0$) and a 1D image plane in x.

In the case of our 1D analysis of flat, textured surfaces, the camera sees a line given by $x_{3D} \sin\theta + z_{3D} \cos\theta = -\rho$.

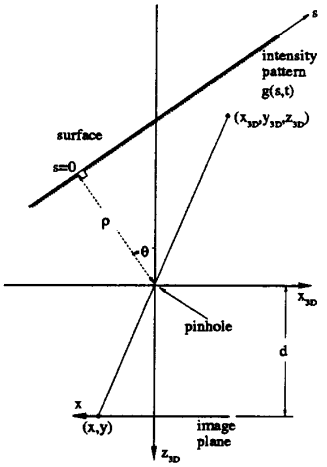

Figure 3: Geometry of 1D image formation through pinhole

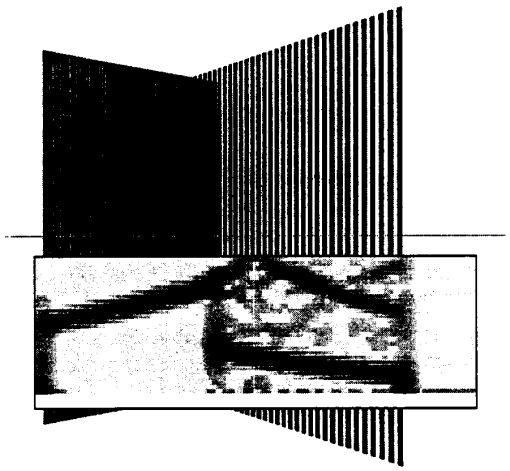

Figure 4: Two slanted plates with sinusoidal and square wave gratings

Points on this line are parameterized by s, where $s = 0$ occurs at the intersection of the line and its perpendicular to the origin. Given an s along the line, the coordinates in the scene are

$$(x_{3D}, z_{3D}) = (-\rho \sin\theta + s\cos\theta, -\rho\cos\theta - s\sin\theta)$$

which projects to the image plane at

$$x = d\frac{-\rho\sin\theta + s\cos\theta}{\rho\cos\theta + s\sin\theta}.$$

Solving for s, we have have the position along the line for a given x on the image plane:

$$s(x) = \frac{-d\rho\sin\theta - x\rho\cos\theta}{x\sin\theta - d\cos\theta}. \quad (1)$$

Suppose that the line has superimposed on it a periodic reflectance pattern given by $g(s) = \cos(2\pi u_l s)$, such that the frequency of the pattern along the line is u_l. If the pattern is projected onto the image plane, we can write the equation of the projected pattern by replacing the s in $g(s)$ with the equivalent value of s given in terms of x in Equation 1. Thus, the projected pattern on the image plane will be given by

$$\cos[2\pi u_l s(x)] = \cos\left[-2\pi u_l \rho \frac{d\sin\theta + x\cos\theta}{x\sin\theta - d\cos\theta}\right].$$

The instantaneous frequency, $u(x)$, of $\cos[2\pi u_l s(x)]$ is defined in the signal processing literature to be the derivative of the argument with respect to x, which is

$$u(x) = \frac{u_l \rho d}{(x\sin\theta - d\cos\theta)^2} \quad (2)$$

in cycles/unit distance. The peak frequency in the spectrogram of the projected cosine will occur at approximately this frequency. In a computer vision application, the known quantities in Equation 2 are d (the pinhole-to-sensor distance), x (the pixel position), and $u(x)$ (the instantaneous frequency from the spectrogram). The unknowns are u_l (the frequency of the pattern along the line), and ρ and θ (the parameters of the line). Since u_l and ρ occur as a product in Equation 2, they cannot be distinguished from each other. This is a manifestation of a familiar effect: a small object (high frequency) at a small distance is indistinguishable from a large object (low frequency) at a large distance. Thus, we treat the product $u_l\rho$ as a single unknown. With θ as the other unknown, we can solve Equation 2 for θ and $u_l\rho$ if we have two or more sets of $(x, u(x))$. The result is a space/frequency formulation of the shape-from-texture paradigm.

2.2 Extracting Shape from the Spectrogram

To demonstrate the use of Equation 2, we will determine parameters of the two plates in Figure 4 based on the spectrogram of the center row. This image was artificially generated with a sinusoidal pattern on the left plate and a square wave pattern on the right plate.

We simplify the spectrogram to $u(x)$, the dominant frequency, determined by finding the maximum value in each column of the spectrogram. Thus, we ignore the pattern's overtones. For a spectrogram window size n, there are $\frac{n+1}{2}$ non-negative frequency values for each point in a 1D signal. Since we use $n = 63$, there are only 32 different possible frequency values, which is too coarse a sampling for the shape analysis we propose. Thus, we compute a "subpixel" frequency value at each point by fitting a quadratic to the peak frequency value and its two vertical neighbors, and then find the maximum of the quadratic. This is done for each column in the spectrogram. Simulations have shown this technique consistently underestimates the actual frequency slightly, and we are currently investigating the reason.

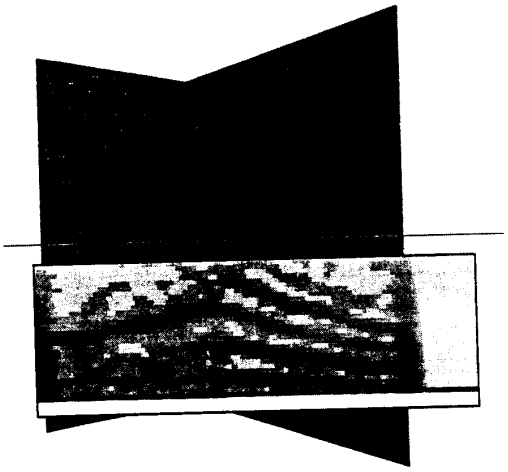

Figure 5: Two slanted plates with Brodatz textures

Each pair of $(x, u(x))$ values from the subpixel spectrogram estimates can be used to calculate a value of $(u_l\rho, \theta)$. In order to reduce the effects of the wavering in the instantaneous frequencies, we calculate each $(u_l\rho, \theta)$ using five pairs of $(x, u(x))$'s placed symmetrically around the point of interest. We then segment the regions by histograming the $(u_l\rho, \theta)$'s, manually picking the peaks, and classifying each $(u_l\rho, \theta)$ pair by finding which peak it is closest to.

With the regions segmented, we calculate the best fit $(u_l\rho, \theta)$ from Equation 2 based on the region's $(x, u(x))$'s using a gradient descent, minimization routine. The results are shown in Table 1. We know the actual values of the parameters from the graphics routine used to generate the images. In this example the errors are quite small.

We performed the same analysis for the textured plates in Figure 5. These plates are geometrically the same as those in Figure 4, but with Brodatz [3] textures mapped on using a computer graphics program. The spectrogram is messier due to irregularities in the texture. The performance figures in Table 1 are based on a manual (perfect) segmentation of the center scan-line. Although the angle estimates for Figure 5 show more error than for Figure 4, the results are still good, with the greatest error being only about 1^o.

2.3 Other Shapes

This method could be extended to other shapes in two different ways. Above we presented a method in which the instantaneous frequencies are fit to a known class of shapes (lines) in order to derive the parameters of the shape. The parameters were those which best fit Equation 2, which describes the instantaneous frequencies on a line. Other equations could be derived which relate instantaneous frequencies to any parameterized shape. Given some *a priori* knowledge of the shapes in the scene, the spectrogram peaks (as well as overtones) could be used to instantiate the shapes' parameters. Alternatively, a program could calculate local surface normals by using the instantaneous frequencies from a small neighborhood along with an equation which relates frequency and surface normal.

Although this method and results are meant to be only illustrative, they show the power of the method for analyzing the effects of 3D shape in images. The spectrogram is a simple, natural method of quantifying the relationship between texture and shape, and it requires no feature detection except for finding frequency peaks.

3 Aliasing and the Spectrogram

Aliasing occurs when a signal is sampled at a rate less than twice its maximum frequency, causing lower-frequency artifacts to appear in the sampled signal. This phenomenon can often be seen on television in images of periodic patterns like striped clothes, automobile grills, or tall buildings. In two dimensional imaging, these artifacts are called *moire patterns*, and they can lead to insidious problems in machine vision. The patterns cannot be detected in single images without detailed *a priori* knowledge of the scene. In these situations there is no hope of recovering the true signal. However, the spectrogram serves as an elegant method of analyzing such behavior.

This is illustrated in Figure 6, which shows a plate with a sinusoidal intensity pattern rotated to the right. Beginning at the left of the plate, the spectrogram shows that the instantaneous frequency is rising as the plate recedes into the distance. At a little less than halfway across the spectrogram, the peak frequency has risen to the top of the spectrogram, which corresponds to the Nyquist frequency (half the spatial sampling frequency). Although the actual frequency on the image plane continues to rise, it appears to decrease after the Nyquist rate has been exceeded. In this region of the image, moire patterns begin to appear as lower-frequency variations caused by the beating of the signal frequency against the sampling frequency. There is another "bounce" on the spectrogram after the apparent peak frequency has fallen to zero. This bouncing would continue if the plate were longer. If the signal had overtone frequencies, these will bounce also, although not at the same places as the fundamental or other overtones.

In an ideal space/frequency representation, there would be no Nyquist limitation, and the spectrogram in Figure 6 would rise without bound. We show in [7] that the actual spectrogram is a folded version of the ideal spectrogram, as shown in Figure 7. The folds occur at integer multiples of the Nyquist frequency. We also present in [7] an algorithm for unfolding the spectrogram based on two images of the same scene taken at slightly different zoom settings of the lens. Thus, the space/frequency representation is an effective means of analyzing local aliasing, something that is impossible with either a pure spatial or pure frequency

	From Figure 4 Periodic Pattern semi-automatic segmentation				From Figure 5 Brodatz Textures manual segmentation			
	Left Plate		Right Plate		Left Plate		Right Plate	
	$u_l\rho$	θ	$u_l\rho$	θ	$u_l\rho$	θ	$u_l\rho$	θ
actual	177.25	50.00^o	40.00	-60.00^o	152.1	50.00^o	47.0	-60.00^o
calculated	172.92	49.75^o	39.31	-59.72^o	141.37	50.82^o	48.27	-58.85^o
error	-2.4%	-0.25^o	-2.4%	0.28^o	-7.1%	0.82^o	2.7%	1.15^o

Table 1: Actual and calculated line parameters

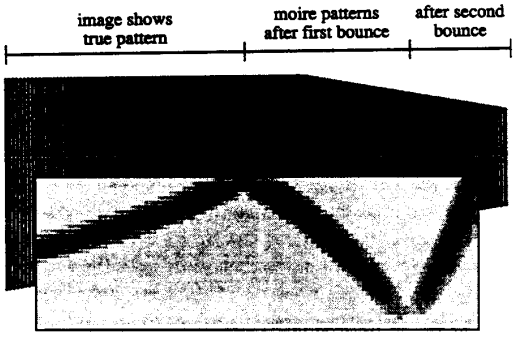

Figure 6: Plate with sinusoid showing aliasing

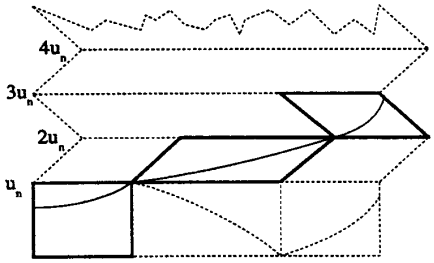

Figure 7: Folding the ideal spectrogram to show aliasing, u_n = Nyquist frequency

representation of the signal.

4 Conclusion

This paper demonstrates the utility of the space/frequency representation for computer vision by showing how the representation can be used for doing shape from texture and analyzing local aliasing. The procedures shown here involve straight-forward applications of geometry and frequency concepts. That these concepts can be applied simultaneously demonstrates the utility of the representation.

The attractiveness of the space/frequency representation lies in the variety of phenomena that it can be used to analyze. We demonstrate these in [7], where we show the versatile abilities of the representation. In addition to the analysis of shape and aliasing shown in this paper, the representation can be used for segmentation, characterizing the effects of lens parameters, and signal matching. We believe that the power of the space/frequency representation will make it possible to develop far more comprehensive methods for low-level vision than the current heuristic techniques allow.

References

[1] Ruzena Bajcsy and Lawrence Lieberman. Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 5:52–67, 1976.

[2] Alan Conrad Bovik, Marianna Clark, and Wilson S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):55–73, January 1990.

[3] Phil Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, 1966.

[4] Fredric J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, January 1978.

[5] David J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302, January 1988.

[6] Y.C. Jau and Roland T. Chin. Shape from texture using the wigner distribution. In *Computer Vision and Pattern Recognition*, pages 515–523. Computer Society Press, June 1988.

[7] John Krumm and Steven A. Shafer. Local spatial frequency analysis for computer vision. Technical Report CMU-RI-TR-90-11, Carnegie Mellon University Robotics Institute, May 1990.

[8] Todd R. Reed and Harry Wechsler. Segmentaion of textured images and gestalt organization using spatial/spatial frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):1–12, January 1990.