Proceedings of the 1996 IEEE
International Conference on Robotics and Automation
Minneapolis, Minnesota - April 1996

Determining Digitizing Distances on Sculptured Surfaces Using Short Time Fourier Transform

Chia-Chun Huang and Yuan F. Zheng
Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210

Abstract

Surface digitization is a method for obtaining three-dimensional measurement of an unknown surface. The performance and quality of digitization are dependent upon the choice of digitizing devices and approaches. An efficient digitizing method not only reduces the elapsed time of the digitizing process, but also improves the precision of measurement. This paper presents a new nonuniform technique for digitizing of sculptured surfaces based on the short time Fourier transform (STFT). The digitizing distances in the x and y directions for each pre-divided block of the surface are adjusted according to the highest spatial frequencies contained in the block which is in contrast to the uniform technique in which the digitizing distance is determined according to the highest spatial frequencies of the entire surface. The speed of the digitizing process is thus improved and meanwhile the accuracy of the surface representation is maintained. Simulations and experimental results are provided to verify the proposed method.

1 Introduction

In the manufacturing industry, "reverse engineering" is becoming an indispensable process for manufacturing of objects with sculptured surfaces. Digitizing an unknown surface is an important step in reverse engineering. Various digitizing techniques are available to measure the depths of discrete points of a surface. These points form a dense mesh which can be used to reconstruct the surface.

The efficiency of the digitizing process is a very important issue and has drawn a great deal of attention in recent years. The fundamental question is that what the digitizing distance should be such that the surface can be accurately represented but the digitizing points are minimal. Much research has been done on this problem. Lin, et al, used the spectral analysis to determine the surface proper sampling spacing under the constraints of permitted aliasing error [1]. Sherrington and Smith used the Fourier analysis method to obtain the frequency information of areal surfaces in the measurement of roughness [2]. Tsukada and Sasajima [3], and Yim and Kim [4] treated this problem from the statistical point of view. Woo and Liang [5] applied the Hammersley sequence for surface measurement and the number of digitizing points can be reduced. All of the above methods consider the whole surface as a block and the digitizing intervals in the x and y axes are determined globally. Once the digitizing intervals are determined according to the highest spatial frequency, they are applied uniformly to the entire surface. This may result in too many digitizing points because there are places on the surface whose spatial frequencies are lower and need longer digitizing distances.

The objective of this research is to reduce the digitizing points by a nonuniform digitizing approach. Most often a sculptured surface has different spatial frequencies in different local areas. Each local area should have its own digitizing density according to its local spatial frequencies.

In this paper, we propose a new approach which divides the entire surface into small windows. The digitizing points are determined for each window according to the spatial frequencies of each window using the short time Fourier transform (STFT) method. The advantages of this approach are evident. Because the digitizing intervals are adjusted according to the highest frequency within each windowed block, the method needs less digitizing points and therefore reduces the processing time. Furthermore, the size of the pre-divided block can be altered in order to get the optimal reduction of digitizing points. On the other hand, the accuracy of the surface representation is preserved because the digitizing distance is selected according to the highest spatial frequency in the block.

The procedure of the proposed approach is in the following order:

1. Obtain a "rough picture" of the sculptured surface by using computer vision. Computer vision can measure the entire surface very quickly with a relatively low precision. However, the information of the spatial frequency is preserved.
2. Perform the short time Fourier transform (STFT) on the rough depth information integrated from computer vision.
3. Find the energy spectrum of the STFT. The highest frequencies which enclose most of the energy of the spectrum in the x and y directions can be obtained. The desired digitizing distances in the x and y directions should be at least twice the highest frequencies according to the Shannon sampling theory [6].
4. Use an accurate digitizing device such as a laser displacement sensor to redigitize the selected points. An accurate digitizing device needs longer time than computer vision to digitize the entire surface. Because the digitizing points have been reduced, the digitizing time is substantially reduced as well and the high precision of digitizing is obtained.

The structure of this paper is as follows. In Section 2, the STFT is presented. The motivation, definition, and selection of the window function of the STFT are described in detail. In Section 3, we illustrate how to find the highest frequencies in the x and y directions and determine the new digitizing distances. An experiment of a model and the error analysis are presented in Section 4. Finally, this paper is concluded in Section 5.

2 Short Time Fourier Transform

2.1 Motivation of the STFT

The STFT was originated for the purpose of detecting nonstationary signals and was widely used in the research of speech and acoustics. For a better and clearer understanding of the groundwork of the STFT, a comparison between the STFT and the conventional Fourier transform (FT) should be made.

The discrete Fourier transform and the corresponding inverse Fourier transform have the following forms:

$$\begin{aligned} F(\omega) &= \sum_{n=-\infty}^{\infty} f(n)e^{-j\omega n} \\ f(n) &= \frac{1}{2\pi}\int_0^{2\pi} F(\omega)e^{j\omega n}d\omega \end{aligned} \qquad (1)$$

where $f(n)$ and $F(\omega)$ are the transform pair in the discrete time and frequency domains, respectively.

From (1), the spectrum $F(\omega)$ is calculated from the time $-\infty$ to ∞. This means that it is necessary to observe the signal $f(n)$ for a very long time to get the frequency spectrum distribution. In the speech applications, this is often accomplished by analyzing the recorded sound from the very far beginning to the present time. Another observation from (1) is that the spectrum only tells the total frequency distribution over the entire time axis (or at least a long time duration). It is unable to distinguish what frequency components have happened at a specific time, or when a particular frequency component was happening. The deficiency of the FT, for example in the acoustic research, is that the FT is obviously uninformative when the musical notes are analyzed because the time and the frequency of the notes have to be considered simultaneously.

The intuitive solution to the mentioned shortcomings of the FT is to split the entire time space into a number of small durations and apply the FT on each one. It seems to be quite helpful and straightforward, but this gives rise to two problems thereafter:

1. How to divide the time axis? What is the length of each time period? Note that the FT requires long durations for the integration and hence the periods should not be too short. It also implies that some preliminary information about the signal $f(n)$ has to be obtained *a priori* to prevent any arbitrary division of the signal.
2. If a fixed length T is specified for the period, each period amounts to a rectangular window with the length T imposed on the original signal $f(n)$. The windowed signal f_w which is centered at τ can be expressed as

$$f_w(n-\tau) = f(n)\cdot \mathrm{rect}(n-\tau) \qquad (2)$$

where

$$\mathrm{rect}(n) = \begin{cases} 1, & -T/2 \le n \le T/2, \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

The multiplication of $f(n)$ and $\mathrm{rect}(n-\tau)$ in the time domain is equivalent to the convolution of $F(\omega)$ and $\mathcal{FT}\{\mathrm{rect}(n-\tau)\}$ in the frequency domain. The FT of $\mathrm{rect}(n-\tau)$ is

$$\mathcal{FT}\{\mathrm{rect}(n-\tau)\} = \frac{\sin[\omega(T+1)/2]}{\sin(\omega/2)} \qquad (4)$$

where T is assumed to be an even number. This convolution brings unwanted high frequency components into the spectrum since there are sharp edges along the two sides of the rectangular window function. This disturbs the true frequency distribution retained in $f(n)$. As we will see in the following section, the rectangular window is the extreme case of the *uncertainty principle* [7]. Hence such a window is not a good choice in most applications. The question is: what is the most suitable window function which can reduce the time and frequency disturbance as much as possible? In other words, what kind of window function is optimal in both time and frequency domains?

2.2 Definition of the STFT

Since the purpose of the STFT is to identify the frequency of the sculptured surface locally instead of the frequency globally, the formal definition of the STFT in 1-D case can be expressed as

$$F(m,\omega) = \sum_{n=-\infty}^{\infty} f(n) \cdot w(n-m) e^{-j\omega n} \tag{5}$$

where $w(n-m)$ is the translated version of the window function $w(n)$ located at m. Therefore, the window function can be moved everywhere along the discrete n-axis, and this amounts to seeing different durations of windowed signal accordingly.

2.3 Gaussian Window

The properties of the STFT are sensitive to the selection of the window function in (5). The window function can be generalized as $w(n)$ which can be moved everywhere in the n-axis, and it is multiplied to $f(n)$ to get the windowed version $f_w(n)$. The width T of the window $w(n)$ and the bandwidth B of its FT, $W(\omega)$, are related according to the "uncertainty principle" (Heisenberg inequality) [8][9], i.e.,

$$T \cdot B \geq \frac{1}{4\pi} \tag{6}$$

where T and B are defined as follows and they can be considered as the standard deviations of the window function with associated expectation values n_0 and w_0 [8]:

$$T \equiv \left(\sum_{n=-\infty}^{\infty} (n-n_0)^2 |w(n)|^2 \right)^{1/2}, \tag{7}$$

$$B \equiv \left(\int_{-2\pi}^{2\pi} (\omega-\omega_0)^2 |W(\omega)|^2 \, d\omega \right)^{1/2} \tag{8}$$

(9)

where

$$n_0 \equiv \sum_{n=-\infty}^{\infty} n|w(n)|^2, \tag{10}$$

$$\omega_0 \equiv \int_{-2\pi}^{2\pi} \omega |W(\omega)|^2 \, d\omega. \tag{11}$$

In other words, when the width of the window is smaller, its bandwidth in frequency becomes larger, and their products are always greater than or equal to the lowest bound $1/4\pi$. Hence the basic issue in the choice of the window function is the tradeoff between the long window length for high frequency resolution and the short window length for zooming in the local variations of the signal at the time of interest. We have already presented an example of rectangular window function. Another extreme case is the Dirac delta function $\delta(n)$ with $T = 1$. This results in infinite length of "1" in the frequency domain.

The Gaussian window $g(n)$ is compact in both time and frequency domains and meets the lowest bound of the Heisenberg inequality. This transform using the Gaussian window was introduced by Gabor [9] and is known as *Gabor transform*:

$$g(n) = (2a)^{1/4} e^{-\pi a n^2}, \tag{12}$$

$$G(\omega) = (2/a)^{1/4} e^{-\pi\omega^2/a}, \tag{13}$$

and

$$T = \sqrt{\frac{1}{4\pi a}}, \quad B = \sqrt{\frac{a}{4\pi}} \tag{14}$$

where a is a constant which can be used to adjust the window width and the bandwidth. Since the Gaussian window has the property of continuity in both time and frequency domains, the interference resulted from the unwanted high frequency can be avoided. In reality, the spatial frequencies of sculptured surfaces might have a large range. To reduce the number of digitizing points, both the digitizing distances and the spatial frequencies have to be taken into consideration. The Gaussian window gives us a reasonable and conservative compromise of time and frequency resolutions. Therefore, the Gaussian window is chosen as the window function of the STFT in our research.

2.4 2-Dimensional STFT

For simplicity, the deduction of the STFT in the previous sub-section is based on 1-D signals. The discrete 2-D STFT which is necessary in the surface digitization can be considered as an expansion of the discrete 1-D STFT:

$$F(m_1,m_2,\omega_1,\omega_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f(n_1,n_2) \cdot w(n_1-m_1, n_2-m_2) e^{-j(\omega_1 n_1 + \omega_2 n_2)}. \tag{15}$$

The Gaussian window function $g(n_1,n_2)$ which is a choice of $w(n_1,n_2)$ can be expressed as

$$g(n_1,n_2) = \sqrt{2a}\, e^{-\pi a(n_1^2+n_2^2)}. \tag{16}$$

A Gaussian window with size 15×15 is shown in Figure 1.

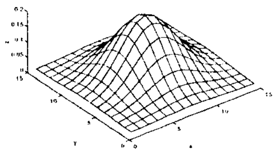

Figure 1: A 15×15 Gaussian window

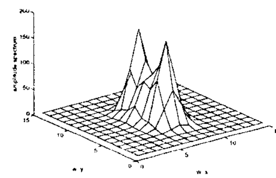

Figure 2: Spectrum of the STFT on a 15×15 block

3 Finding the Highest Frequencies of Bandlimited Signals

After applying the 2-D STFT with the Gaussian window on the sculptured surface, the magnitude of the spectrum of each block can be obtained. The square of the magnitude of each component in the spectrum represents the energy at the corresponding spatial frequencies. The center of the spectrum is the d.c. term and when spreading out from the center the frequency increases. Note that for a real signal, the spectrum is symmetric about the center. In practice, there is rarely truly bandlimited signal and the high frequency component is often existing more or less if not impossible. But from the above induction, we can use the ratio of energy as a reasonable bandlimited criterion for all the frequencies components [10]. For a $N \times N$ spectrum map where N is an odd number, the bandwidths B_x and B_y in the directions x and y respectively can be calculated as the minimum arguments which satisfy:

$$\frac{\sum_{k_1=(N+1)/2-B_x}^{(N+1)/2+B_x} \sum_{k_2=(N+1)/2-B_y}^{(N+1)/2+B_y} |X(k_1,k_2)|^2}{\sum_{k_1=1}^{N} \sum_{k_2=1}^{N} |X(k_1,k_2)|^2} \leq 1-\delta \tag{17}$$

where the frequency spectrum $X(k_1, k_2)$ is the discrete version of the FT using the FFT. The denominator is the sum of the energy of all the frequency components in the window, and δ is a suitably small constant which can be expressed as in decibel (db). For example, if the sum of $|X(k_1, k_2)|^2$ is in the order of 1.0, δ can be selected between 0.0001 (40 db down) and 0.01 (20 db down). As δ is smaller, the accuracy is improved and meanwhile the number of digitizing points will nevertheless increase. Therefore, the selection of δ depends on the requirement of efficiency and accuracy of a specific digitization. Here we can use 20 db as a criterion for the bandlimited signals, i.e., 99% of the total energy is concentrated in the area surrounded by the bandlimited frequencies in the x and y directions.

As an example, Figure 2 shows that the highest frequencies of the discrete STFT of a signal in the x and y directions which enclose 99% of the total energy are $3\pi/7$ and $5\pi/7$. The frequency Ω in the discrete-time STFT is related to the frequency ω by the following equation:

$$\Omega = \omega / D \tag{18}$$

where D denotes the sampling distance. Therefore, the new digitizing distance can be obtained as

$$D_s = \frac{1}{f_s} = \frac{1}{2f_H} = \frac{\pi}{\Omega_H} \tag{19}$$

where the subscript H represents the highest frequency. Using (19) and if the original digitizing distance D is equal to 3 mm, the new digitizing distances D_s in the x and y directions are 7 mm and 4.2 mm, respectively. They are both larger than the original digitizing distance. This means that the number of the digitizing points is reduced.

Once the new digitizing points are calculated, the surface can be reconstructed by a low-pass filter. Here, the choice of the low-pass filter makes one face the same dilemma as encountering in the choice of sampling windows. Since the distribution of the digitizing points is irregular globally but uniform within each window, two categories of reconstruction or interpolation methods can be used:

1. Interpolate each window and concatenate all the windows together. Fundamentally, one has to use the 2-D interpolation functions to do the convolution in each window [11]. Between two neighboring windows, the boundary points are calculated using some smoothing functions, e.g., averaging.

2. Irregularly interpolate the entire surface. For example, biharmonic spline interpolation proposed by Sandwell can be used for this purpose [12]. This method is used for minimum curvature interpolation of irregularly distributed data points. But unfortunately, it is inefficient and very demanding of computing resources.

4 Experiment

4.1 Equipment Setup

The following setup was used in the experiment to verify the proposed approach:

1. A photometric stereo approach [13] was used in computer vision to obtain the initial depth information of the surface. A coordinate measuring machine (CMM) was used to precisely position the point light source which was used in the photometric stereo approach.
2. A high-intensity halogen bulb was used as the point light source. In order to make the light uniformly distributed in all the directions which is a requirement in the photometric stereo approach, the bulb was wrapped by white-colored plastic tapes.
3. A metal goose model coated with white paint was used as a test model. The paint was used to avoid the specular effect of the surface. The size of the model is about 4"×3".
4. Since the photometric stereo was determined by the magnitude of illumination, many images under the same illumination conditions were taken and averaged to reduce the noise variance.
5. The images were stored in GIF (Graphics Interchange Format) with size 240×320 and 256 gray levels.

4.2 Results

Three different illuminated images were taken when the point light source was placed at the top, right, and left sides of the model, respectively. One of the images is shown in Figure 3.

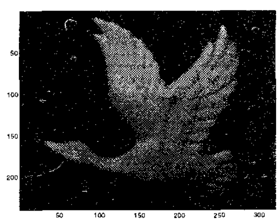

Figure 3: Image of Goose

The photometric stereo approach calculated the gradients of every element of the surface using these three images. By integrating the gradients of the surface, the depth information of the surface can be obtained [14]. The needle diagram of the surface and the integrated depth map are shown in Figures 4 and 5, respectively. It should be noted that the sampling points were evenly distributed at this stage and the distribution was based on the highest frequency of the model. For the goose model, the digitizing distance was 0.4 mm which was called the pre-selected digitizing distance, and the total digitizing points were $240 \times 320 = 76,800$.

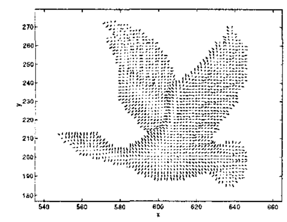

Figure 4: Needle diagram of goose model (decimated by 5)

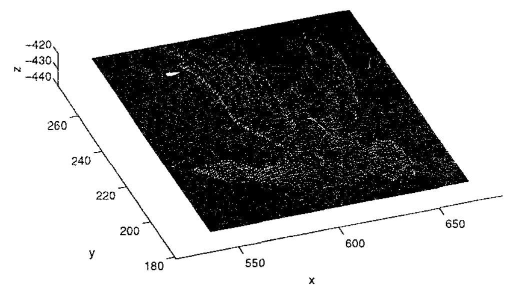

Figure 5: Integrated 3-D depth map

The window size T was chosen to be 16 and this divides the entire surface into 300 windows. The STFT was applied to each window. When the bandlimit criterion was chosen to be 90 db, the new digitizing points were obtained for each window as shown in Figure 6.

In Figure 6, one may notice that there are darker and lighter squares. The darkest square contains 16×16 digitizing points while the lighter squares contain less digitizing points. The total digitizing points were reduced to 24,301.

After measuring the selected digitizing points by a laser displacement sensor (LDS) mounted on the CMM and reconstructing the model using the 2-D bilinear interpolation, the goose model is illustrated in Figure 7.

To evaluate the performance of the STFT approach, we used the same LDS to evenly digitize the entire surface with the pre-selected digitizing distance 0.4

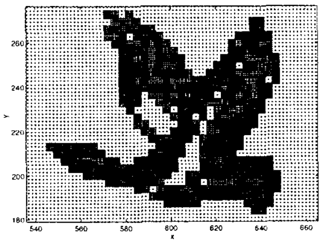

Figure 6: New digitizing points determined by the STFT (90 db)

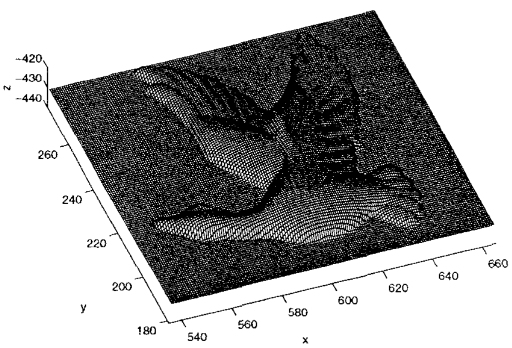

Figure 7: Bilinear reconstruction of the goose using new digitizing points

mm. It took almost 32 hours to complete the digitizing (240×320). In comparison, the STFT approach only needed 10.12 hours to complete the task. Since every point was digitized using the LDS which was highly accurate, we were able to evaluate the precision of the STFT approach. The reconstructed model was compared with the measured data at every pre-selected digitizing point. We found that the maximum error was 4.958×10^{-2} mm and the average mean square error was 3.9396×10^{-5} mm. This indicates that the STFT approach was both efficient and accurate.

The choice of the window length T is empirical in the experiment. In fact, this is the compromise between the window length and the frequency resolution as stated in Section 2. In the experiment, if the window length is chosen too small, the part of the signal that is covered by the window behaves like an impulse which generates a very wide frequency spectrum. This means that we have to choose all the points within the small window. On the other hand, if the window is too large, the frequency spectrum spans a wide range which defeats the purpose of the STFT.

5 Conclusions

In this paper, we have proposed an STFT based approach to improve the efficiency of surface digitizing. The STFT is applied to sculptured surfaces for the purpose of reducing the digitizing points. The reduced digitizing points are still able to represent the original surface with a high accuracy. This creates an efficient way to digitize and represent a sculptured surface. In the experiment, a goose model was digitized by a computer vision method first. Then the STFT approach was applied to the model. By doing so, the number of digitizing points was reduced while the accuracy was still maintained.

Acknowledgment

This work was supported by the Engineering Research Center/Net Shape Manufacturing at The Ohio State University.

References

[1] T. Lin, L. Blunt, and K. Stout, "Determination of proper frequency bandwidth for 3d topography measurement using spectral analysis. part i: isotropic surfaces," *Wear*, vol. 166, no. 2, pp. 221–232, July 1993.

[2] I. Sherrington and E. Smith, "Areal fourier analysis of surface topography, part 1: Computational methods and sampling considerations," *Surface Topography*, pp. 43–68, 1990.

[3] T. Tsukada and K. Sasajima, "An optimum sampling interval for digitizing surface asperity profiles," *Wear*, vol. 83, pp. 119–128, 1982.

[4] D. Yim and S. Kim, "Optimum sampling interval for r_a roughness measurement," in *Proc. Instn. Mechanical Engineerings, Part C*, pp. 139–142, 1991.

[5] T. Woo and R. Liang, "Dimensional measurement of surfaces and their sampling," *Computer Aided Design*, vol. 25, no. 4, pp. 233–239, April 1993.

[6] C. Shannon, "Communication in the presence of noise," in *Proc. IRE*, January 1949.

[7] A. Messiah, *Quantum Mechanics*. Amsterdam: North-Holland, 1961.

[8] G.Kaiser, *A Friendly Guide to Wavelets*. Boston: Birkhäuser, 1994.

[9] D. Gabor, "Theory of communication," *Journal of the IEE*, vol. 93, pp. 429–457, 1946.

[10] A. Oppenheim, *Applications of digital signal processing*. Prentice-Hall, 1978.

[11] A. Jain, *Fundamentals of digital image processing*. Prentice-Hall, 1989.

[12] D. Sandwell, "Biharmonic spline interpolation of geos-3 and seasat altimeter data," *Geophysical research letters*, vol. 14, no. 2, pp. 139–142, February 1987.

[13] R. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[14] N. Kolagani and J. S. Fox, "Photometric stereo using point light sources," in *Proc. 1992 IEEE, International Conference on Robotics and Automation*, pp. 1759–1764, 1992.