A Simple and Fast Spline Filtering Algorithm for Surface Metrology

Hao Zhang[1,2]**, Daniel Ott**[2]**, John Song**[2]**, Mingsi Tong**[2,3]**, and Wei Chu**[2]

[1]College of Mechanical and Electronic Engineering,
Nanjing Forestry University,
Nanjing, 20877, China

[2]National Institute of Standards and Technology,
Gaithersburg, MD 20899 USA

[3]School of Mechatronics Engineering,
Harbin Institute of Technology,
Harbin, 150001, China

hao.zhang@nist.gov
daniel.ott@nist.gov
jun-feng.song@nist.gov
mingsi.tong@nist.gov
wei.chu@nist.gov

Spline filters and their corresponding robust filters are commonly used filters recommended in ISO (the International Organization for Standardization) standards for surface evaluation. Generally, these linear and non-linear spline filters, composed of symmetric, positive-definite matrices, are solved in an iterative fashion based on a Cholesky decomposition. They have been demonstrated to be relatively efficient, but complicated and inconvenient to implement. A new spline-filter algorithm is proposed by means of the discrete cosine transform or the discrete Fourier transform. The algorithm is conceptually simple and very convenient to implement.

1. Introduction

A typical engineering surface consists of a range of spatial frequencies. The high frequency or short wavelength components are referred to as roughness, the medium frequencies as waviness and low frequency components as form [1, 2]. Filtration is a critical process used to isolate form, waviness, and surface roughness from each other as well as from finer, higher frequency content such as instrument noise [3]. At present, the Gaussian filter, the spline filter and their corresponding robust filters are recommended as ISO standard profile filters [4-7]. Compared with the widely used Gaussian filter, the spline filters have advantages such as end preserving, efficient calculation and good form following.

Generally, the spline filter is implemented using an iterative procedure based on a Cholesky matrix decomposition [2, 8]. It is a widely used and relatively efficient method for solving the spline filter. In this paper, a new approach for computing the spline filter is proposed that relies on the discrete cosine transform (DCT) or the discrete Fourier transform (DFT) rather than matrix decomposition. This new algorithm is conceptually more simple, efficient, and, in practice, more convenient to implement.

Journal of Research of the National Institute of Standards and Technology

2. The Spline Filters

2.1 Linear Spline Filter

In ISO/TS 16610-22 [5], the standard spline filter is given as

$$(I + \beta\alpha^2 P + (1 - \beta)Q)w = z \qquad (1)$$

where I is the identity matrix, z is the vector of sampled data values, w is the vector of output data values, and P and Q are the coefficient matrices. The spline filter of Eq. (1) is obtained as the solution matrix of the classic variational method in Eq. (2) [8-9].

$$\min(\|z - w\|^2 + \beta\alpha^2\|D_1 z\|^2 + (1 - \beta)\|D_2 z\|^2) \qquad (2)$$

where $\|\cdot\|$ denotes the Euclidean norm (L$_2$ norm), and D_1 and D_2 stand for 1st and 2nd differential operators. Generally, Eq. (2) is described as a linear combination of a least squares approach and a penalty component consisting of two differential terms. More specifically, the least squares term aims to guarantee a close match between the result w and the profile z [10], and the penalty component helps to ensure appropriate smoothness of the filtered result. The scalars α and β are used to determine the balance between closeness to the original data and amount of smoothness, where $\alpha = \left(2\sin(\pi/\lambda_c)\right)^{-1}$ (λ_c is the cut-off wavelength) and $0 \leq \beta \leq 1$ [5]. Since all terms of Eq. (2) are differentiable, the matrix equation of Eq. (1) is obtained directly by simple mathematical derivation. Here it should be noted that the boundary conditions, which are dictated by P and Q, vary depending on the type of surface profile, which may be classified as periodic or non-periodic (see Eqs. (3) and (4)).

$$P = \begin{cases} \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} & \text{non-periodic} \\[2em] \begin{pmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{pmatrix} & \text{periodic} \end{cases} \qquad (3)$$

$$Q = \begin{cases} \begin{pmatrix} 1 & -2 & 1 & & & \\ -2 & 5 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix} & \text{non-periodic} \\[2em] \begin{pmatrix} 6 & -4 & 1 & & & 1 & -4 \\ -4 & 6 & -4 & 1 & & & 1 \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & 1 & -4 & 6 & -4 & 1 \\ 1 & & & 1 & -4 & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{pmatrix} & \text{periodic} \end{cases} \qquad (4)$$

Journal of Research of the National Institute of Standards and Technology

2.2 Robust Spline Filter

It is well known that the linear filter is vulnerable to outliers [11]. Although there is no agreement on a universal and formal definition of an outlier, it is usually regarded as the observations that lie abnormally far from others, such as scratches and deep valleys on a surface. The roughness profiles or surfaces will be distorted when the linear filtering algorithm is applied to this original sampling data containing outliers. The resulting filtered data will contain excessively swollen or sunk areas due to being "pulled down or up" in the vicinity of abnormal points. This phenomenon will certainly cause a serious error when performing the evaluation or analysis of a practical surface.

In order to solve this problem, the concept of the robustness of a filtering algorithm is introduced. In a robust filter, the effect of outliers on surrounding data is mitigated. Both L_1-norm and L_2-norm based robust spline filters have been demonstrated [7, 12]. A robust spline filter based on the L_2-norm [12] is obtained by the direct expansion of Eq. (1), that is:

$$(\boldsymbol{\delta} + \beta\alpha^2\boldsymbol{P} + (1-\beta)\alpha^4\boldsymbol{Q})\boldsymbol{w} = \boldsymbol{\delta}\mathbf{z} \qquad (5)$$

where $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the same coefficient matrices as in Eqs. (3) and (4). $\boldsymbol{\delta}$ contains the additional weights corresponding to each data point of $\boldsymbol{z}$. $\boldsymbol{\delta}$ is updated with a specified weighting function by using current residuals from iteration to iteration, until the residuals remain unchanged [13]. Several weighting functions are available such as Huber, Hampel, and Andrews. Indeed, the most widely used weighting function in surface metrology is the Tukey function [6].

3. Simplified Algorithm

3.1 The Spline Filter for Non-Periodic Profiles

As previously mentioned, the standard spline filter is endowed with different solution matrices based on the periodicity of the profile and how this is handled at boundaries. Let us first focus on the matrix $\boldsymbol{Q}$ under the non-periodic condition, which is also called the natural boundary, that is, when the 2nd differential equation equals zero. If we instead redefine the boundary condition as the 1st differential equation equalling zero, the original matrix $\boldsymbol{Q}$ can be simplified and rewritten in the following form

$$\boldsymbol{Q} = \begin{pmatrix} 2 & -3 & 1 & & & & & \\ -3 & 6 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -4 & 6 & -4 & 1 & \\ & & & 1 & -4 & 6 & -3 \\ & & & & 1 & -3 & 2 \end{pmatrix}. \qquad (6)$$

It can be easily verified that $\boldsymbol{Q} = \boldsymbol{D}^T\boldsymbol{D}$,

$$\boldsymbol{D} = \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix} \qquad (7)$$

where $\boldsymbol{D}$ is a tridiagonal matrix and $\boldsymbol{D}^T$ represents the transpose of $\boldsymbol{D}$.

Furthermore, an eigen decomposition of $\boldsymbol{D}$ yields $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ [13]. $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{D}$ [14]:

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \text{ with } \lambda_i = -2 + 2\cos((i-1)\pi/n)$$

where n is the number of sampled points.

Since $\boldsymbol{U}$ is a unitary matrix, Eq. (1) can be written as

$$\boldsymbol{w} = (\boldsymbol{I} - \beta\alpha^2\boldsymbol{D} + (1-\beta)\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{z} = \boldsymbol{U}(\boldsymbol{I} - \beta\alpha^2\boldsymbol{\Lambda} + (1-\beta)\boldsymbol{\Lambda}^2)^{-1}\boldsymbol{U}^T\boldsymbol{z} \qquad (8)$$

According to Reference [15], $\boldsymbol{U}^T$ is a type-2 DCT matrix and $\boldsymbol{U}$ is an inverse DCT matrix. Let us define $\boldsymbol{\Gamma} = \mathrm{diag}(\Gamma_1, \ldots, \Gamma_n)$ with $\Gamma_i = [1 - \beta\alpha^2\lambda_i + (1-\beta)\lambda_i^2]^{-1}$, Eq. (8) can be rewritten in the form

$$\boldsymbol{w} = \mathrm{DCT}^{-1}(\boldsymbol{\Gamma}\mathrm{DCT}(\boldsymbol{z})) \qquad (9)$$

where DCT^{-1} represents the inverse discrete cosine transform.

In fact, DCT is a Fourier-related transform similar to DFT, a technique for mapping the "time domain" to the "frequency domain." Both DCT and DFT express a signal in terms of a sum of sinusoids with different frequencies and amplitudes. The distinction between DFT and DCT is that, DFT uses both cosine and sine functions, while DCT uses only cosine functions [16]. Boundary conditions have a direct impact on the form of discrete transfer functions such that different combinations of the discrete boundary conditions give rise to eight different types of DCT. Of these eight DCT representations, the most commonly used transform is type-2 [16].

3.2 The Spline Filter for Periodic Profiles

For the case of a periodic profile, the boundary conditions will require that matrix $\boldsymbol{Q}$ refer to the classic assumption of circular data defined as $z_i = z_{i+n}$, that is

$$\boldsymbol{Q} = \begin{pmatrix} 6 & -4 & 1 & & & 1 & -4 \\ -4 & 6 & -4 & 1 & & & 1 \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ 1 & & & 1 & -4 & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{pmatrix}. \qquad (10)$$

One can obtain the same description $\boldsymbol{Q} = \boldsymbol{D}^T\boldsymbol{D}$, where $\boldsymbol{D}$ is also a circulant matrix,

$$\boldsymbol{D} = \begin{pmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{pmatrix}. \qquad (11)$$

For this case, a DCT is not able to fulfill the spline filter equation because the boundary definition for any type of DCT does not comply with a periodic boundary of $z_i = z_{i+n}$. However, it is interesting that this boundary condition is exactly the same as the boundary assumption of the DFT. According to signal processing theory, a discrete spectrum necessarily comes from a periodic signal in the time domain. Hence, it is rational to apply the fast Fourier transform (FFT) to the computation of the spline filter given a periodic boundary. As expected, following the related derivation shown in Section 3.1, we obtain

$$\boldsymbol{w} = \mathrm{FFT}^{-1}(\boldsymbol{\Gamma}\mathrm{FFT}(\boldsymbol{z})) \qquad (12)$$

where FFT^{-1} stands for the inverse fast Fourier transform.

4. Robust Simplified Algorithm

As previously mentioned, the sensitivity of the linear filter to outliers may result in the mean line deviating severely from the actual distribution. The occurrence of dropouts, e.g., resulting from an instrument's inability to measure certain data features, is frequent in practice [13]. Fortunately, the weighting function is demonstrated to be an effective method for handling such dropouts. This function serves the purpose of giving outliers a low weight and dropouts a weight of zero, while allocating a relatively high weight to high-quality data.

For convenience, Eq. (5) can be rewritten as

$$(I + \beta\alpha^2 P + (1 - \beta)\alpha^4 Q)w = (I - \delta)w + \delta z. \tag{13}$$

This implicit formula can be solved using an iterative procedure

$$w^{k+1} = (I + \beta\alpha^2 P + (1 - \beta)\alpha^4 Q)^{-1}(\delta(z - w^k) + w^k) \tag{14}$$

where w^{k+1} refers to w^k at the kth iteration step. Furthermore, referring to Eqs. (9) and (12) for the appropriate boundary conditions, we can easily obtain the new equations

$$w^{k+1} = \text{DCT}^{-1}(\boldsymbol{\Gamma}\text{DCT}(\delta(z - w^k) + w^k)) \quad \text{for the non-periodic case,} \tag{15}$$

or

$$w^{k+1} = \text{FFT}^{-1}(\boldsymbol{\Gamma}\text{FFT}(\delta(z - w^k) + w^k)) \quad \text{for the periodic case.} \tag{16}$$

The robustness to outliers is realized by iterating the weighted function of Eq. (15) or (16) and updating the weights using robust estimators for the current residuals $(z - w^k)$. It should be noted that during one iteration, the DCT and FFT only require $n\log(n)$ operations for a profile with n data points, whereas Eq. (1) requires a Cholesky factorization with a computational complexity of $O(n^3)$ [13].

## 5. Experiments and Comparison

Comparing filtering results for a standard surface is helpful for highlighting the advantages or shortcomings of particular filtering algorithms. The comparison to be presented here will use both simulated and practical profiles.

Figure 1 shows an experiment that assesses the linear filtering techniques. A simulated profile composed of 10 harmonic components with frequencies ranging from $0.1/\lambda_c$ to $25/\lambda_c$ is created, where the cut-off wavelength λ_c equals 1600 points. The DCT based linear spline filter and standard spline filter were applied to this simulated profile. The filtering results and differences between the filtered profiles are shown in Figs. 1 (a) and (b) respectively. Figure 1 (b) clearly shows that the mean lines obtained by applying these two filters are almost identical, except at the ends of the surface. The identical results over most of the surface demonstrate the validity and feasibility of the new filtering algorithm. Also, the distortion occurring on both ends demonstrates the effect of the 1st differential boundary condition. In Fig. 1 (a) the actual maximum deviation at the edges is 0.0145 μm, and the largest error is less than 2.02 % relative to the peak-to-peak value of the simulated surface. In fact, for a general measurement, this small error does not significantly impact the assessment of the profile as a whole.

In Fig. 2, a simulated periodic profile is used to validate the spline filtering algorithm based on the FFT. Both Eq. (12) and the standard spline filter with periodic boundary conditions are implemented. As expected, their filtered profiles are identical, even at the boundaries. This perfect result further demonstrates the accuracy and effectiveness of the new algorithm.
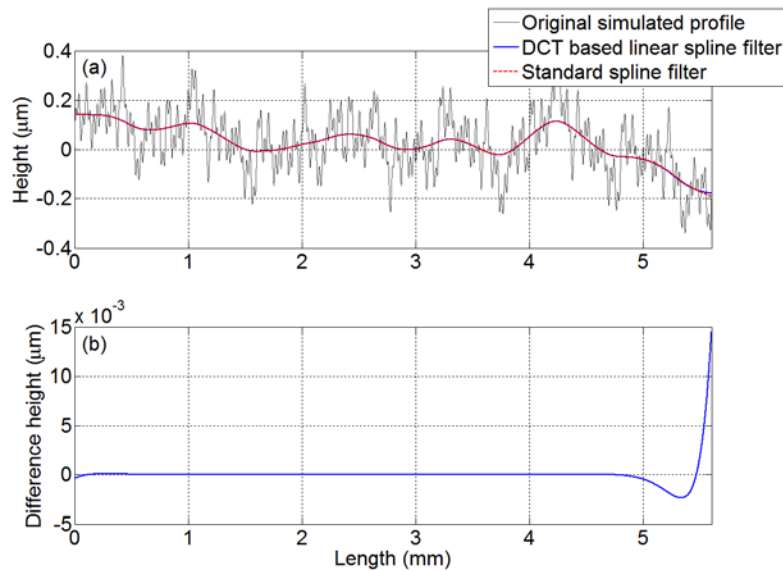
Fig. 1. Simulated profile and mean lines: (a) DCT based linear spline filter and standard spline filter under cut-off wavelength 0.8 mm; (b) Difference of the mean line of the DCT based spline filter relative to the standard filter.
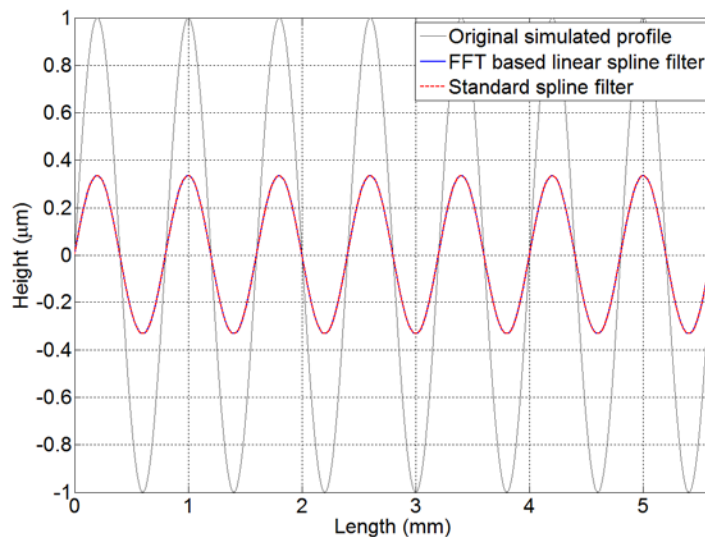

Fig. 2. FFT based linear spline filter and standard spline filter for a periodic function using cut-off wavelength of 0.8 mm. The results from the FFT based algorithm and traditional matrix decomposition are indistinguishable.

Figure 3 shows a more complicated profile with a 5.6 mm length. It is found that there are many outliers such as scratches, deep valleys and sharp spikes distributed along this profile. The DCT based robust spline filter, the L_2-norm based robust spline filter and the standard spline filter are applied to obtain the mean lines with a cut-off wavelength of 0.8 mm. It should be noted that the two robust filters can also easily handle missing data and dropouts if these points are assigned values that deviate far from the profile such that they appear as outliers. From the result in Fig. 3, it is evident that the robust algorithms are resistant to the outliers and able to acquire the satisfactory mean lines that follow the practical form very well.
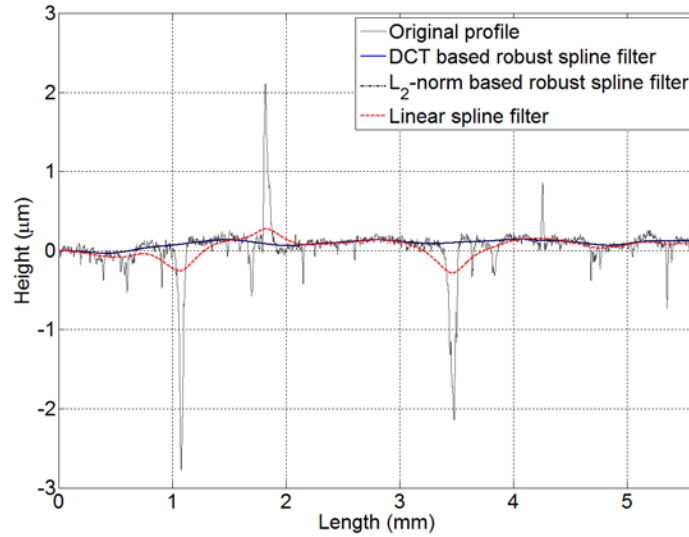
Fig. 3. DCT based robust spline filter, L_2-norm based robust spline filter and standard spline filter with a cut-off wavelength of 0.8 mm applied to data with outliers. In the graph, the filtered curves for DCT based robust spline filter and L_2-norm based robust spline filter are indistinguishable in most parts.

In order to further demonstrate the efficacy of the newly proposed filter, an example consisting of a profile that may be encountered in practice is shown in Fig. 4. This practical engineering profile is 14.3 mm in length with sampling interval 0.125 μm. In this figure, the mean lines determined by the DCT based robust spline algorithm, the L_2-norm based robust spline filter, and the second-order Gaussian regression filter with a cut-off wavelength of 0.8 mm are presented concurrently. The robustness of the DCT based algorithm throughout the entire profile is verified again, but a larger difference between the robust spline filters and the Gaussian regression filter is exhibited in this example. These differences are caused by the dissimilar transmission characteristics of the filters and the second-order polynomial fitting principle in the Gaussian regression filter [17].
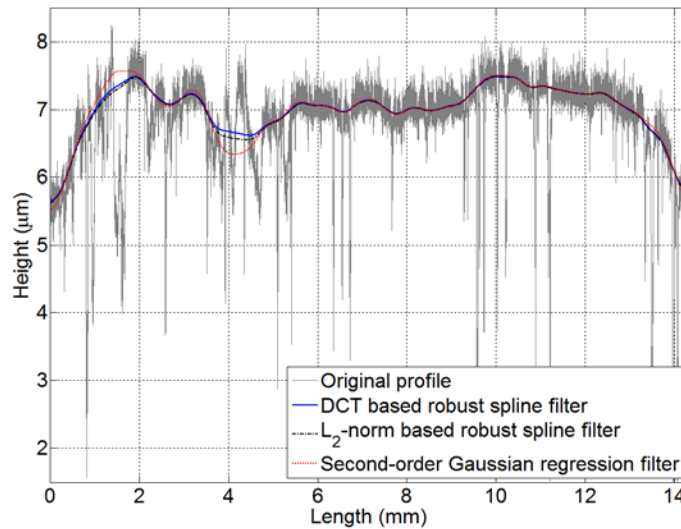
Fig. 4. DCT based robust spline filter, L_2-norm based robust spline filter and Gaussian regression filter with a cut-off wavelength of 0.8 mm applied to a practical profile.

For this example, all of the filtering algorithms were implemented in Matlab and executed on an ordinary, commercially available desktop computer. For the test profile in Fig. 4, which contains 114,400 data points, the second-order Gaussian regression filter requires multiple hours to attain the robustness requirement. In comparison, the L_2-norm based robust spline filter requires 1160 ms to finish the task, and the newly developed DCT robust algorithm takes only 720 ms to filter this profile. While the speed enhancements of the proposed method are only marginal for the profiles investigated here, it is possible that more dramatic improvements can be realized for larger or higher dimension arrays. Nevertheless, the primary benefit of this method is the relative simplicity of computing Eq. (15) and (16) without the requirement of any matrix operation and manipulation.

6. Conclusion

In this paper, several new algorithms for the spline filters are proposed based on the DCT and FFT. Both the linear and robust spline filters for periodic and non-periodic profiles are considered. In the resulting algorithms, the DCT and FFT which are utilized in the filtering process require only $O(n\log(n))$ computations, whereas the traditional algorithm based on Cholesky factorization requires $O(n^3)$. In addition to the potential scaling benefits, the DCT and FFT produce a more straightforward and simple to implement algorithm than the standard spline filter based on matrix decomposition. Meanwhile, although an approximation for the non-periodic boundary condition is utilized, the error caused by the end effects does not affect the surface evaluation significantly. For periodic profiles, the new and faster FFT based algorithm showed no differences when compared to the traditional spline filter.

In brief, these new algorithms for generating spline filters, based on the discrete cosine and discrete Fourier transforms, provide measureable improvements in computational speed and demonstrate excellent robustness to outliers and dropouts that are expected in real world data. Comparison of this new approach with traditional methods directly demonstrates the utility of these new techniques.

Acknowledgments

7. References

[1] J. Raja, B. Muralikrishnan, S. Fu. Recent advances in separation of roughness, waviness and form. Prec. Eng. 2002; 26(2):222-235. http://dx.doi.org/10.1016/S0141-6359(02)00103-4

[2] H. Zhang, Y. B. Yuan, W. Y. Piao. The spline filter: a regularization approach for the Gaussian filter. Prec. Eng. 2012; 36(4):586-592. http://dx.doi.org/10.1016/j.precisioneng.2012.04.008

[3] ASME B46.1-2009. Surface texture (surface roughness waviness, and lay). New York: American Society of Mechanical Engineers 2010.

[4] ISO 16610-21. Geometrical product specifications (GPS)-filtration-part 21: linear profile filters: Gaussian filters; 2011.

[5] ISO/TS 16610-22. Geometrical product specifications (GPS)-filtration-part 22: linear profile filters: spline filters; 2006.

[6] ISO/TS 16610-31. Geometrical product specifications (GPS)-filtration-part 31: Robust profile filters: Gaussian regression filters; 2006.

[7] ISO/TS 16610-32. Geometrical product specifications (GPS)-filtration-part 32: Robust profile filters: spline filters; 2006.

[8] H. Zhang, M. Tong, W. Chu. An areal isotropic spline filter for surface metrology. Journal of Research of NIST 2015; 120:64-73. http://dx.doi.org/10.6028/jres.120.006

[9] I. J. Schoenberg. Spline functions and the problem of graduation. Proceedings of the National Academy of Sciences of the United States of America 1964; 52:947-950. http://dx.doi.org/10.1073/pnas.52.4.947

[10] D. Janecki. A generalized L2-spline filter. Measurement 2009; 42(6):937-943. http://dx.doi.org/10.1016/j.measurement.2009.01.020

[11] H. Li, C. F. Cheung, X. Jiang, W. Lee, S. To. A novel Gaussian filtering method for the characterization of surface generation in ultra-precision machining. Prec. Eng. 2006; 30(4):421-430. http://dx.doi.org/10.1016/j.precisioneng.2006.01.005

[12]   T. Goto, J. Miyakura, K. Umeda. A robust spline filter on the basis of L_2-norm. Prec. Eng. 2005; 29(2):151-161.
        http://dx.doi.org/10.1016/j.precisioneng.2004.06.004
[13]   D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. Computational Statistics and
        Data Analysis 2010; 54(4): 1167-1178.   http://dx.doi.org/10.1016/j.csda.2009.09.020
[14]   W. C. Yueh. Eigenvalues of several tridiagonal matrices. Applied Mathematics E-notes 2005; 5:66-74.
[15]   G. Strang. The Discrete Cosine Transform. SIAM Review 1999; 41(1): 135-147.
        http://dx.doi.org/10.1137/S0036144598336745
[16]   V. Udayashankara. Real time digital signal processing: fundamentals, algorithms and implementation using TMS processor.
        Prentice-Hall of India Private Limited, New Delhi, India (2010): 233-252.
[17]   B. Muralikrishnan, J. Raja. Computational Surface and Roundness Metrology. Springer, London, UK (2009): 67-76.

About the authors: Hao Zhang, Daniel Ott, Mingsi Tong, and Wei Chu are the guest researchers of the Surface & Nanostructure Metrology Group of the Semiconductor and Dimensional Metrology Division (SDMD) of the Physical Measurement Laboratory (PML) at NIST. John Song is a project leader of the Surface & Nanostructure Metrology Group of SDMD of PML at NIST. The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.