Precision Engineering 36 (2012) 586–592

Contents lists available at SciVerse ScienceDirect

Precision Engineering

journal homepage: www.elsevier.com/locate/precision

The spline filter: A regularization approach for the Gaussian filter

Hao Zhang [a,b,*], Yibao Yuan [b], Weiying Piao [b]

[a] College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China
[b] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China

ARTICLE INFO

Article history:
Received 3 July 2011
Received in revised form 15 March 2012
Accepted 21 April 2012
Available online 5 May 2012

Keywords:
Approximating spline filter
Cascade algorithm
Gaussian filter
Surface metrology

ABSTRACT

The Gaussian filter described in the ISO 11562 standard has become the most widely used filtering technique in surface metrology. However, this filter is always plagued by the large distortions called end effects at the boundaries of the filtered result. In order to alleviate the end effects, the spline filter based on natural cubic splines is incorporated into ISO standard as a substitute. So there exist two kinds of linear profile filters with different transmission characteristics which also lead to different mean lines for the assessment of the same surface. A new spline algorithm for determining the Gaussian filtered mean line is deduced using the central limit theorem. The filter uses the cascade method of the approximating spline filter, and therefore can approximate the transmission characteristic of the Gaussian filter with high accuracy. It is proved that the transmission characteristic relative deviation of the cascade approximating spline filter from the Gaussian filter is only 0.3% when the cascade order approaches infinity. With this theorem, it is easy to achieve the unification of the international standard ISO 16610-21 and ISO 16610-22.

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]. The mean line is the reference line used for surface assessment. In surface metrology, the determination of the mean line is a fundamental procedure for describing and assessing surface characteristics [3]. This procedure is usually accomplished using digital filters to wipe off the short wavelength components and to preserve the long wavelength components [4].

In 1996, the ISO 11562 standard proposed the Gaussian filtered mean line as the reference line for profiles measured with contact (stylus) instruments [5]. Though the ISO 11562 standard has recently been replaced by the ISO 16610-21 standard [6], which develops some concepts of the Gaussian filter and a finite support implement, the main theory of this linear profile filter does not change. The Gaussian filter is still recognized as an optimal filter because of its zero-phase characteristic and its minimum product of time width and frequency width [3]. However, it has been known that the Gaussian filter causes excessive end effects, and the Gaussian filtered mean line at the boundary region cannot be used for the assessment. Generally, the first half and the last half cutoff length of the mean line are discarded to eliminate end effects [7]. It means that two halves of cutoff length of valid profile data are also abandoned to ensure the surface assessment accuracy.

To overcome this problem, in 2006, the ISO/TS 16610-22 standard recommended the spline filter as a further substitute for the Gaussian filter [8,9]. The spline filter has two characteristics that are different from other filters. One is the matrix factorization algorithm which results in a higher efficiency of the filtering calculation. The other is the optional boundary conditions during the algorithm. The natural boundary condition and the periodic boundary condition may be selected according to different profiles, such as general engineering surfaces or periodic profiles. In this way, the end effects of the filtered dataset can be effectively alleviated.

From the discussions above, the spline filter has significant advantages over the Gaussian filter. However, in practical use, the spline filter has a different transmission characteristic from the Gaussian filter. Therefore, there exist two sets of standard linear profile filters, and corresponding two kinds of assessment results for the same profile. It has to say that the differences between these two filters' transmission characteristic bring a big problem to the assessment of surface, that is, the assessments of the same profile with two current international standards are not consistent. In fact, the transmission characteristic of the Gaussian filter is sufficient and it should be noted that the Gaussian filter will be implemented in the machine manufacture for a long time in the future. Hence, there is an urgent need to unify the filtering characteristics of the Gaussian filter and the spline filter, or in other words, to

* Corresponding author at: College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China. Tel.: +86 13851425380.
E-mail address: zhanghaowo79@sohu.com (H. Zhang).

0141-6359/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.precisioneng.2012.04.008

approximate the transmission characteristic of the Gaussian filter with the spline filter.

In this work, a novel cascade spline filter for determining the Gaussian filtered mean line is proposed based on the central limit theorem. Not only can the spline advantages such as end effect alleviation, zero-phase characteristic and high computational efficiency be inherited, but also the filtering characteristic of the Gaussian filter can be ensured by this approach.

2. Approximating spline filter

The spline filter was introduced by M. Krystek into surface measurement in 1996 [10,11]. In view of its advantages, the spline filter became a commonly used profile filter and ultimately was included into the international standard. Generally, the physical interpretation of a spline filter is described as, "to find $s(x_i)$ minimizing the L2-norm of residual errors under the collateral condition of minimizing the bending energy of spline", and is formulated by the following expression [9]:

$$\varepsilon = \sum_{i=1}^{N}(y_i - s(x_i))^2 + \mu \int_{x_1}^{x_N} \left| \frac{d^2 s(x)}{dx^2} \right|^2 dx \to \text{Min} \quad (1)$$

where y_i is the measured data under a constant sampling interval Δx, $s(x)$ is the output, i is the index of dataset and N is the total number of measured data. The first component of Eq. (1) is supposed to guarantee that the mean line s is to approximate the profile y. The second component called the spline bending energy is the cost guaranteeing a suitable smoothness of the mean line [12,13]. So Eq. (1) shows a tradeoff between the fidelity and the smoothness of the data. μ, Named Lagrange constant, is a free scalar parameter to control this tradeoff. The solution process is actually a low-pass filter.

The spline filter utilizes the matrix factorization algorithm instead of the convolution operation used by the conventional phase-correct filters, and the ends of the measured profile are still usable without distortion. However, the transmission characteristic of the spline filter is different from that of the Gaussian filter, which means there exist two kinds of surface assessment systems that are inconvenient for the actual measurement.

In 1989, Johannes proposed a regularization approach for the implementation of an approximate Gaussian filter [14]. An item similar to velocity is added into Eq. (1), which can improve its transmission characteristic to fit that of the Gaussian filter.

$$\varepsilon = \sum_{i=1}^{N}(y_i - s(x_i))^2 + \mu \int_{x_1}^{x_N} \left| \frac{d^2 s(x)}{dx^2} \right|^2 + \tau \left| \frac{ds(x)}{dx} \right|^2 dx \to \text{Min} \quad (2)$$

where τ is the approximating coefficient. With adjustment of the tension τ, a set of filters with different amplitude transmission characteristic is constructed. Among these filters, there is a solution likewise close to the filtering characteristic of the Gaussian filter. The solution is called approximating spline filter, in fact, it is a spline approach to implement the Gaussian filter.

Assume that the data points are equidistant and spaced with the unit distance which is suitable for most situations. If we approximate each order derivative of $s(x_i)$ with the form of difference, such as

$$\frac{d^2}{dx} s(x_i) = s(x_{i-1}) - 2s(x_i) + s(x_{i+1}) \quad (3)$$

$$\frac{ds(x_i)}{dx} = s(x_i) - s(x_{i-1}) \quad (4)$$

and denote $s(x_i)$ by s_i, then Eq. (2) can be written as the discrete functional form:

$$\varepsilon = \sum_{i=1}^{N}(y_i - s_i)^2 + \mu \sum_{i=1}^{N}\{(s_{i-1} - 2s_i + s_{i+1})^2 + \tau(s_i - s_{i-1})^2\} \to \text{Min} \quad (5)$$

To determine the solution which minimizes Eq. (5), the partial derivative operation with respect to s_i is performed, that is

$$\frac{\partial \varepsilon}{\partial s_i} = 0 \quad (6)$$

Like the spline interpolation, the boundary condition is necessary to the spline filter, and the boundary condition incurs an obvious influence on the filtering effect. According to ISO/TS 16610-22, the spline filter has two kinds of boundary conditions, nature (non-periodic) condition and periodic condition, which are used in filtering open profiles and closed profiles respectively. The equations for non-periodic spline filter and periodic spline filter are described in different forms because of the boundary conditions.

The nature boundary condition is

$$\nabla^2 s_1 = \nabla^2 s_N = 0 \quad (7)$$

where ∇ represents difference symbol. From Eqs. (5)–(7), the following equations are derived:

$$\begin{cases} \dfrac{\partial \varepsilon}{\partial s_1} = -2(y_1 - s_1) + 2\mu\{s_3 - (2+\tau)s_2 + (1+\tau)s_1\} = 0 \\ \dfrac{\partial \varepsilon}{\partial s_2} = -2(y_2 - s_2) + 2\mu\{s_4 - (4+\tau)s_3 + (5+2\tau)s_2 - (2+\tau)s_1\} = 0 \\ \dfrac{\partial \varepsilon}{\partial s_i} = -2(y_i - s_i) + 2\mu(\nabla^4 s_i - \tau\nabla^2 s_i) = 0 \\ \dfrac{\partial \varepsilon}{\partial s_{N-1}} = -2(y_{N-1} - s_{N-1}) + 2\mu\{s_{N-3} - (4+\tau)s_{N-2} + (5+2\tau)s_{N-1} - (2+\tau)s_N\} = 0 \\ \dfrac{\partial \varepsilon}{\partial s_N} = -2(y_N - s_N) + 2\mu\{s_{N-2} - (2+\tau)s_{N-1} + (1+\tau)s_N\} = 0 \end{cases} \quad (8)$$

For the period data, the boundary condition is given by

$$s_i = s_{i+N}$$

and the more concise equations are derived:

$$\frac{\partial \varepsilon}{\partial s_i} = -2(y_i - s_i) + 2\mu(\nabla^4 s_i - \tau\nabla^2 s_i) = 0 \quad (9)$$

Finally, Eqs. (8) and (9) can be written in the following unified matrix form:

$$(I + \mu Q)S = Y \quad (10)$$